VPLS : Virtual Private LAN Service

Jean-Marc Uzé
Juniper Networks
Espace 21 - 31 Place Ronde - 92986 Paris la Défense - France
juze@juniper.net
14 Octobre 2003

Résumé

VPLS signifie Virtual Private LAN Service. VPLS est un service Ethernet multipoint-à-multipoint (MP2MP) et utilise IP ainsi qu'un mécanisme de tunnel (généralement MPLS) pour apporter une connectivité entre plusieurs sites à travers un nuage IP, comme si ces sites étaient reliés par un même LAN Ethernet. L'organisme connecté obtient un simple service Ethernet de la part de l'opérateur (MAN, backbone national…) dont le réseau apparaît comme un large domaine de broadcast, et le client ne voit pas les détails de l'émulation du LAN. VPLS permet à l'opérateur de service de délivrer un service d'interconnexion de LAN entre plusieurs sites au niveau d'un réseau métropolitain ou à travers un réseau longue distance (éventuellement reliant des réseaux métropolitains distants) en exploitant la capacité de déploiement de IP/MPLS à grande échelle. Le trafic du client est commuté sur la base de l'adresse MAC des paquets Ethernet, et le service VPLS peut transporter IPv4 et IPv6 ainsi que des protocoles expérimentaux, ou à l'opposé des protocoles hérités d'autres technologies comme IPX et DECnet.
Cet article présente la technologie VPLS développée dans le cadre de l'IETF (Internet Engineering Task Force). Une attention particulière est portée sur les aspects service de bout en bout traversant plusieurs domaines IP, comme c'est généralement le cas pour les réseaux de Recherche & Enseignement.

Mots clefs

VPLS, Ethernet, multipoint-à-multipoint, IP, BGP, MPLS, Inter-domaine

1 Introduction

L'Ethernet est la technologie LAN (Local Area Network) la plus largement déployée dans les réseaux actuels. Ces dernières années les standards Ethernet ont évolué de manière très significative, d'une part en ce qui concerne le débit géré par cette technologie, avec une croissance impressionnante de 10 Mb/s à 10 Gb/s, et d'autre part par les améliorations du protocole pour étendre l'usage de l'Ethernet au contexte de réseau longue distance (WAN : Wide Area Network), appelé souvent Métro Ethernet. Les opérateurs qui proposent ce type de service offrent généralement des connexions Ethernet point-à-point entre les différents sites du client sur une zone métropolitaine. Cependant, le but ultime du Métro Ethernet est de proposer une solution permettant d'aller au-delà pour finalement offrir une connectivité multipoint-à-multipoint sur une zone métropolitaine ou entre plusieurs zones métropolitaines à travers un réseau longue distance.
En d'autres mots cela consiste à pouvoir offrir un service à un ensemble de sites (par exemple des universités ou éventuellement un sous-ensemble comme un laboratoire de chaque campus universitaire) de sorte que tous ces sites apparaissent comme étant sur le même LAN Ethernet, indépendamment du fait qu'ils soient sur la même zone métropolitaine ou sur des zones distantes.
Une des solutions les plus largement défendues pour apporter ces services est connue sous le nom de Virtual Private LAN Service (VPLS), qui propose une connectivité Ethernet multipoint-à-multipoint intra ou inter-zone métropolitaine au-dessus d'un service IP/MPLS.

2 Offres actuelles de service Ethernet

Les offres actuelles de services Ethernet sont de manière générale assez limitées. De nombreux opérateurs de service supportent simplement des connexions point-à-point apportant des accès Internet dédiés ou des interconnexions privées entre sites d'une même zone métropolitaine. Certains opérateurs supportant peu de clients ont déployé des services LAN Ethernet multipoint-a-multipoint au sein d'une zone métropolitaine ; ainsi plusieurs sites d'une même zone métropolitaine sont connectés comme si ils étaient sur un LAN unique, utilisant les VLAN IDs pour une séparation logique du trafic. Cependant, étant donné que la plupart de ces opérateurs délivrant des services Ethernet métropolitain ont construit leur infrastructure sur la base de commutateurs Ethernet, un certain nombre de problèmes inhérents à ce type de solution apparaissent aujourd'hui sur de larges réseaux métropolitains.
D'une part ces réseaux sont limités en nombre de clients pouvant être supportés : ils ne peuvent en effet supporter que 4096 VLAN IDs. Un VLAN ID est requis par client, et comme les VLAN IDs ont une signification globale, ils doivent être uniques au sein du réseau de l'opérateur. Il est évident qu'il n'est pas possible d'envisager d'apporter ce service de LAN à travers plusieurs réseaux métropolitains distants avec ce type d'architecture, car cela nécessiterait de construire un réseau de niveau 2 Ethernet encore plus large. Il existe une approche consistant à encapsuler des VLANs dans d'autres VLANs, mais le manque de standardisation dans ce domaine oblige à un déploiement mono-constructeur.
D'autre part ces réseaux à base de commutateurs offrent généralement des solutions de QoS très limitées, voire pas de possibilité de différencier les paquets. Enfin l'opérateur de service dépend du protocole de Spanning Tree (STP) pour traiter la résilience des liens (ou plus exactement les problèmes de boucle dans le réseau), ce qui s'avère peu optimum en termes de partage de charge et de temps de convergence.
C'est pourquoi le déploiement de services Ethernet multipoint-à-multipoint traversant plusieurs réseaux métropolitains éventuellement reliés par des réseaux nationaux et européens n'est pas réaliste aujourd'hui avec cette approche.

Intéressons-nous un moment au cas des entreprises : étant donné que de nombreuses entreprises multi-sites sont limitées par des connexions point-à-point à travers plusieurs zones métropolitaines, elles optent généralement pour une topologie réseau « hub and spoke » au niveau de la connectivité longue distance. Pour une entreprise traditionnelle, les sites distants sont reliés aux sites centraux (typiquement le siége), eux-mêmes reliés aux ressources stratégiques tels que les centres de calculs ou de stockage des données. D'après une étude récente réalisée par Infonetics Research, « User Plans for WAN and Internet Access, US/Canada 2002 », il apparaît qu'une entreprise moyenne dispose de 20 sites, ceci incluant le siège, les centres de données, et les bureaux distants. Cela veut dire que le besoin en liaisons pour une entreprise moyenne correspond à 20 liaisons primaires et 20 liaisons de secours pour son réseau en architecture « hub and spoke ». Il y a un certain nombre d'inconvénients à ce type d'architecture :

- L'architecture surcharge l'opérateur du service avec la gestion de nombreuses liaisons point-à-point, nécessitant des ressources en personnel et des coûts opérationnels.
- Chaque fois qu'un site distant est ajouté, il est nécessaire de configurer à la fois l'équipement CPE[1] du nouveau site et l'équipement CPE du site central
- Une panne au niveau du hub implique une panne totale du réseau de l'entreprise (une entreprise peut contourner ce problème en déployant plusieurs hubs, ce qui renforce les deux premiers inconvénients cités ci-dessus)
- Le trafic site à site nécessite fréquemment de traverser deux fois le réseau de l'opérateur de liaisons, en passant par le site hub, ce qui implique un besoin toujours croissant en bande passante pour les connexions du hub.
- Des délais de transaction peuvent apparaître dès lors que des congestions se manifestent au niveau du hub lorsque plusieurs sites distants accèdent aux ressources centrales via le réseau et excèdent la limite en bande passante du hub.
- L'utilisation du protocole de Spanning Tree (STP) pour gérer à la fois la fonction de détection des boucles ainsi que la fonction de convergence rapide peut causer des difficultés. En effet, bien que STP soit efficace pour détecter des boucles Ethernet dans le réseau, il s'avère limité en terme de convergence en cas de rupture de lien.

Dans un contexte de réseaux de Recherche et Enseignement, les même problèmes peuvent se poser pour le raccordement de communautés particulières, par exemple des sites administratifs d'un organisme, ou des laboratoires distants travaillant sur un projet commun de recherche et nécessitant une infrastructure réseau fédératrice (virtuellement) dédiée. D'autre part il peut s'agir d'une infrastructure réseau métropolitaine et collective supportant des communautés très distinctes (Recherche & Enseignement, Santé, Culture, pépinières d'entreprises, mairies et collectivités locales, etc…).

[1] Customer Premises Equipment

3 VPLS : Virtual Private LAN Service

VPLS délivre un service Ethernet multipoint-à-multipoint pouvant être indifféremment délivré au niveau d'une infrastructure métropolitaine ou sur des réseaux longue distance, et apporte une connectivité entre plusieurs sites comme si ces sites étaient reliés par un même LAN Ethernet.
Contrastant avec l'offre de service Ethernet multipoint-à-multipoint apportés par une infrastructure d'opérateur basée sur des commutateurs Ethernet, VPLS utilise une infrastructure IP/MPLS.
Du point de vue de l'opérateur, l'utilisation des protocoles de routage IP/MPLS au lieu d'une technologie Ethernet basée sur le Spanning Tree Protocol, ainsi que l'utilisation des labels (étiquettes) MPLS au lieu des identités VLAN, apportent à l'infrastructure de l'opérateur une souplesse et une capacité de déploiement de services Métro Ethernet à grande échelle.

Ce nouveau service permet aux réseaux de Recherche et Enseignement d'envisager une approche innovante pour rapprocher les chercheurs et enseignants souvent dispersés sur des sites distants, parfois à l'échelle européenne. Le réseau peut devenir un outil étendant les capacités du LAN traditionnel Ethernet, en s'affranchissant de la distance et du nombre de réseaux intermédiaires (domaines IP). Ainsi il apporte une synergie supplémentaire favorisant le partage et l'échange des ressources et connaissances au niveau régional, national, européen et au-delà.

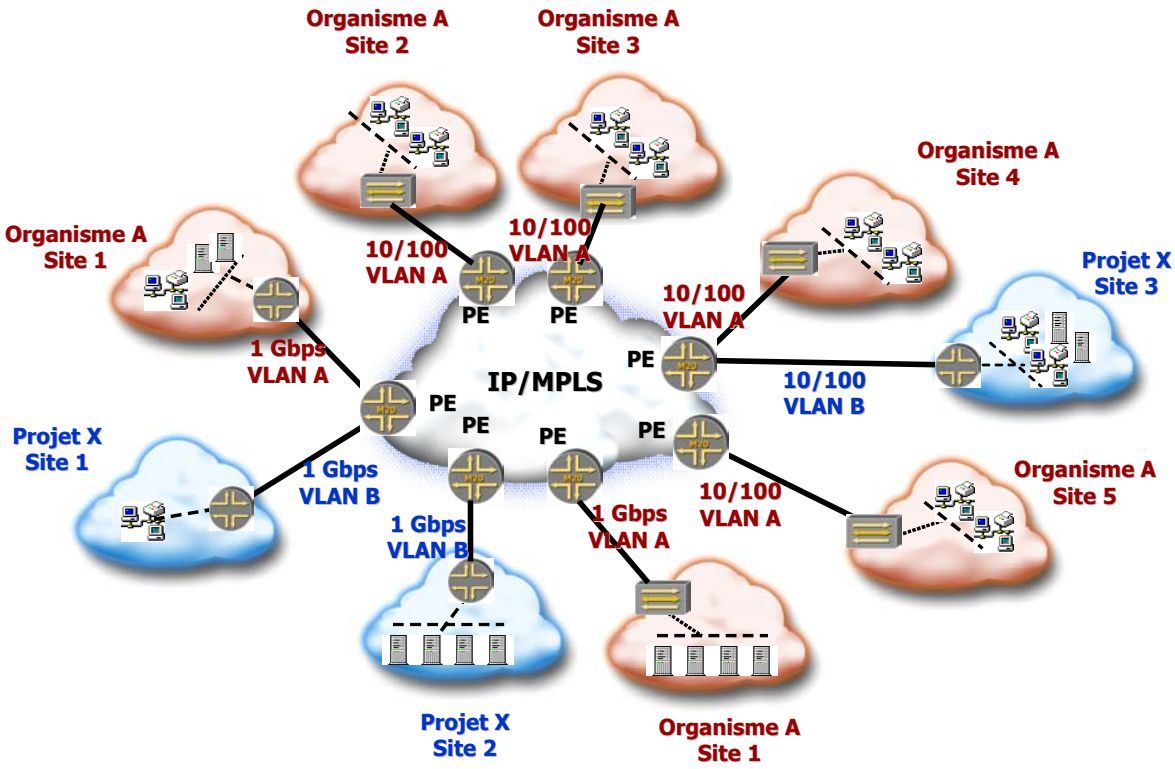

Figure 1 – *VPLS délivre un service Ethernet multipoint-à-multipoint qui peut s'étendre à plus d'une zone métropolitaine*

Le réseau de l'opérateur de services IP/MPLS est constitué d'un point de vue fonctionnel de routeurs de périphérie PE (Provider Edge) et de routeurs de cœur P (Provider). Dans la pratique, il est évident que des routeurs peuvent avoir à la fois le rôle d'un PE et d'un P, en particulier dans les réseaux de Recherche et Enseignement qui sont rarement hiérarchisés. Avant de détailler les spécificités de VPLS, il est important de noter que chaque PE a un tunnel établi vers chacun des autres PE. Ce tunnel peut être de type MPLS, généré de manière dynamique avec LDP[2] ou RSVP-TE[3], ou de manière statique dans certains cas particuliers. Mais ces tunnels peuvent aussi être de type GRE ou IPSec, ce qui permet de ne pas imposer l'usage de MPLS pour la commutation dans le cœur du réseau. Le point fondamental est que dans cette architecture, toute l'intelligence (la connaissance des VPNs) se trouve uniquement en périphérie, et les routeurs de cœur n'ont aucune information sur les VPNs ou les routes IP extérieures au réseau de l'opérateur. Ce modèle assure une distribution de l'intelligence dans les PE (un PE n'aura connaissance d'un VPN que si il en a besoin, c'est à dire si il connecte directement des clients à ce VPN), et la commutation dans le cœur est assurée par les routeurs P traversés par des tunnels (MPLS ou GRE ou IPSec) reliant les routeurs PE. Ce modèle sert de base aux VPNs MPLS de niveau 2 ou 3, et donc aussi à VPLS que nous allons décrire dans la suite de cet article. Nous considérons comme point de départ que la configuration décrite ci-dessus est déployée.

Chaque routeur PE (Provide Edge) en périphérie du réseau IP/MPLS de l'opérateur intègre les fonctionnalités VPLS. Chaque domaine VPLS est constitué d'un certain nombre de PE (les PE connectant un ou plusieurs sites de ce domaine VPLS), chacun activant une instance VPLS pour ce domaine. Un maillage complet de Label Switched Paths (LSPs) doit être construit entre toutes les instances VPLS sur chacun des PE pour un domaine VPLS donné. En fonction de l'implémentation VPLS utilisée, lorsqu'un PE ou une instance VPLS est ajouté, l'effort nécessaire pour établir un maillage de LSP peut varier considérablement.

Une fois le maillage construit, l'instance VPLS d'un PE est prête à recevoir des trames Ethernet d'un site client, et peut commuter ces trames sur le LSP approprié en fonction de l'adresse MAC destination. Cela est possible car VPLS permet au routeur PE d'agir comme un commutateur Ethernet avec une table d'adresses MAC par instance VPLS. En d'autres termes, l'instance VPLS sur le routeur PE dispose d'une table MAC alimentée par apprentissage des adresses MAC sources lorsque les trames Ethernet entrent par des ports physiques ou logiques dédiés, exactement de la même façon qu'avec un commutateur Ethernet traditionnel.

Une fois que la trame Ethernet arrive sur un port d'entrée connectant le client, l'adresse MAC destination est recherchée dans la table MAC et la trame est transmise sans altération (si, bien sur, l'adresse MAC correspondante se trouve dans la table MAC) à l'intérieur du LSP qui va la délivrer au PE adéquat connectant le site distant visé. Si l'adresse MAC n'est pas dans la table d'adresses MAC, la trame Ethernet est répliquée et transmise à tous les ports logiques associés à cette instance VPLS (ports locaux ou LSPs vers les autres PE du domaine VPLS), excepté le port d'entrée par lequel la trame est arrivée. Il est aussi important de noter que les PEs traitant les paquets en sortie (c'est à dire venant du cœur du backbone et destinés à sortir vers les sites clients) ne transmettent jamais des trames Ethernet en retour vers le cœur pour éviter toute boucle Ethernet dans le réseau. Le PE de sortie inspecte l'adresse MAC source et l'ajoute à sa table MAC avec le PE d'entrée du réseau identifié par le label utilisé par la trame Ethernet. La trame Ethernet est ainsi délivrée au site client, puis arrive à destination. Les adresses MAC n'ayant pas été utilisées après un certain temps sont automatiquement éliminées de la table, exactement comme sur un commutateur Ethernet.
Du point de vue du client, le service apporté par l'opérateur correspond à l'équivalent d'une connexion de chaque site sur un commutateur Ethernet dédié, mais sans limitation de distance entre les sites.

[2] Label Distribution Protocol (Protocole de distribution des étiquettes MPLS)
[3] Reservation Protocol for Traffic Engineering

4 Implémentations VPLS

Aujourd'hui deux drafts principaux sont débattus au niveau de l'IETF et sont considérés par la majorité des constructeurs pour développer leur solution. Le premier est http://www.ietf.org/internet-drafts/draft-kompella-ppvpn-vpls-02.txt, que nous appellerons « draft Kompella », proposé par Kireeti Kompella. Le second est http://www.ietf.org/internet-drafts/draft-ietf-l2vpn-vpls-ldp-00.txt, que nous appellerons « draft Lasserre-VKompella », proposé par Marc Lasserre et Vach Kompella. Chacun de ces deux modèles peut être décrit par deux caractéristiques fondamentales :

- Découverte automatique (Auto-Discovery) : quelle méthode est utilisée par chacun des routeurs de périphérie (PE) pour participer à un domaine VPLS et pour trouver les autres PEs partenaires ?
- Signalisation : quel protocole est utilisé pour établir des tunnels MPLS et distribuer des labels (étiquettes) entre les PEs pour assurer le démultiplexage ?

Modèle d'implémentation de VPLS	Découverte Automatique	Signalisation
Draft Kompella VPLS	BGP	BGP
Draft Lasserre-VKompella VPLS	Non défini	LDP

4.1 Découverte Automatique

La découverte automatique est une composante essentielle pour aider l'opérateur à déployer un service à grande échelle tout en minimisant les coûts opérationnels, car ce mécanisme permet la création automatique d'un maillage complet de LSP pour un domaine VPLS donné[4].

Afin de bien comprendre le mécanisme de découverte automatique, considérons qu'un nouveau routeur PE est ajouté en périphérie du réseau de l'opérateur, et que le nécessaire ait été fait pour assurer sa connectivité avec les autres PEs par des tunnels (MPLS ou GRE ou IPSec). Selon l'approche BGP proposée dans le draft Kompella, une simple session BGP (plus précisément MP-iBGP) est établie entre le PE et le Route Reflector[5] (distributeur de routes BGP). Ensuite le nouveau PE va se joindre à un domaine VPLS dés lors que l'instance VPLS est configurée sur ce routeur PE, et que un ou plusieurs ports connectant le client sur ce PE est associé à cette instance VPLS. Chaque domaine VPLS est identifié par une Route Target (attribut de la communauté étendue de BGP) particulière, qui est spécifiée dans la configuration de l'instance VPLS. Une fois cette opération réalisée, le PE informe via le Route Reflector de son appartenance à ce domaine VPLS aux autres PEs faisant partie de ce domaine VPLS. En effet, l'annonce BGP transporte la Route Target[6] (« cible de la route») de la communauté BGP étendue) configurée pour ce domaine VPLS, et c'est précisément cette communauté qui identifie l'annonce d'appartenance de nouveau PE au domaine VPLS. Par la suite tous les PEs ont pris connaissance du nouveau PE, et les PEs membres de ce domaine VPLS ont toutes les informations nécessaires pour établir des LSPs avec ce nouveau PE de manière automatique.

Etant donné que le draft Lasserre-VKompella ne spécifie pas de méthode de découverte automatique, l'opérateur de service doit informer le nouveau PE de manière explicite quels PEs font partie du même domaine VPLS. Pour chaque instance VPLS présente dans un PE, l'opérateur doit configurer ce PE avec les adresses de tous les PEs faisant partie du même domaine VPLS. Cette information peut être stockée de plusieurs manière : dans une base de données LDAP, un système de configuration ou un simple cahier/fichier. Cependant toutes ces options requièrent une intense activité opérationnelle et sont soumises à d'éventuelles erreurs humaines.
Diverses options ont été discutées pour ajouter une découverte automatique au draft Lasserre-VKompella :

[4] En rappel, il ne s'agit pas ici de la création de tunnels pour la commutation (ou routage) de l'ensemble des paquets entre les PE (cela étant supposé être déjà établi), mais de la création de tunnels de second niveau, spécifiques à un domaine VPLS donné, et permettant aux PEs de gérer le plan de données (commutation) pour les sites connectés à ce domaine VPLS.
[5] Le Route Reflector a pour fonction essentielle de distribuer les routes des PEs de manière centralisée (par un routeur choisi) et ainsi augmenter le facteur d'échelle d'une architecture iBGP (internal BGP). Pour assurer une redondance, un PE peut établir des sessions BGP avec d'autres Route Reflectors. Bien sur, dans le cas ou l'opérateur n'utilise pas de Route Reflector, nous aurons un maillage complet de sessions BGP entre les PEs, tels que défini par les règles d'ingénierie du protocole BGP.
[6] La Route Target permet d'ajouter à l'annonce BGP une « couleur » indiquant l'appartenance du PE à un domaine VPLS donné. Les PEs qui reçoivent l'annonce ont une Routing Policy (règle de routage) leur indiquant d'importer dans l'instance VPLS (appartenant au domaine) les routes associées à cette Route Target.

- Utiliser le protocole BGP[7] : cette option implique l'usage de deux protocoles différents pour gérer VPLS, l'un pour la découverte automatique (BGP), l'autre pour la signalisation (LDP), ce qui nous amènerait à plus de complexité, plus de maintien d'états dans le réseau, et plus d'interactions inter-protocoles.
- Modifier LDP pour réaliser la découverte automatique : cette solution impliquerait d'ajouter à LDP les mécanismes équivalents à BGP, à savoir les communautés et Route Target, ce qui pose la question de la justification de développer des mécanismes qui existent déjà dans le protocole BGP.
- Développer une solution basée sur DNS ou une autre base de données centralisée éventuellement propriétaire : ce type de solution centralisée devra être attentivement analysée, en particulier concernant sa capacité d'usage à grande échelle.

Il est intéressant de noter que c'est le même problème de découverte automatique qui a été rencontré plusieurs fois par le passé, et que la réponse dans la plupart des cas a finalement pointé vers BGP. Par exemple cela concernait les discussions pour les routeurs virtuels et les VPNs de niveau 3 définis dans la RFC 2547bis. Une raison souvent évoquée contre BGP et que cette solution est « trop complexe ». La réalité est que BGP est aussi simple à utiliser par un opérateur de service qu'un autre protocole comme OSPF, mais peut être plus complexe à implémenter par les équipementiers, ce qui peut générer une certaine motivation anti-BGP.

4.2 Signalisation

Le modèle VPLS supporté par le draft Kompella préconise l'usage de BGP pour la signalisation, c'est à dire la distribution des labels (étiquettes) pour l'établissement des LSPs entre les PEs pour un domaine VPLS particulier. Par contre le modèle défini dans le draft Lasserre-VKompella est basé sur l'usage de LDP comme mécanisme de signalisation. Pour chacun des modèles, le principe consiste a échanger des labels pour établir les LSPs entre les PEs pour un domaine VPLS donné.
De nombreuses discussions ont lieu sur les avantages et inconvénients de chaque solution.

Concernant l'usage de BGP comme protocole de signalisation, les arguments contre cette solution sont généralement :
- BGP est trop complexe à utiliser,
- BGP nécessite une pré-allocation de blocs de labels.

On peut s'interroger sur le poids de cet argumentaire : d'une part, de nombreux opérateurs (principalement des opérateurs commerciaux) ont aujourd'hui deployé des VPNs de niveau 3 selon la RFC 2547 (et son extension draft 2547bis) basée sur BGP. D'autre part, la prédéfinition des blocs n'a pas d'impact direct sur les ressources jusqu'à ce que les labels eux-mêmes soient réellement alloués et configurés. Et puis le principe de pré-allocation n'est pas obligatoire, c'est à dire qu'il est toujours possible de ne pas pré-allouer de bloc, et dans ce cas cela fonctionne comme avec LDP, c'est à dire manuellement.

D'un autre coté, il y a un certain nombre de limitations avec la signalisation LDP.
Premièrement, étant donné que le draft Laserre VKompella ne définit pas de mécanisme de découverte automatique, il est nécessaire à chaque fois qu'un PE joint un domaine VPLS de manuellement (action de l'opérateur) rechercher les autres PEs faisant partie du même domaine VPLS. Une fois cette information obtenue, il faut construire un maillage complet de sessions LDP entre ce PE et tous les autres PEs faisant partie de ce domaine VPLS. Ce grand nombre de LSPs (maillage à facteur exponentiel) est nécessaire car LDP n'a pas l'avantage de pouvoir bénéficier d'une architecture comme BGP et ses Route Reflectors. Pour les opérateurs offrant ce type de service à peu de clients ayant peu de sites, la charge générée par LDP n'est probablement pas notable. Cependant cette charge augmente de manière très significative avec la croissance du service.

[7] en plus de LDP qui est la solution définie dans le draft Lasserre-VKompella pour la signalisation, voir chapitre 4.2

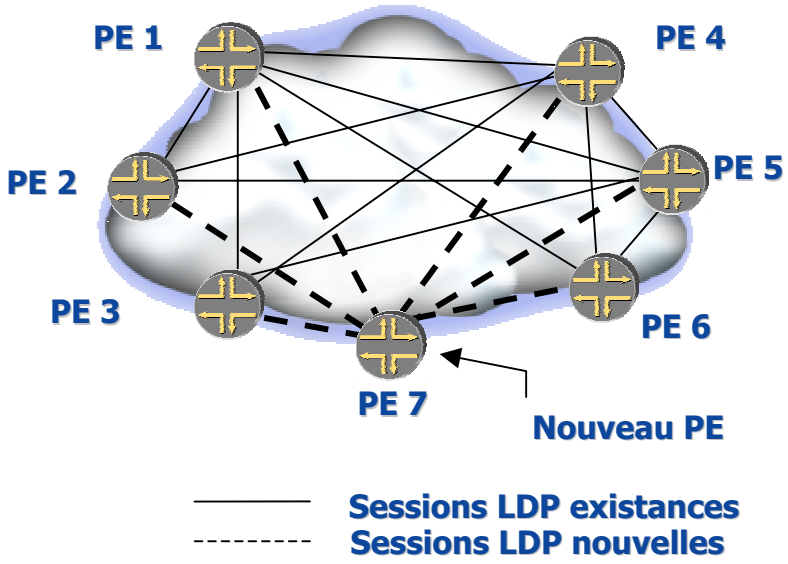

Figure 2 – *l'utilisation de LDP requiert un maillage complet de sessions LDP, qui doivent être établies chaque fois que l'on ajoute un nouveau PE.*

Deuxièmement, cet enjeu opérationnel consistant à gérer un nombre de sessions LDP d'ordre fonction de N puissance 2 devient encore plus significatif si l'opérateur prend la décision d'authentifier les sessions de signalisation LDP avec MD5 pour renforcer la sécurité de son réseau. Car dans un cas de maillage complet de sessions LDP, les clefs MD5 doivent être configurées à chaque extrémité de toutes les sessions LDP.
Troisièmement, étant donné le risque lié à la transmission éventuelle des informations à tous les routeurs en cas de mauvaise configuration du protocole, il est très utile de pouvoir rapidement détecter des erreurs et l'origine de ces erreurs. Le problème avec LDP est qu'une connexion PE1-PE2 pour un domaine VPLS donné peut avoir un différent numéro ID que la connexion PE1-PE3 pour ce même domaine. De plus, ce numéro ID n'est pas structuré et donc rend assez difficile toute identification.
Quatrièmement, dans le cas où un domaine VPLS doit être étendu à plusieurs Autonomous System (Système Autonome ou AS)[8], la signification globale des 32 bits utilisés par le VCID de la signalisation LDP requiert une coordination manuelle intensive par les équipes opérationnelles des différents AS. En d'autres termes, si un domaine VPLS s'étend sur 3 ASs, les trois opérateurs devront utiliser le même VCID LDP pour ce VPLS. Enfin, si le draft Lasserre-VKompella préconise un jour l'utilisation de BGP pour la découverte automatique, alors cela nécessitera une synchronisation entre BGP et LDP.

Le draft Kompella qui base la signalisation sur le protocole BGP ne souffre pas des limitations décrites ci-dessus. En effet, lorsqu'un PE est ajouté, une seule session BGP entre lui et le Route Reflector est requise. Si la session nécessite d'être authentifiée avec MD5, alors seuls les 2 équipements à chaque extrémité de la session BGP requièrent une configuration des clefs MD5. Lorsqu'une nouvelle instance VPLS est configurée sur ce PE, il avertit alors sa disponibilité du service via le Route Reflector, qui redistribue l'information pour tous les PEs concernés. Ensuite, la signalisation BGP construit automatiquement le maillage de LSPs pour cette instance VPLS. Toutes les connexions vers un domaine VPLS donné sont identifiées par un ID unique (le Route Target), et qui dispose d'une structure offrant d'office un partitionnement hiérarchique. Ce dernier point rend tout diagnostic de dysfonctionnement beaucoup plus simple pour l'équipe opérationnelle. De plus, si le domaine VPLS doit s'étendre sur plusieurs Systèmes Autonomes (ASs), pouvant être chacun géré par un opérateur différent (ce qui est un cas très courant dans les réseaux de Recherche et Enseignement), l'utilisation d'une Route Target identifiant un VPLS particulier simplifie grandement les aspect opérationnels, car chaque AS peut assigner une Route Target particulière pour ce VPLS selon ses propres règles internes. Cela est possible car la Route Target (de la communauté étendue) intègre le numéro de Système Autonome et ce numéro est globalement unique car alloué par un organisme officiel (par exemple le RIPE en Europe).

[8] La problématique inter-domaine pour des services de bout en bout sera étudiée dans le chapitre 5

Enfin, l'utilisation de BGP a la fois pour la découverte automatique et la signalisation implique une synchronisation entre les actions de découverte et de signalisation, et les équipes techniques n'ont besoin de maîtriser et de gérer qu'un seul protocole. Ce protocole BGP est par ailleurs déjà utilisé pour la découverte automatique et la signalisation des VPN de niveau 3 de type 2547bis, la solution de service VPN IP la plus déployée au monde à ce jour.

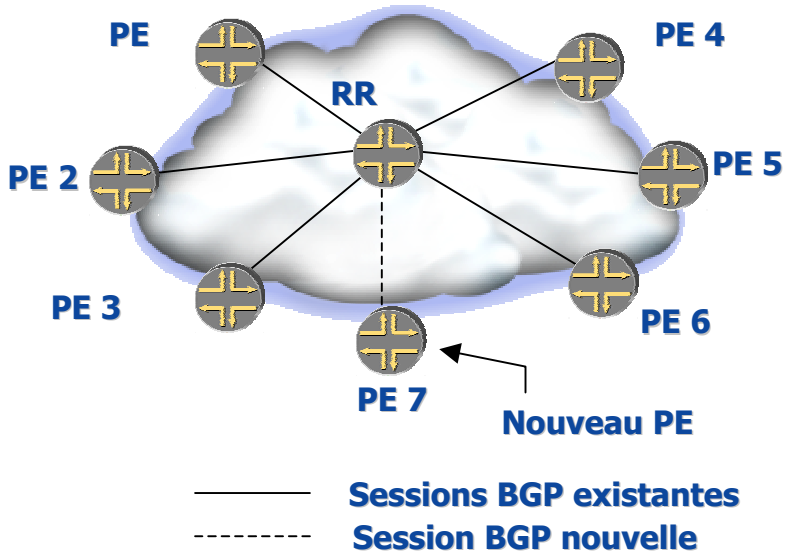

Figure 3 – *Une approche simplifiée ou les sessions BGP doivent être configurées uniquement entre le nouveau PE et le Route Reflector (ou les RR en cas de redondance), de manière identique au déploiement d'un service VPN de niveau 3 à la 2547bis. De plus, le nombre de domaines VPLS n'impacte pas le nombre de sessions BGP.*

5 Le bout en bout avec l'inter-domaine

Les réseaux de Recherche et Education ont pour vocation de pouvoir délivrer des services réseaux servant la communauté distribuée sur de nombreux domaines IP distincts, que ce soit au niveau métropolitain, régional, national ou Européen. Quel que soit le projet ou la communauté considérée, il y a une probabilité non négligeable que des sites (ou VLANs) soient connectés via plusieurs domaines distincts. Nous entendons par domaine IP un Système Autonome AS (au sens BGP), qui peut correspondre directement à un domaine administratif (par exemple un réseau métropolitain). Il se peut aussi qu'un réseau administratif contienne plusieurs ASs (pour les gros réseaux).

L'approche du draft Kompella étant basé sur BGP, il hérite toutes les capacités d'extension des services VPN au delà d'un AS unique. En effet les solutions carrier's carriers et Inter-AS/Inter-provider sont décrites dans le draft 2547bis (extensions de la RFC 2547). Nous présentons ici l'option C du draft, qui est la plus adaptée pour déployer des services VPLS à grande échelle. Il est important de noter que la description technique qui suit n'est pas spécifique au service VPLS. En effet, ces mécanismes s'appliquent de manière identique pour les services de VPN IP, Layer 2 (draft Kompella Layer 2 VPN) et VPLS (draft Kompella VPLS).
Prenons l'exemple d'un domaine VPLS que nous voulons étendre sur 3 domaines différents, tels que décrit dans la figure 4.

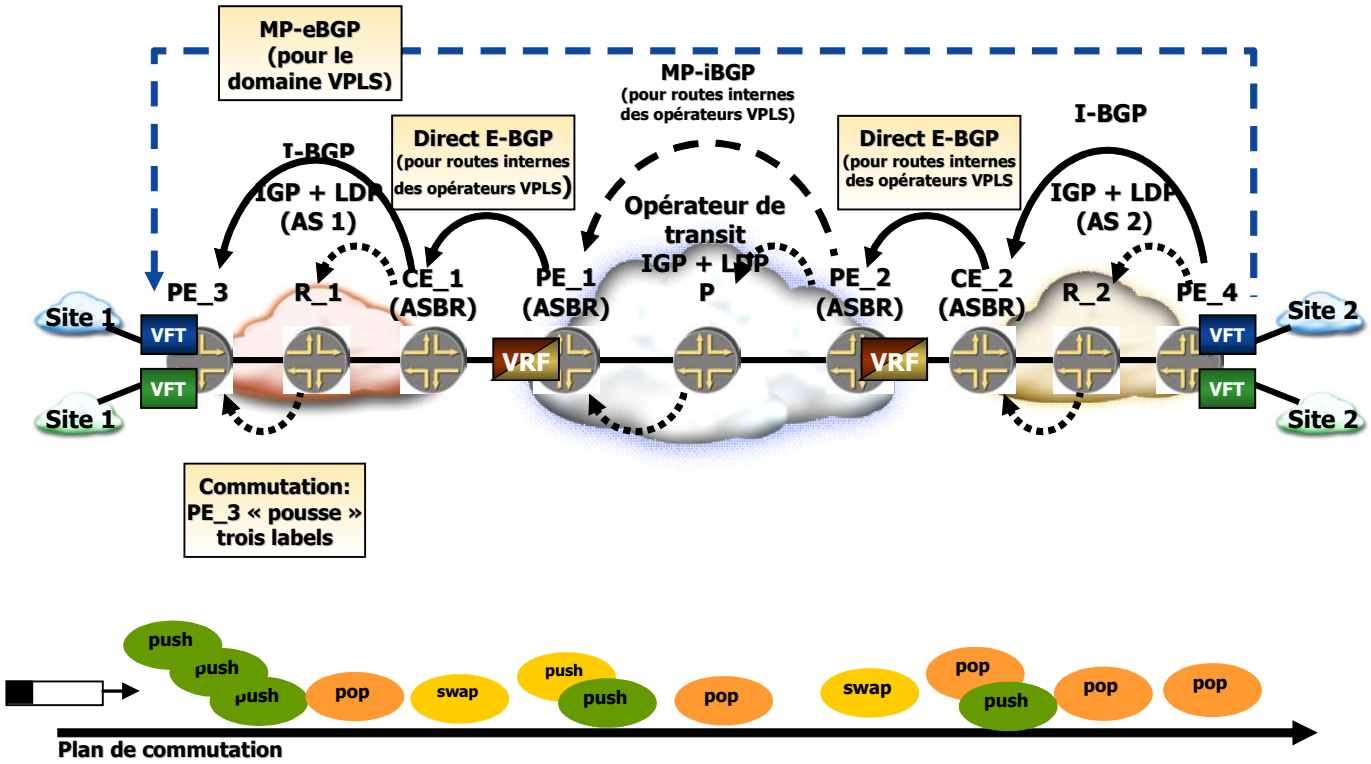

Figure 4 – *VPN inter-domaine : option C de la 2547Bis*

Pour étendre le modèle de l'AS unique à plusieurs ASs, il faut pouvoir établir des LSPs entre PEs qui ne sont pas dans le même AS, afin d'obtenir une commutation de bout en bout (sans routage) entre PE3 et PE4.
Une solution consiste pour les opérateurs de service VPLS (AS 1 et AS 2) à échanger leurs adresses internes (adresse des PEs supportant des instances VPLS) sous forme de VPN de niveau 3 par l'opérateur de transit. Ces routes sont maintenues dans des tables VRF[9] dédiées. Les sessions BGP entre AS1 (ou AS2) et l'opérateur de transit se font en E-BGP et chaque route est échangée avec un label ce qui permet de construire le plan de commutation. L'opérateur de transit gère ces routes sous forme de VPN de niveau 3 pour assurer une isolation avec le reste des routes (par exemple de l'Internet). Des sessions iBGP entre PEs au sein de chaque AS permettent une distribution locale (intra-AS) des routes externes, c'est à dire appartenant à l'AS distant. Ces échanges de routes avec iBGP se font aussi en y affectant des labels associés. Le résultat de cette opération apporte une connectivité MPLS (LSP) de bout en bout entre PE3 et PE4.

PE3 et PE4 contiennent les informations locales de chaque instance VPLS pour lesquelles des sites sont directement connectés. A chaque instance VPLS correspond une VPLS Forwarding Table (VFT). Il est maintenant possible d'établir une session Multi-hop MP-eBGP entre PE3 et PE4 selon le draft Kompella, et ainsi assurer la découverte automatique et la signalisation requises pour activer le domaine VPLS de bout en bout. Pour cette partie, la seule différence par rapport au chapitre 4 est l'activation d'une session Multi-hop MP-eBGP au lieu d'une session iBGP.

Au niveau du plan de commutation, PE3 recevant un paquet Ethernet par le site 1 du domaine VPLS bleu et à destination du site 2, va encapsuler la trame avec 3 labels MPLS. Le premier (de l'intérieur vers l'extérieur) correspond au label identifiant l'instance VPLS au niveau de PE4. Le deuxième label correspond au LSP permettant la commutation entre PE3 et PE4 au niveau de PE1, CE1, PE2 et CE2. Et le troisième label correspond au LSP permettant la commutation au sein de AS1 (routeur R1). Dans la figure 4, la fonction de « push » correspond à l'ajout d'un label MPLS, la fonction « pop » correspond à la suppression d'un label, et la fonction de « swap » correspond à l'échange de label[10].

[9] VPN Routing and Forwarding table
[10] le label n'ayant qu'une fonction locale, doit changer à chaque saut de commutation

Il est intéressant de noter que seul le premier label (label intérieur) est spécifique au service VPLS (plus précisément l'instance VPLS du PE de sortie). La solution peut en effet ne pas s'appuyer sur un service MPLS généralisé pour gérer le plan de commutation. Car il est possible d'établir un tunnel GRE (ou IPSEC) entre PE3 et PE4, et d'utiliser uniquement le label en charge du service VPLS. La solution choisie dépendra de nombreux facteurs et sera fonction notamment de l'échelle de déploiement recherchée.

Enfin l'exemple ci-dessus est basé sur des connexions BGP directement entre PEs pour bien comprendre les mécanismes mis en jeu. Cependant, pour des besoins de déploiement à grande échelle, le modèle s'appuiera fortement sur l'usage de Route Reflector, ce qui limite fortement les charges opérationnelles.
Dans ce cas, il n'y a qu'une seule session BGP entre les deux Route Reflectors (AS1 et AS2). L'AS intermédiaire n'a bien sûr pas connaissance du service VPLS (que ce soit l'état MLPS par domaine VPLS ou les tables d'adresses MAC par instance VPLS). La signalisation réalisée entre les deux Route Reflectors est équivalente (mais plus efficace) qu'un maillage complet de sessions BGP entre les PEs de l'AS1 et l'AS2. Lorsque le PE 3 envoie du trafic en mode « broadcast » (c'est à dire avant d'avoir appris l'adresse MAC et sa réelle destination), tous les PEs de l'AS1 et l'AS2 faisant partie du domaine VPLS concerné reçoivent le paquet.

La solution proposée dans la Figure 4 est basée sur un choix d'ingénierie, mais d'autres solutions peuvent être étudiées. Par exemple :

- L'opérateur de transit pourrait proposer un service IP/MPLS mais sans VPN de niveau 3. Dans ce cas les routes internes de chaque opérateur de service VPLS seront échangées (avec des labels associés pour construire les LSPs) par la table de routage globale de l'opérateur de transit. Cette solution n'offre pas de différence au niveau du plan de commutation mais a des incidences sur le plan de contrôle. En effet, elle sous-entend que les opérateurs de service VPLS acceptent de voir partager leurs adresses de PE internes dans une table de routage globale de l'opérateur de transit, éventuellement avec des règles de routage (policy routing) pour ne pas propager ces routes à d'autres domaines.
- L'opérateur de transit pourrait proposer un service de connexion point à point de niveau 2 (par exemple Layer 2 VPN) reliant CE1 et CE2. Dans ce cas, les deux opérateurs de service VPLS se retrouvent « directement » connectés et peuvent échanger avec eBGP leurs routes internes (avec labels associés) sans impliquer l'opérateur de transit au niveau 3.

Il apparaît à travers ces exemples que plusieurs solutions (en réalité combinaisons de plusieurs services basés sur les mêmes principes) sont envisageables.

L'exemple de la Figure 5 permet d'étendre le modèle de la Figure 4 à une situation plus complexe où des universités sont connectées via des réseaux métropolitains ou régionaux, nationaux et le backbone paneuropéen.

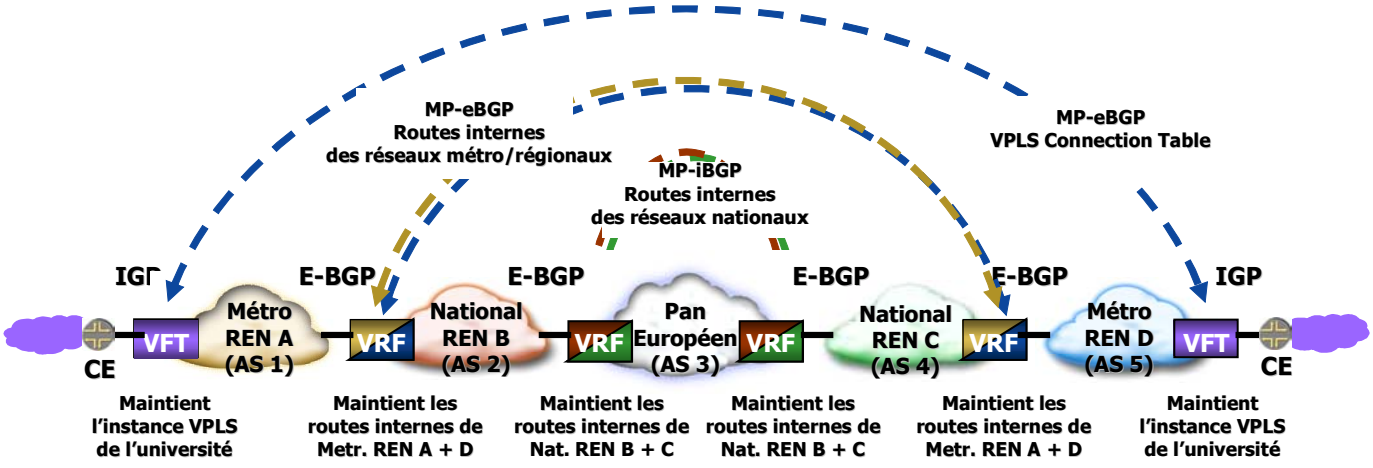

Figure 5 – *Exemple d'opérations Inter-domaine récursives*

Les mêmes mécanismes précédemment décrits peuvent s'appliquer dans ce cas de figure de manière récursive, avec un usage préconisé de Route Reflectors. Le sigle REN correspond à Research & Education Networks (réseaux de Recherche et Enseignement).

Le draft Lasserre-VKompella ne traite pas la problématique inter-domaine et précise que cela sera étudié dans une version future.

6 Conclusion

VPLS apporte un service Ethernet multipoint-à-multipoint sur une infrastructure IP/MPLS au niveau métropolitain ou longue distance. Différentes approches sont étudiées dans le cadre de l'IETF et des implémentations opérationnelles sont aujourd'hui proposées par les équipementiers. Les services Ethernet sont largement utilisés dans le cadre des réseaux Recherche et Enseignement, en particulier dans les réseaux métropolitains. La technologie permet aujourd'hui d'étudier et envisager la délivrance de services Ethernet innovants par et pour la communauté Recherche et Enseignement, non seulement au niveau local, mais aussi de manière plus globale dans un cadre multi-domaines.

Il y a aujourd'hui peu de solutions VPLS disponibles.
Juniper Networks délivre une solution VPLS basée sur le draft Kompella depuis la version JUNOS 5.7[11] pour les routeurs de la série M et T. L'implémentation VPLS de Juniper Networks est aujourd'hui complète et destinée aux réseaux de production. Cela veut dire que la solution intègre les mécanismes de découverte automatique, l'échange des labels, l'apprentissage des adresses MAC et fonction d'inondation, et fonctionne aussi en environnement inter-domaine (inter-AS et inter-opérateur).
Ce dernier point mériterait une réflexion plus approfondie de la part des réseaux et institutions de Recherche & Enseignement. En effet, comme nous l'avons vu dans le chapitre 5, plusieurs options sont envisageables avec la technologie existante. Cette étude basée sur une bonne coordination des entités intéressées devrait à la fois étudier les interfaces techniques entre les domaines (définir le service de transit, préciser les informations échangées et la politique de routage) et les interfaces opérationnelles.
Des changements importants ont été réalisés avec BGP ces dernières années (permettant entre autres l'activation inter-domaine des services multicast et IPv6), et une nouvelle phase démarre pour prolonger les services de VPNs à plusieurs domaines, à la fois pour des VPNs de niveau 3, de niveau 2 point a point, et Ethernet multipoint-a-multipoint (VPLS). Comme pour le Multicast ou IPv6, cela demanderait une révision au niveau local ou global des interfaces d'interconnexion entre les réseaux de Recherche et Enseignement.

Références

[1] Kompella K., et al, "Virtual Private LAN Service", draft-kompella-ppvpn-vpls-02.txt, Mai 2003.
[2] Lasserre M., Kompella V., et al, "Virtual Private LAN Services over MPLS", draft-ietf-l2vpn-vpls-ldp-00.txt, Juin 2003.
[3] Rosen E., Rekhter Y., BGP/MPLS VPNs, RFC2547, Mars 1999
[4] Rosen E., Rekhter Y., et. al., "BGP/MPLS IP VPNs", draft-ietf-l3vpn-rfc2547bis-01.txt, Mai 2003.
[5] Capuano M., VPLS : Scalable Transparent LAN Services (White Paper Juniper Networks), 2003
[6] Semeria Chuck, RFC 2547bis: BGP/MPLS VPN. Hierarchical and Recursive Applications (White Paper Juniper Networks), 2001

[11] Tous les routeurs de la série M et T de Juniper Networks utilisent le même logiciel JUNOS (code binaire identique). Tous les paquets (IPv4, IPv6, MPLS, etc…) sont routés/commutés (mais aussi filtrés, comptés, limités, etc…) en hardware par des ASICs spécifiques.