

# Des services spécifiques pour les projets et les infrastructures de recherche (LHCONE) : Multi-Domain Multi-Point VPN et MPLS-TE

## Xavier Jeannin

GIP RENATER  
23-25, rue Daviel  
75013 PARIS

## Nicolas Garnier

GIP RENATER  
23-25, rue Daviel  
75013 PARIS

## Jérôme Bernier

Centre de Calcul IN2P3/CNRS -  
Domaine scientifique de La Doua  
43 bd du 11 Novembre 1918  
69622 Villeurbanne Cedex

## Résumé

*La recherche et ses infrastructures (grille de calcul, télescope ...) sont mises en œuvre au sein de collaborations internationales. En plus des données, d'autres usages du réseau se développent comme l'échange de jobs, de machines virtuelles et il est vital de maintenir ces services opérationnels pour garantir la production scientifique. Les réseaux privés virtuels (VPN) renforcent ces infrastructures en fournissant une isolation et une protection du réseau, et mettent ainsi en œuvre le concept de zone protégée d'échange « scientist DMZ ». Grâce à cette zone, on économise les coûteux pare-feu, ce qui permet d'atteindre des performances de débits jusqu'alors inatteignables.*

*En isolant les trafics au sein d'un VPN, les réseaux nationaux d'éducation et de recherche (NREN) peuvent y appliquer des traitements particuliers comme la mise en place de bandes passantes dédiées, de redondance, la détermination des chemins utilisés, etc. L'utilisation du Traffic Engineering MPLS est désormais disponible sur RENATER et optimise les ressources réseau de RENATER tout en permettant un déploiement très flexible du VPN.*

*La méthode dite « back-to-back » a été utilisée pour interconnecter les différents L3VPN (IP) déployés dans chacun des NRENs qui composent le LHCONE, le L3VPN multi-domaine de la communauté High Energy Physics. La mise en place du LHCONE a permis d'augmenter considérablement les échanges inter-sites et d'améliorer la production scientifique. Enfin, les VPNs multi-domaines constituent un nouveau service de RENATER utilisable par de nombreux projets éducatifs et scientifiques. RENATER poursuit son travail par la mise en place de nouveaux L3VPN multi-domaine plus rapides à déployer, le service MD-VPN (Multi-Domaine VPN).*

## Mots-clefs

*Multi-Point VPN, MPLS, Traffic Engineering*

## 1 Introduction

Les grandes expériences scientifiques ou les infrastructures scientifiques (grille de centres de calcul, télescopes, ...) sont désormais mises en œuvre au sein de collaborations internationales, c'est-à-dire que par nature, elles sont réparties sur plusieurs domaines d'administration réseau. Ces projets nécessitent d'échanger des données, des « jobs », de machines virtuelles et ils ont de plus besoin de garder leurs services opérationnels pour le maintien en production de leur infrastructure (par ex. système de fichier distribué, des programmes de supervision et de réservation de ressources, ...).

Ces projets peuvent avoir des besoins :

- Liés à la confidentialité et la sécurité des échanges de données notamment dans le cas d'un partenariat sur des brevets avec des entreprises à vocation commerciale
- De performance ; l'utilisation de pare-feu (deep inspection) à très haut débit (10Gbps) dégrade les performances et peut être complètement bloquante dans certains cas
- Spécifiques réseau liés au débit, à la latence ou à la gigue

En isolant les trafics au sein d'un VPN, les fournisseurs réseau (SP) peuvent appliquer des traitements particuliers aux trafics de ce VPN, par exemple : mise en place de bandes passantes dédiées, détermination des chemins utilisés, mise en place de redondance ...

La mise en place d'un VPN permet

- La création d'une DMZ pour la communauté utilisatrice (scientist DMZ) (meilleure sécurité et performance)
  - Les utilisateurs gèrent eux-mêmes ce réseau virtuel selon les règles qui conviennent le mieux à leur recherche ou à leur besoin
- De renforcer la sécurité des sites et ainsi de soulager l'équipe technique réseau des sites de certaines tâches
- D'établir entre les utilisateurs et les fournisseurs réseau un partenariat rapproché qui permet de mieux adresser les besoins des utilisateurs

Dans cet article, nous allons présenter à travers l'exemple de la mise en place d'un L3VPN multi-domaines pour la communauté HEP (High Energy Physics) nommé LHCONe [2 - 3], les possibilités qu'un tel outil peut avoir pour la communauté recherche et éducation (partenariat international inter-université, projets de recherche internationaux). L'utilisation du *Traffic Engineering*, désormais disponible sur RENATER sera décrite. Ce service de VPN multi-domaine s'est révélé très utile pour la communauté HEP. Il est important de **faire la promotion de ce nouveau service auprès des autres communautés scientifiques**. RENATER travaille au déploiement d'une nouvelle version plus VPN multi-domaine qui seront plus facile et plus rapide à déployer.

## 2 LHCONe exemple d'un L3VPN multi-domaines

### 2.1 Problématique

Afin de pouvoir réaliser les énormes quantités de calculs nécessaires à leur recherche, les expériences de LHC (ATLAS, CMS, LHCb, ALICE) ont mis en place une organisation complexe divisée en service (service d'authentification, catalogue, gestion des jobs, service de transfert de données ...), le réseau est considéré comme un de ces services au sein de cette organisation.

Le *data workflow* décrit comment les informations sont échangées entre les centres de calcul, les sites. Le premier modèle de « *data workflow* », MonArch, était hiérarchique, schématiquement les données étaient transférées depuis le CERN (Tier 0) vers des grands sites nommés Tier 1, c'est le travail du réseau dédié LHCONe qui d'acheminer les données depuis le CERN vers les Tiers 1 et entre Tiers 1. Les autres sites (Tiers 2 et 3) utilisaient principalement les données de leur Tiers 1 national ou régional. Désormais, les expériences (ATLAS, CMS, LHCb, ALICE) veulent pouvoir utiliser les données depuis n'importe quel site. D'autre part, de nouveaux modes de calcul tendent à ce que les jobs utilisent les données à distance. Enfin, la taille des données manipulées par les expériences est plus importante que prévu. L'ensemble de ces points a débouché sur la création d'un réseau dédié LHCONe qui compléterait le LHCONe et viserait le transfert de données entre les 12 Tiers 1 et les 130 Tiers 2, et aussi entre Tiers 2 eux-mêmes.

### 2.2 Design du LHCONe, un back-to-back MD-MP-L3VPN

Le LHCONe est un réseau L3VPN multi-point et multi-domaine. Les différents domaines (NRENs ou service provider SP) utilisent plusieurs technologies différentes en fonction des technologies présentes dans leur domaine et en fonction des moyens et des besoins de leurs sites utilisateurs. La plupart des NRENs ont utilisé un réseau en overlay au-dessus de leur propre infrastructure. RENATER a, pour sa part, utilisé une infrastructure dédiée pour la majeure partie des sites avec la fourniture de deux peerings un à Paris (10 Gbps) et un à Genève (20 Gbps) avec le LHCONe international. Les sites qui n'ont pas un trop gros besoin de bande passante sont connectés en overlay grâce à des L2VPNs point à point (pseudowire) dont le chemin est déterminé grâce à l'usage de MPLS-Traffic Engineering.

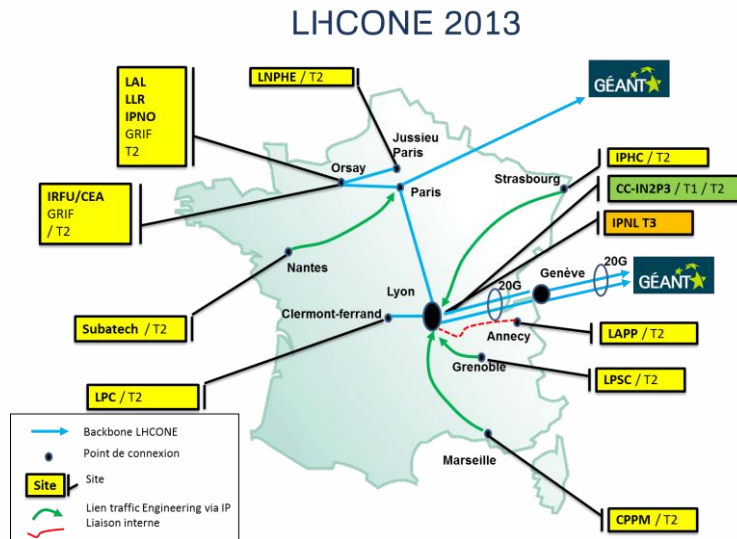


Figure 1 : LCHONE France

Les différents NRENs sont interconnectés entre eux par des peerings BGP simples parfois multiples. Au niveau signalisation, il n’y a pas d’échange entre les domaines autres que des échanges de routes IPv4 (IPv6 est en cours de déploiement) des utilisateurs du VPN via eBGP VPNv4. Cette architecture est appelée back-to-back.

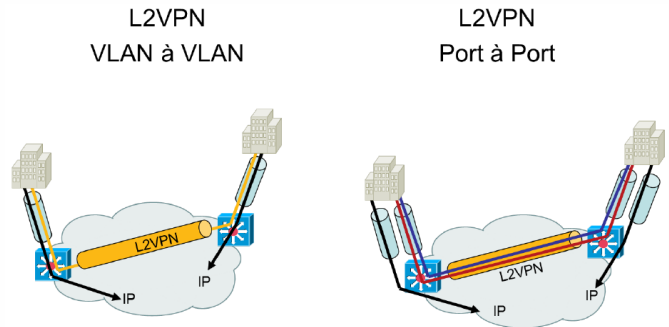
### 3 LHCONE France

#### 3.1 Bref rappel sur les VPN MPLS/BGP

RENATER propose des solutions d’interconnexion privée de niveau 2, L2VPN Ethernet et de niveau 3 L3VPN MPLS.

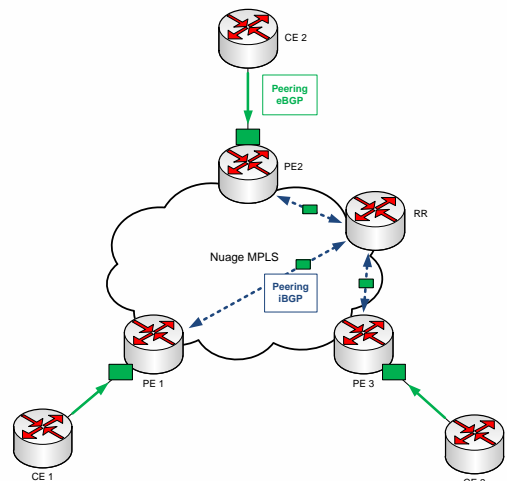
- Les **L2VPN** reposent sur la technologie « Virtual Private Wire Service » (VPWS), ou « Pseudo-Wire » (L2VPN-PW) qui construit un circuit Ethernet point à point à partir du cœur MPLS. Deux types de L2VPN PW peuvent être configurés : VLAN-à-VLAN et port-à-port.

Dans le cas d’un L2VPN VLAN-à-VLAN où les 2 sites interconnectés souhaitent communiquer sur plusieurs VLAN, on utilisera la solution du QinQ (IEEE 802.1ad) qui permet de dissocier les VLAN clients du VLAN de transport.



- Les **L3VPN** reposent sur la technologie MPLS-VPN. Le mode utilisé est « any-to-any », chaque site connecté au L3VPN peut dialoguer directement avec un autre site, sans devoir passer par un site central. Les informations de routage sont propagées via BGP aux routeurs d’extrémité.

La solution L3VPN MPLS présente sur RENATER-5 repose sur l’exploitation de l’en-tête MPLS, où l’on définit des VPN discriminés par un label supplémentaire (en plus du label de commutation). Chaque VPN possède sa propre table de routage IP dans le concept de Routage et Transfert Virtuel « Virtual Routing and Forwarding » (VRF) impliquant une



notion de « Route Distinguisher » et « Route target » (RD et RT). (RFC 4364) – [1].

Le protocole de routage utilisé est BGP. Celui-ci échange les routes annoncées dans le VPN via son extension Multi-Protocol VPNv4 BGP (MPBGp). Comme pour le fonctionnement traditionnel de BGP, il est possible d'utiliser ou non un route reflector (RR).

### 3.2 Architecture

Les sites peuvent être raccordés au backbone LHCONE soit directement si le débit le nécessite, soit en overlay via des pseudowires. Dans ce second cas, deux pseudowires sont montés entre le site et les routeurs de Paris et Lyon. On choisit la sortie privilégiée (soit Paris, soit Genève), ce qui permet de partager la charge tout en offrant une bonne redondance via BGP. Le chemin de ce pseudowire est indépendant de l'IGP de RENATER grâce à la mise en place du MPLS-TE et peut être choisi sur des liens peu occupés du backbone d'où une optimisation des ressources réseau de RENATER.

#### Connexions des sites Tiers 2 via L2VPN PseudoWire

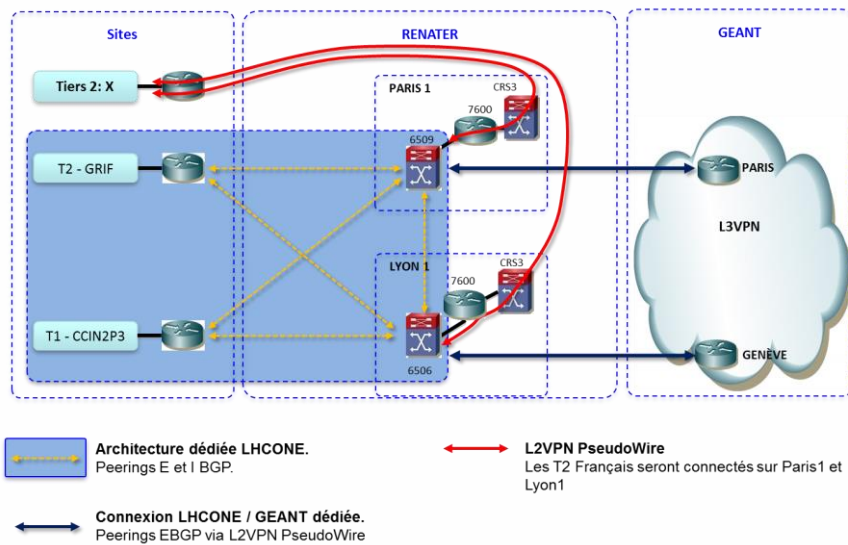


Figure 2 : raccordement de site via un pseudowire

Le réseau LHCONE est principalement un réseau dédié. Les sites raccordés directement au backbone bénéficient de la même manière des deux sorties ce qui leur offre une bonne résilience et permet la répartition de charge.

#### LHCONE FR BGP policy

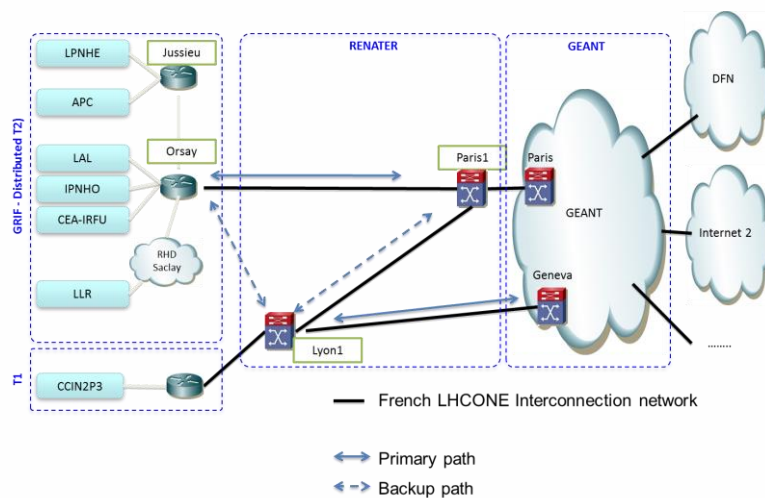


Figure 3 : routage du backbone LHCONE

### 3.3 Utilisation du MPLS-TE pour la connexion des Tiers 2

L'interconnexion des sites Tiers 2 au réseau LHCONe est effectuée via des L2VPN-PW vlan-à-vlan. Plusieurs points sont à souligner quant à ces connexions :

- La consommation moyenne par site Tier 2 est estimée début 2012 à plus d'1 Gbit/s. On constate en septembre 2013 une utilisation moyenne de 4 à 6 Gb/s sur certains T2, ainsi que des points réguliers à 10 Gb/s.
- La cohabitation des L2VPN-LHCONe avec les autres utilisateurs de RENATER est possible sur le réseau de production IP/MPLS 10 Gb/s. Cependant, le risque de saturation et de manque de disponibilité du réseau à cause des L2VPN-LHCONe n'est pas négligeable.

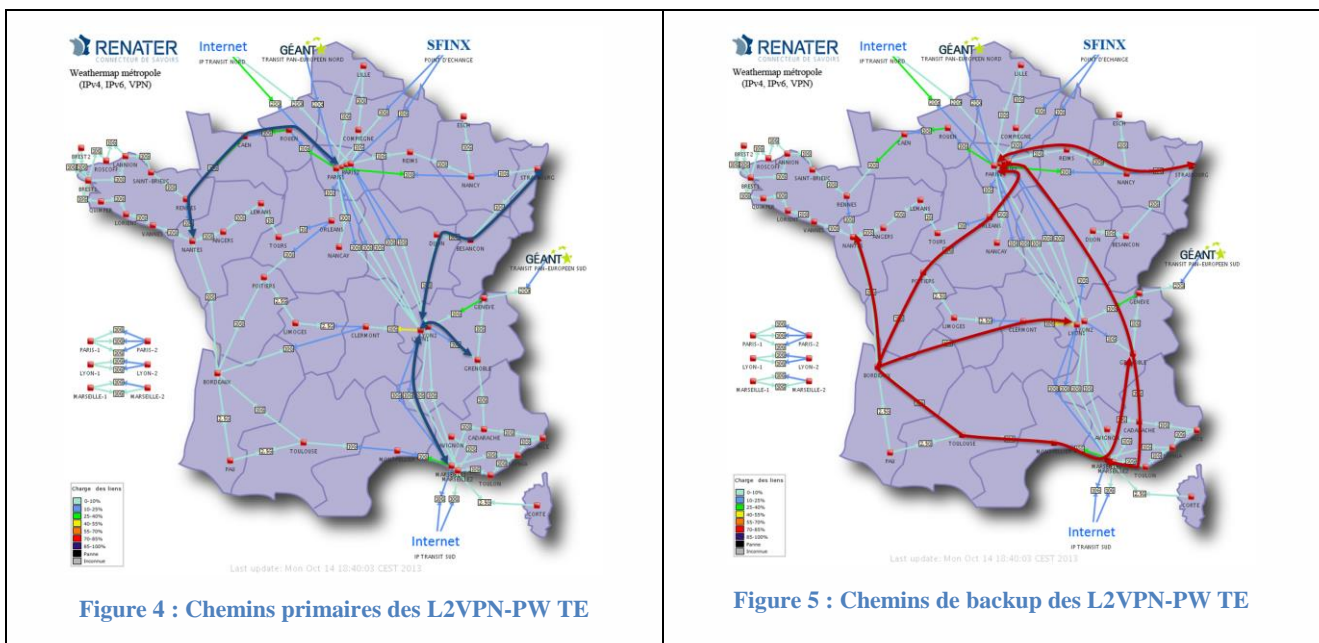
Pour des questions de fiabilité, tous les NR sont au moins 2-connectés aux autres NR. Dans la plupart des cas, on peut parler d'une liaison principale, et d'une liaison de secours.

- Les liaisons de secours sont peu utilisées par la politique de routage actuelle.
- Les liaisons de secours sont aussi performantes que les principales (10 Gb/s).

Des mécanismes d'ingénierie de trafic MPLS-TE sont alors utilisés pour orienter les liaisons L2VPN-PW vers le « LHCONe France », au travers de ces liaisons de secours. Le MPLS-TE est utilisé pour l'instant en mode tactique dans RENATER. Les avantages sont les suivants :

- Utilisation/rentabilisation des liaisons de secours.
- Allègement, non saturation des liaisons principales du réseau RENATER.

La mise en place des mécanismes MPLS-TE a fait l'objet d'un projet d'étude, de conception et déploiement entre octobre 2011 et juin 2012. Nous avons étroitement collaboré avec les équipes techniques de Cisco pour valider la compatibilité avec le backbone de production de RENATER.



### 3.4 Déploiement dans les sites de l'IN2P3 et politique de routage

Du point de vue des sites utilisateurs, la disponibilité d'un réseau privé de type L3VPN est très importante dans le cadre de grands projets nationaux et internationaux comme le LHC.

Elle permet en effet de créer une communauté d'utilisateurs de confiance, en fait un ensemble de sous-réseaux qui adoptent la même politique de sécurité. Ce réseau privé permet de séparer le trafic de cet espace de confiance du trafic de l'Internet généraliste, avec 2 conséquences principales: un traitement de la sécurité différent et la possibilité de mettre en place des liaisons dédiées, notamment afin de faire passer les très gros volumes de transferts de données. La suppression de coûteux pare-feux est aussi un avantage et permet de meilleure performance à très hauts débits.

Un site éligible à s'interconnecter à ce L3VPN doit disposer d'une connexion de niveau 2 jusqu'aux points de présence Paris et Lyon, pour pouvoir établir ses peerings BGP. Si on dispose d'un lien direct sur l'équipement de RENATER, on va pouvoir rajouter facilement des vlans, sinon le réseau métropolitain/régional utilisé pour connecter le site à RENATER devra pouvoir nous mettre à disposition de tels vlans. RENATER se chargera alors de propager ses vlans jusqu'aux routeurs du L3VPN. On pourra alors établir nos sessions BGP entre le site et les routeurs du backbone LHCONE Français et définir une politique de routage et de secours sur ces liens.

Sur le réseau L3VPN LHCONE, il a été décidé au niveau international d'appliquer la politique de routage suivante:

- on annonce uniquement les sous-réseaux dédiés au LHC
- on accepte tous les sous-réseaux qui nous sont annoncés

Afin de garantir notre espace de confiance, seuls les trafics entre les sous-réseaux dédiés au LHC sont autorisés sur le L3VPN. Il faudra donc faire très attention au routage et notamment éviter tout routage asymétrique.

Ainsi il faut s'assurer que :

- le trafic provenant des postes de travail (souvent cible de virus et autres cheval de Troie) n'utilisera pas ce réseau, même pour contacter des machines sur LHCONE
- que seules les machines dédiées, par exemple aux transferts de fichiers, utiliseront ce L3VPN.

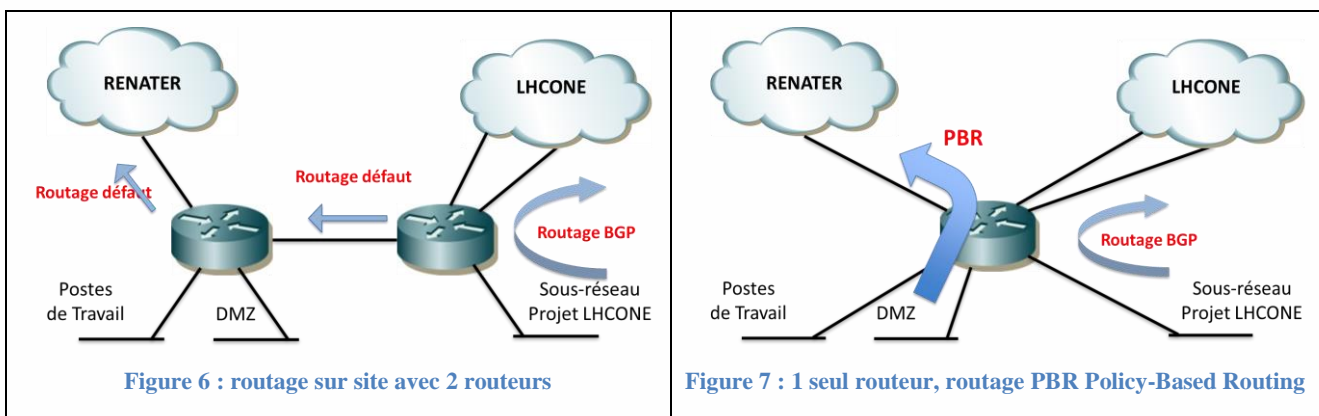
Comme la liste des sites connectés sur LHCONE est dynamique (nouveaux sites, problèmes réseaux, ...) il faut se baser sur BGP pour le routage de nos données. Afin d'assurer la séparation des trafics, la solution la plus simple est de disposer de 2 routeurs, l'un routant le trafic des sous-réseaux dédiés au projet, l'autre le trafic du reste du laboratoire.

Malheureusement ceci n'est pas toujours possible, et surtout onéreux, donc nous devons utiliser des techniques de PBR Policy-Based Routing pour s'assurer de la séparation des trafics.

Une possibilité est de forcer le routage des paquets IP avec l'algorithme suivant:

- si je ne suis pas originaire d'un sous-réseau du projet LHCONE
- et que je ne suis pas à destination d'un sous-réseau local
- alors ma prochaine destination est la sortie sur l'Internet généraliste

Ainsi un paquet originaire d'un sous-réseau lambda ne pourra utiliser LHCONE, et un paquet originaire d'un sous-réseau du projet LHCONE suivra la table de routage et utilisera préférentiellement LHCONE ou par défaut le routage par l'Internet généraliste.



Avec cette configuration on s'assure de la séparation des trafics, du passage de nos paquets par le pare-feu quand le trafic n'est pas dans notre L3VPN, du secours automatique de nos trafics projet sur les différents liens qui nous connectent au LHCONE, ou à travers le réseau généraliste en cas d'indisponibilité du réseau LHCONE.

Suivant la taille du site, il est ainsi très facile de mutualiser tous nos trafics sur le même lien physique qui nous connecte à RENATER, puis en cas de très gros volumes de transferts, de séparer ces trafics sur des liens distincts.

## 4 LHCONE international

### 4.1 Contexte

Le LHCONE est une interconnexion de L3VPN en mode back-to-back. Le principe est que chaque domaine considère les autres comme un client du VPN et chaque domaine envoie les routes clientes du VPN qu'il connaît. Cette architecture offre une très grande indépendance aux membres du VPN et un très bon niveau de sécurité, ce qui s'est avéré très utile dans ce contexte. Néanmoins, cette architecture est très lente à mettre en place car il faut négocier la connexion à chaque peering d'un nouveau membre, c'est le principal défaut de cette architecture.

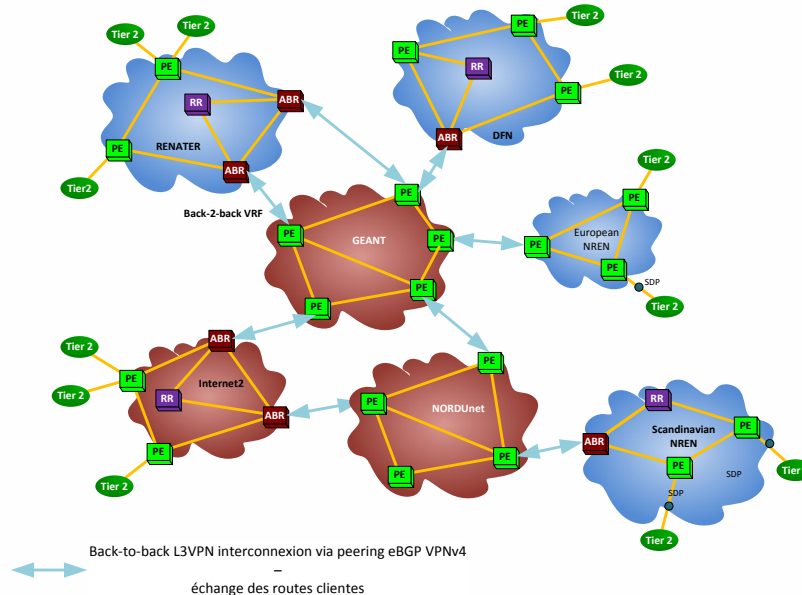


Figure 8 : Back-to-back interconnexion

### 4.2 Exploitation

Le LHCONE est dirigé de manière informelle par les utilisateurs conseillés par les NRENs. Les décisions sont prises lors de 2 à 3 réunions annuelles sur le mode du consensus. Il en résulte souvent un certain flou sur les choix opérationnels, en matière de politique de sécurité et de routage.

- Le modèle opérationnel n'est pas bien défini et est toujours en cours d'élaboration. Pour les opérations de dépannage, qui sont les plus cruciales, c'est le modèle classique appliqué pour les peerings IP entre NRENs qui est utilisé.
- Le monitoring est décentralisé, il est néanmoins possible aux utilisateurs de consulter plusieurs weathermap souvent fournies par leur propre NREN [4 - 5], mais au-delà, il faut qu'ils aient connaissance de la topologie du réseau.
- La politique de routage souffre aussi de la faiblesse de la gouvernance et il est difficile de bannir certaines pratiques nuisibles au fonctionnement du réseau
  - Les utilisateurs Français pour leur part publient leurs préfixes dans la base RIPE afin que les autres partenaires puissent faire des contrôles de sécurité
- En manière de sécurité, il a été demandé de filtrer en entrée du LHCONE les préfixes valides ; cette politique de sécurité repose sur le fait que tous les réseaux l'appliquent
  - RENATER filtre en fonction des préfixes publiés dans la base RIPE par ces utilisateurs

## 5 Bilan du LHCONE

Le succès du LHCONE se traduit par une augmentation très importante de son usage et une amélioration des performances proposées aux utilisateurs. Le nombre de site connectés au LHCONE dans le monde est désormais supérieur à 100, en France 13 laboratoires CNRS/CEA sont connectés.

Les sites sont classés par les projets scientifiques notamment en fonction de leur accès réseau. Ce classement est primordial pour les sites car un bon classement offre la possibilité de réaliser beaucoup plus de calculs pour la grille donc **d'améliorer la performance scientifique du site**. Les sites français ont bénéficié d'une forte amélioration de leur connexion grâce au LHCONE et ont ainsi obtenu un classement nettement meilleur.

laboratoires	NR RENATER	Type de connexion
SUBATECH	Nantes	L2VPN-PW
LAL - LLR - IPNO (GRIF)	Orsay	Backbone LHCONE
IRFU-CEA (GRIF)	Orsay	Backbone LHCONE
LPNHE	Jussieu	Backbone LHCONE
IPHC	Strasbourg	L2VPN-PW
CC-IN2P3	Lyon1	Backbone LHCONE
LPSC	Grenoble	L2VPN-PW
IPNL - LAPP	Connexion via le CC	Backbone LHCONE
LPC	Clermont-Ferrand	Backbone LHCONE
CPM	Marseille1	L2VPN-PW

Les améliorations fournies par le LHCONE se sont traduites **par une meilleure performance du travail des utilisateurs** en termes de nombre de jobs traités et correctement terminés.

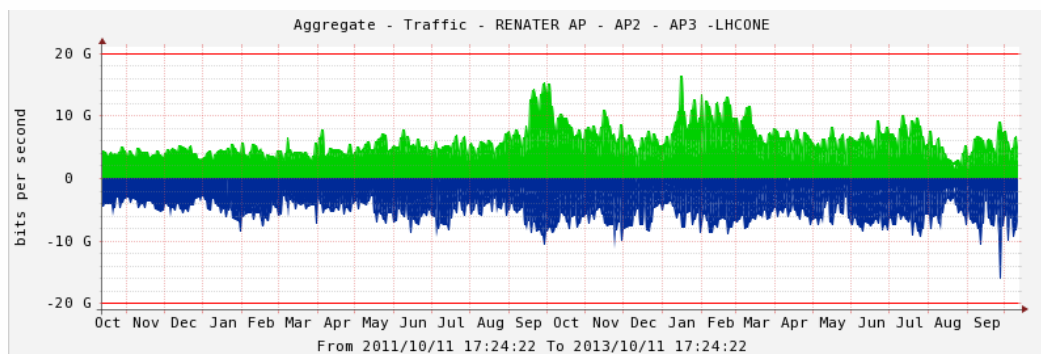


Figure 9 : graphe cumulé des interfaces RENATER-LHCONE vers GEANT-LHCONE

## 6 Conclusion et perspectives pour les L3VPNs multi-domaines

Le déploiement du LHCONE est un succès. Les VPNs multi-domaines sont un nouveau service fournis par les NRENs pour les utilisateurs scientifiques que de nouvelles communautés scientifiques comme ITER ou PRACE pourraient utiliser. RENATER pilote au niveau européen la conception et le déploiement d'une nouvelle solution de service de multi-domaine VPN dit sans couture (seamless) qui permettra à terme de déployer des VPN L2, L3, point-à-point et multi-point très rapidement à travers l'ensemble du réseau GEANT et RENATER. Les Réseaux Régionaux pour offrir ainsi à leurs utilisateurs ces nouvelles fonctionnalités.

L'introduction de mécanisme de Traffic-Engineering sur le backbone de RENATER offre aussi de nouvelles fonctionnalités et perspectives. La gestion automatisée des mécanismes MPLS-TE et leur association à des mécanismes de qualité de service seront étudiées au sein de backbone RENATER.

## 7 Bibliographie

1. RFC 4364 BGP/MPLS IP Virtual Private Networks (VPNs)
2. <http://lhcone.net>
3. <https://twiki.cern.ch/twiki/bin/view/LHCONE/WebHome>
4. [http://pasillo.renater.fr/weathermap/weathermap\\_lhcone\\_france.html](http://pasillo.renater.fr/weathermap/weathermap_lhcone_france.html)
5. <http://netstat.in2p3.fr/weathermap/prov-new.html>