Groupe de travail Réseau
Request for Comments : 4761
Catégorie : Sur la voie de la normalisation
Traduction Claude Brière de L'Isle

V. Kompella, éd., Alcatel-Lucent
Y. Rekhter, éd., Juniper Networks
janvier 2007

Service de LAN privé virtuel (VPLS) utilisant BGP pour l'auto découverte et la signalisation

Note de l'IESG
Le groupe de travail L2VPN a produit deux documents séparés, la RFC 4762 et le présent document, qui effectuent en fin de compte des fonctions similaires de manières différentes. Il faut savoir que chaque méthode est couramment appelée "VPLS" bien qu'elles soient distinctes et incompatibles entre elles.

Résumé
Le service de LAN privé virtuel (VPLS, *Virtual Private LAN Service*) aussi appelé service de LAN transparent et service de réseau commuté privé virtuel, est une offre utile de fournisseur de services. Le service offre un réseau privé virtuel (VPN, *Virtual Private Network*) de couche 2 ; cependant, dans le cas de VPLS, les consommateurs dans le VPN sont connectés par un LAN Ethernet multipoint, à la différence des VPN de couche 2 usuels, qui sont en point à point par nature.

Le présent document décrit les fonctions requises pour offrir des VPLS, un mécanisme pour signaler un VPLS, et les règles pour transmettre les trames VPLS à travers un réseau à commutation de paquets.

Table des matières

1. Introduction

Le service de LAN privé virtuel (VPLS, *Virtual Private LAN Service*), aussi appelé service de LAN transparent et service de réseau commuté privé virtuel, est une offre de service utile. Un LAN privé virtuel apparaît sous (presque) tous les aspects comme un LAN Ethernet aux consommateurs d'un fournisseur de services. Cependant, dans un VPLS, les consommateurs ne sont pas tous connectés à un seul LAN ; les consommateurs peuvent être dispersés sur une zone métropolitaine ou une grande zone. Par essence, un VPLS agglutine ensemble plusieurs LAN individuels à travers un réseau à commutation de paquets pour apparaître et fonctionner comme un seul LAN [RFC4664]. Ceci est accompli en incorporant les fonctions d'apprentissage de l'adresse MAC, d'arrosage, et de transmission dans le contexte de pseudo-filaires qui connectent ces LAN individuels à travers le réseau à commutation de paquets.

Le présent document détaille les fonctions nécessaires pour offrir le VPLS, et ensuite décrit un mécanisme pour l'auto découverte des points d'extrémité d'un VPLS ainsi que pour signaler un VPLS. Il décrit aussi comment les trames VPLS sont transportées sur des tunnels à travers un réseau à commutation de paquets. Les mécanismes d'auto découverte et de signalisation utilisent BGP comme protocole de plan de contrôle. Le présent document discute aussi brièvement les options de déploiement, en particulier, la notion de fonctions de découplage à travers les appareils.

D'autres approches incluent : [VPN-Tunnel], qui permet de construire un VPN de couche 2 avec Ethernet comme interconnexion ; et la [RFC4447], qui permet d'établir une connexion Ethernet à travers un réseau à commutation de paquets. Toutes deux offrent cependant des services Ethernet en point à point. Ce qui distingue VPLS des deux approches ci-dessus est que VPLS offre un service multipoints. Un mécanisme pour établir des pseudo-filaires pour VPLS utilisant le protocole de distribution d'étiquettes (LDP, *Label Distribution Protocol*) est défini dans la [RFC4762].

1.1 Domaine d'application de ce document

Le présent document a quatre parties majeures : définir un modèle fonctionnel de VPLS ; définir un plan de contrôle pour l'établissement d'un VPLS ; définir le plan des données pour VPLS (encapsulation et transmission des données) ; et définir les diverses options de déploiement.

Le modèle fonctionnel sous-jacent à VPLS est décrit dans la Section 2. Cela décrit le service offert, les composants du réseau qui interagissent pour fournir le service, et de façon générale, leurs interactions.

Le plan de contrôle décrit dans ce document utilise BGP multi protocoles [RFC4760] pour établir le service de VPLS, c'est-à-dire, pour l'auto découverte des membres VPLS et pour l'établissement et la suppression des pseudo-filaires qui constituent une certaine instance de VPLS. La Section 3 se concentre sur cela, et aussi décrit comment est établi un VPLS qui traverse des frontières de système autonome, ainsi que comment est traité le multi rattachements. Utiliser BGP comme plan de contrôle pour les VPN n'est pas nouveau (voir [VPN-Tunnel], [RFC4364], et [RFC5195]) : ce qui est décrit ici se fonde sur les mécanismes proposés dans la [RFC4364].

Le plan de transmission et les actions que doit effectuer un routeur participant côté fournisseur (PE, *Provider Edge*) pour offrir le service de VPLS sont décrits dans la Section 4.

Dans la Section 5, la notion d'opération "découplée" est définie, et l'interaction de PE découplés et non découplés est décrite. Le découplage permet un déploiement plus souple de VPLS.

1.2 Conventions utilisées dans ce document

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

2. Modèle fonctionnel

Il va être décrit en se référant à la figure suivante.

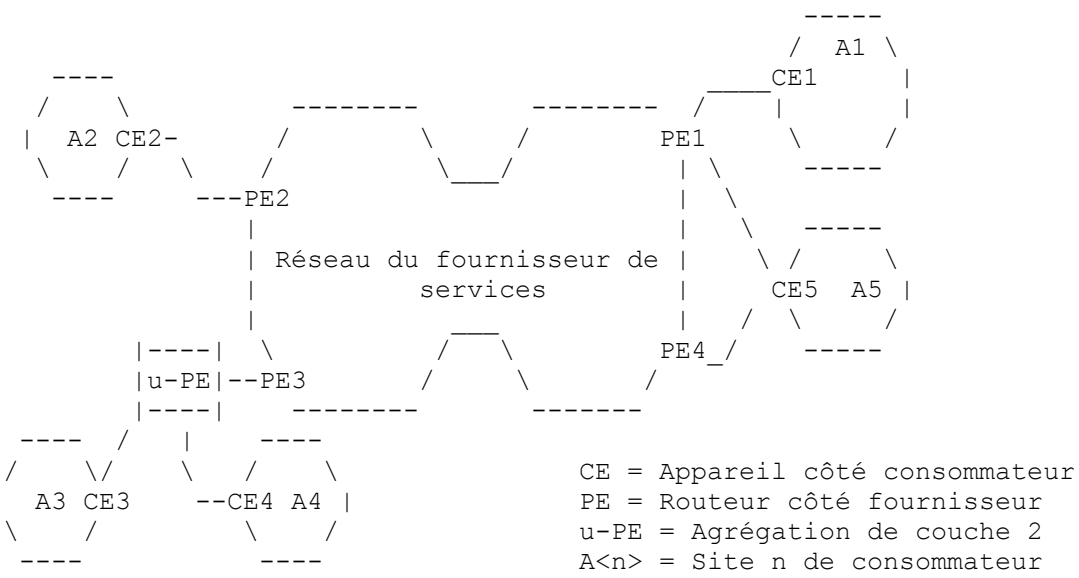

Figure 1 : Exemple de VPLS

2.1 Terminologie

Une terminologie similaire à celle de la [RFC4364] est utilisée : un réseau de fournisseur de service (SP, *Service Provider*) avec des routeurs seulement fournisseur (P, *Provider-only*) et côté fournisseur (PE, *Provider Edge*) , et des appareils de côté consommateur (CE, *Customer Edge*). Ici, cependant, il y a un concept supplémentaire, celui d'un appareil PE de couche 2, "u-PE", utilisé pour l'agrégation de couche 2. La notion de u-PE est plus décrite dans la Section 5. Les appareils PE et u-PE ont "connaissance de VPLS", ce qui signifie qu'ils savent qu'un service de VPLS est offert. Le terme de "VE" se réfère à un appareil côté VPLS, qui pourrait être un PE ou un u-PE.

À l'opposé, l'appareil CE (qui peut être possédé et utilisé par le SP ou par le consommateur) est sans connaissance de VPLS ; pour autant que le CE est concerné, il est connecté aux autres CE dans le VPLS via un réseau commuté de couche 2. Cela signifie qu'il ne devrait pas y avoir de changement à un appareil CE, ni au matériel, ni au logiciel, afin d'offrir le VPLS.

Un appareil CE peut être connecté à un PE ou un u-PE via des commutateurs de couche 2 qui sont sans connaissance de VPLS. Du point de vue de VPLS, de tels commutateurs de couche 2 sont invisibles, et donc on n'en parlera pas plus. Un u-PE peut de plus être connecté à un PE via des appareils de couche 2 et de couche 3 ; en en discute dans une autre section.

Le terme "démultiplexeur" se réfère à un identifiant dans un paquet de données qui identifie le VPLS auquel le paquet appartient ainsi que le PE d'entrée. Dans ce document, le démultiplexeur est une étiquette MPLS.

Le terme "VPLS" va se référer au service aussi bien qu'à une instanciation particulière du service (c'est-à-dire, un LAN émulé) ; cela devrait être clair d'après le contexte.

2.2 Hypothèses

Le réseau du fournisseur de services est un réseau à commutation de paquets. Les PE sont supposés être (logiquement) complètement maillés avec des tunnels sur lesquels les paquets qui appartiennent à un service (comme VPLS) sont encapsulés et transmis. Ces tunnels peuvent être des tunnels IP, comme des tunnels d'encapsulation générique d'acheminement (GRE, *Generic Routing Encapsulation*) ou des tunnels MPLS, établis par le protocole de réservation de ressources – ingénierie du trafic (RSVP-TE) ou LDP. Ces tunnels sont établis indépendamment des services offerts sur eux ; la signalisation et l'établissement de ces tunnels ne sont pas discutés dans le présent document.

L'"arrosage" et "l'apprentissage" d'adresse MAC (voir la Section 4) font partie intégrante de VPLS. Cependant, ces activités sont privées pour un appareil de SP, c'est-à-dire, dans le VPLS décrit ci-dessous, aucun appareil de SP ne demande à un autre appareil de SP d'arroser des paquets ou d'apprendre des adresses MAC en son nom.

Tous les PE qui participent à un VPLS sont supposés être pleinement maillés dans le plan des données, c'est-à-dire, il y a un pseudo-filaire bidirectionnel entre chaque paire de PE participant à ce VPLS, et donc chaque PE (d'entrée) peut envoyer un paquet VPLS au ou aux PE de sortie directement, sans avoir besoin d'un PE intermédiaire (voir le paragraphe 4.2.5.) Cela exige que les PE VPLS soient logiquement pleinement maillés dans le plan de contrôle afin qu'un PE puisse envoyer un message à un autre PE pour établir les pseudo-filaires nécessaires. Voir au paragraphe 3.6 une discussion sur des solutions de remplacement pour réaliser un maillage logique complet dans le plan de contrôle.

2.3 Interactions

VPLS est un "service de LAN" en ce que les appareils CE qui appartiennent à une certaine instance VPLS peuvent interagir à travers le réseau de SP comme si ils étaient connectés par un LAN. VPLS est "privé" en ce que les appareils CE qui appartiennent à des VPLS différents ne peuvent pas interagir. VPLS est "virtuel" en ce que plusieurs VPLS peuvent être offerts sur un réseau à commutation de paquets commun.

Les appareils PE interagissent pour "découvrir" tous les autres PE participant au même VPLS, et pour échanger des démultiplexeurs. Ces interactions sont dirigées par le contrôle, et non par les données.

Les u-PE interagissent avec les PE pour établir les connexions avec les PE distants ou les u-PE dans le même VPLS. cette interaction est dirigée par le contrôle.

Les appareils PE peuvent participer simultanément à des VPLS et des VPN IP [RFC4364]. Ces services sont indépendants, et les informations échangées pour chaque type de service sont gardées séparément car les informations d'accessibilité de couche réseau (NLRI, *Network Layer Reachability Information*) utilisées pour cet échange ont des identifiants de famille d'adresse (AFI, *Address Family Identifier*) et des identifiants de famille d'adresse suivants (SAFI, *Subsequent Address Family Identifier*) différents. Par conséquent, une mise en œuvre DOIT conserver une mémorisation d'acheminement séparée pour chaque service. Cependant, plusieurs services peuvent utiliser les mêmes tunnels sous-jacents; l'étiquette de VPLS ou de VPN est utilisée pour démultiplexer les paquets appartenant à des services différents.

3. Plan de contrôle

Il y a deux fonctions principales du plan de contrôle VPLS : l'auto découverte, et l'établissement/suppression des pseudo-filaires qui constituent le VPLS, souvent appelée signalisation. Les paragraphes 3.1 et 3.2 décrivent ces fonctions. Ces deux fonctions sont accomplies par une seule annonce de mise à jour BGP ; le paragraphe 3.3 décrit comment cela est fait en détaillant le fonctionnement du protocole BGP pour VPLS. Le paragraphe 3.4 décrit l'établissement des pseudo-filaires qui s'étendent sur les systèmes autonomes. Le paragraphe 3.5 décrit le traitement du multi rattachements.

3.1 Auto découverte

La découverte se réfère au processus par lequel on trouve tous les PE qui participent à une certaine instance de VPLS. Un PE peut être configuré avec les identités de tous les autres PE dans un certain VPLS ou il peut utiliser un protocole pour découvrir les autres PE. Ce dernier est appelé l'auto découverte.

La première approche est très consommatrice de configuration, en particulier parce qu'elle exige que les PE qui participent à un certain VPLS soient complètement maillés (c'est-à-dire, que chaque PE dans un VPLS donné établit des pseudo-filaires avec chaque autre PE de ce VPLS). De plus, quand la topologie d'un VPLS change (c'est-à-dire, quand un PE est ajouté, ou retiré du VPLS) la configuration de VPLS sur tous les PE dans ce VPLS doit être changée.

Dans l'approche de l'auto découverte, chaque PE "découvre" quels autres PE font partie d'un certain VPLS au moyen d'un protocole, dans le cas présent BGP. Cela permet à la configuration de chaque PE de consister seulement en l'identité de l'instance de VPLS établie sur ce PE, et non l'identité de chaque autre PE dans cette instance de VPLS -- elle est auto-découverte. De plus, quand la topologie d'un VPLS change, seule la configuration des PE affectés change ; les autres PE trouvent automatiquement le changement et s'adaptent.

3.1.1 Fonctions

Un PE qui participe à une certaine instance de VPLS V doit être capable de dire aux autres PE dans le VPLS V qu'il est aussi membre de V. Un PE doit aussi avoir un moyen de déclarer qu'il ne participe plus à un VPLS. Pour ce faire, le PE doit avoir un moyen d'identifier un VPLS et un moyen de communiquer avec tous les autres PE.

Les appareils u-PE ont aussi besoin de savoir ce qui constitue un certain VPLS ; cependant, ils n'ont pas besoin du même niveau de détail. Le ou les PE auxquels un u-PE est connecté donnent au u-PE une abstraction du VPLS ; ceci est décrit à la Section 5.

3.1.2 Spécification du protocole

Le mécanisme spécifique de l'auto découverte décrit ici se fonde sur [VPN-Tunnel] et la [RFC4364] ; il utilise les communautés étendues de BGP [RFC4360] pour identifier les membres d'un VPLS, en particulier, la communauté Cible de chemin (RT, *Route Target*) dont le format est décrit dans la [RFC4360]. La sémantique de l'utilisation des cibles de chemin est décrite dans la [RFC4364] ; son utilisation dans VPLS est identique.

Comme on a supposé que les VPLS sont pleinement maillés, une seule cible de chemin suffit pour un VPLS V donné, et a pour effet que la RT est l'identifiant du VPLS V.

Un PE annonce (normalement via I-BGP) qu'il appartient au VPLS V en annotant ses NLRI pour V (voir le paragraphe suivant) avec la RT, et agit sur elle en acceptant les NLRI provenant des autres PE qui ont la RT. Un PE annonce qu'il ne participe plus à V en retirant toutes les NLRI qu'il avait annoncées avec la RT.

3.2 Signalisation

Une fois la découverte effectuée, chaque paire de PE dans un VPLS doit être capable d'établir (et supprimer) les pseudo-filaires avec chaque autre, c'est-à-dire, échanger (et supprimer) les démultiplexeurs. Ce procès est appelé signalisation. La signalisation est aussi utilisée pour transmettre certaines caractéristiques des pseudo-filaires qu'un PE établit pour un certain VPLS.

On se rappelle qu'un démultiplexeur est utilisé pour distinguer plusieurs flux de trafic différents portés sur un tunnel, chaque flux représentant éventuellement un service différent. Dans le cas de VPLS, le démultiplexeur ne dit pas seulement à quel VPLS spécifique appartient un paquet, mais aussi identifie le PE d'entrée. La première information est utilisée pour transmettre le paquet ; la dernière information est utilisée pour apprendre les adresses MAC. Le démultiplexeur décrit ici est une étiquette MPLS. On notera cependant que les tunnels de PE à PE n'ont pas besoin d'être des tunnels MPLS.

Utiliser un message Update BGP distinct pour envoyer un démultiplexeur à chaque PE distant exigerait que le PE d'origine envoie N messages pour N PE distants. La solution décrite dans ce document permet à un PE d'envoyer un seul (commun) message Update qui contient les démultiplexeurs pour tous les PE distants, au lieu de N messages individuels. Faire ainsi réduit la charge du plan de contrôle à la fois chez le PE d'origine et chez les réflecteurs de chemin BGP qui peuvent être impliqués dans la distribution de ce Update aux autres PE.

3.2.1 Blocs d'étiquettes

Pour réaliser cela, on introduit la notion de "blocs d'étiquettes". Un bloc d'étiquettes, défini par une base d'étiquette LB *(label base)* et une taille de bloc VE (VBS, *VE block size*) est un ensemble contigu d'étiquettes {LB, LB+1, ..., LB+VBS-1}. Voici comment les blocs d'étiquettes fonctionnent. Tous les PE au sein d'un VPLS donné reçoivent des identifiants VE uniques au titre de leur configuration. Un PE X qui souhaite envoyer une mise à jour de VPLS envoie les mêmes informations de bloc d'étiquettes à tous les autres PE. Chaque PE receveur déduit l'étiquette destinée au PE X en ajoutant son identifiant VE (unique) à la base d'étiquette. De cette manière, chaque PE receveur obtient un démultiplexeur unique pour le PE X pour ce VPLS.

Cette simple notion est améliorée par le concept d'un décalage de bloc VE (VBO, *VE block offset*). Un bloc d'étiquettes défini par <LB, VBO, VBS> est l'ensemble {LB+VBO, LB+VBO+1, ..., LB+VBO+VBS-1}. Donc, au lieu d'un seul grand bloc d'étiquettes pour couvrir tous les identifiants de VE dans un VPLS, on peut avoir plusieurs blocs d'étiquettes, chacun avec une base d'étiquette différente. Cela rend la gestion de bloc d'étiquettes plus facile, et aussi permet au PE X d'approvisionner en douceur un PE qui se joint à un VPLS avec un identifiant de VE qui n'est pas couvert par l'ensemble de blocs d'étiquettes que le PE X a déjà annoncé.

Quand un PE démarre, ou est configuré avec une nouvelle instance de VPLS, le processus BGP peut souhaiter attendre de recevoir plusieurs annonces pour cette instance de VPLS provenant d'autres PE pour améliorer l'efficacité de l'allocation de blocs d'étiquettes.

3.2.2 NLRI de VPLS BGP

Les NLRI de VPLS BGP décrites ci-dessous, avec de nouvelles AFI et SAFI (voir la [RFC4760]) sont utilisées pour échanger les membres et les démultiplexeurs du VPLS.

Les NLRI de VPLS BGP ont les éléments d'information suivants : un identifiant de VE, un décalage de bloc VE, une taille de bloc VE, et une base d'étiquette. Le format des NLRI de VPLS est donné ci-dessous. L'AFI est l'AFI L2VPN (25), et le SAFI est le SAFI VPLS (65). Le champ Longueur est en octets.

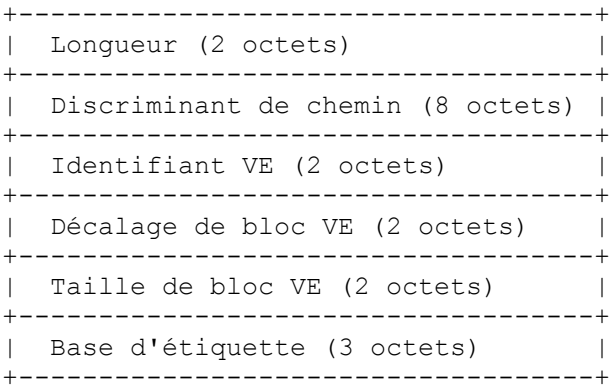

Figure 2 : NLRI BGP pour informations de VPLS

Un PE qui participe à un VPLS doit avoir au moins un identifiant de VE. Si le PE est le VE, il a normalement un identifiant de VE. Si le PE est connecté à plusieurs u-PE, il a un identifiant de VE distinct pour chaque u-PE. Il peut de plus avoir un identifiant de VE pour lui-même, si il agit lui-même comme VE pour ce VPLS. Dans ce qui suit, on appellera le PE qui annonce les NLRI de VPLS le PE-a, et on supposera que le PE-a possède l'identifiant de VE V (appartenant au PE-a lui-même ou à un u-PE connecté au PE-a).

Les identifiants de VE sont normalement alloués par l'administrateur du réseau. Leur portée est locale pour un VPLS. Un identifiant de VE donné devrait appartenir à un seul PE, sauf si un CE est multi rattachements (voir au paragraphe 3.5).

Un bloc d'étiquettes est un ensemble d'étiquettes de démultiplexeur utilisé pour atteindre un certain identifiant de VE. Des NLRI BGP de VPLS avec l'identifiant de VE V, le décalage de bloc de VE VBO, la taille de bloc de VE VBS, et la base d'étiquette LB communiquent ce qui suit à leurs homologues :

bloc d'étiquettes pour V : étiquettes de LB à (LB + VBS - 1) et
ensemble de VE distant pour V : de VBO à (VBO + VBS - 1).

Il y a une correspondance biunivoque entre l'ensemble de VE distant et le bloc d'étiquettes : l'identifiant de VE (VBO + n) correspond à l'étiquette (LB + n).

3.2.3 Établissement et suppression de PW

Supposons que PE-a fait partie du VPLS foo et fait une annonce avec l'identifiant de VE V, le décalage de bloc de VE VBO, la taille de bloc de VE VBS, et la base d'étiquette LB. Si PE-b fait aussi partie du VPLS foo et a l'identifiant de VE W, le PE-b fait ce qui suit :

1. il vérifie si W fait partie de l'ensemble de VE distants du PE-a : si VBO ≤ W < VBO + VBS, alors W fait partie de l'ensemble de VE distants de PE-a. Sinon, PE-b ignore ce message, et saute le reste de cette procédure.

2. il établit un PW pour PE-a : l'étiquette de démultiplexeur pour envoyer le trafic du PE-b au PE-a est calculée par (LB + W - VBO).

3. il vérifie si V fait partie d'un "ensemble de VE distants" que le PE-b a annoncé, c'est-à-dire, PE-b vérifie si V appartient à un ensemble de VE distants que le PE-b a annoncé, disons avec le décalage de bloc de VE VBO', la taille de bloc de VE VBS', et la base d'étiquette LB'. Sinon, le PE-b DOIT faire une nouvelle annonce comme décrit au paragraphe 3.3.

4. Il établit un PW à partir de PE-a : l'étiquette de démultiplexeur sur lequel PE-b devrait attendre du trafic provenant de PE-a est calculée comme : (LB' + V - VBO').

Si Y retire des NLRI pour V qu'utilisait X, alors X DOIT supprimer son extrémité de pseudo-filaire entre X et Y.

3.2.4 Signalisation des capacités de PE

L'attribut étendu suivant, "Communauté étendue d'informations de couche 2", est utilisé pour signaler des informations de contrôle sur les pseudo-filaires à établir pour un certain VPLS. La valeur de communauté étendue est à allouer par l'IANA (la valeur utilisée actuellement est 0x800A). Ces informations incluent le type Encaps (type d'encapsulation sur les pseudo-filaires) les fanions de contrôle (informations de contrôle concernant les pseudo-filaires) et l'unité maximum de transmission (MTU) à utiliser sur les pseudo-filaires.

Le type Encaps pour VPLS est 19.

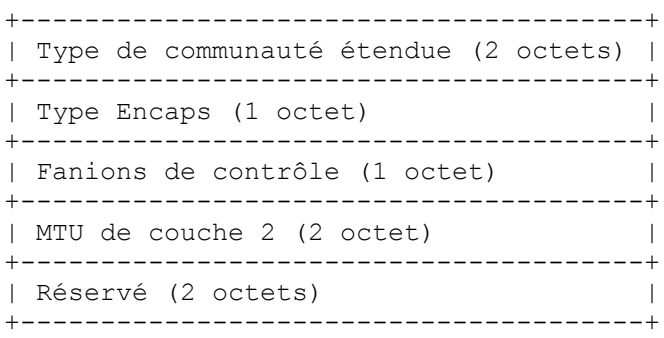

Figure 3 : Communauté étendue d'informations de couche 2

Figure 4 : octet des fanions de contrôle

Par référence à la Figure 4, les bits suivants de fanions de contrôle sont définis ; les bits restants, désignés par MBZ, DOIVENT être réglés à zéro à l'envoi et DOIVENT être ignorés à la réception de cette communauté.

Nom	Signification
C	Un mot de contrôle [RFC4448] DOIT ou NE DOIT PAS être présent à l'envoi de paquets VPLS à ce PE, selon que C est 1 ou 0, respectivement.
S	La livraison en séquence des trames DOIT ou NE DOIT PAS être utilisée à l'envoi de paquets VPLS à ce PE, selon que S est 1 ou 0, respectivement.

3.3 Fonctionnement de VPLS BGP

Pour créer un nouveau VPLS, disons le VPLS foo, un administrateur de réseau doit prendre un RT pour VPLS foo, disons RT-foo. Il va être utilisé par tous les PE qui desservent le VPLS foo. Pour configurer un certain PE, disons le PE-a, comme faisant partie du VPLS foo, l'administrateur de réseau a seulement à choisir un identifiant de VE V pour le PE-a. (Si le PE-a est connecté aux u-PE, le PE-a peut être configuré avec plus d'un identifiant de VE ; dans ce cas, ce qui suit est fait pour chaque identifiant de VE). Le PE peut aussi être configuré avec un différenciateur de chemin (RD, *Route Distinguisher*) sinon, il génère un RD unique pour le VPLS foo. Disons que le RD est RD-foo-a. Le PE-a génère alors un bloc initial d'étiquettes et un VE distant établi pour V, défini par le décalage de bloc de VE (VBO, *VE Block Offset*), la taille de bloc de VE (VBS, *VE Block Size*) et la base d'étiquette LB. Ce peut être vide.

Le PE-a crée alors des NLRI BGP de VPLS avec le RD RD-foo-a, l'identifiant de VE V, le décalage de bloc de VE VBO, la taille de bloc de VE VBS et la base d'étiquette LB. À cela, il attache une communauté étendue d'informations de couche 2 et un RT, RT-foo. Il règle le prochain bond BGP pour ces NLRI comme lui-même, et annonce ces NLRI à ses homologues. Le protocole de couche réseau associé à l'adresse réseau du prochain bond pour la combinaison <AFI=L2VPN AFI, SAFI=VPLS SAFI> est IP ; cette association est exigée par la Section 5 de la [RFC4760]. Si la valeur de la longueur du champ Prochain bond est 4, alors le prochain bond contient une adresse IPv4. Si cette valeur est 16, le prochain bond contient une adresse IPv6.

Si PE-a entend d'un autre PE, disons PE-b, une annonce BGP de VPLS avec le RT-foo et l'identifiant de VE W, alors le PE-a sait que le PE-b est un membre du même VPLS (auto découverte). Le PE-a doit alors établir sa partie d'un pseudo-filaire VPLS entre le PE-a et le PE-b, en utilisant les mécanismes du paragraphe 3.2. De façon similaire, le PE-b va avoir découvert que le PE-a est dans le même VPLS, et le PE-b doit établir sa partie du pseudo-filaire VPLS. Donc, la signalisation et l'établissement du pseudo-filaire sont aussi réalisés avec le même message Update.

Si W n'est dans aucun ensemble de VE distants que PE-a a annoncé pour l'identifiant de VE V dans VPLS foo, PE-b ne va pas être capable d'établir sa part du pseudo-filaire avec le PE-a. Pour traiter cela, PE-a peut choisir de retirer sa ou ses vieilles annonces qu'il avait faites pour le VPLS foo, et annoncer un nouvel Update avec un plus grand ensemble de VE distants et un bloc d'étiquettes correspondant qui couvre tous les identifiants de VE qui sont dans le VPLS foo. Ceci, cependant, peut causer des perturbations de service. Une autre solution pour le PE-a est de créer un nouvel ensemble de VE distants et le bloc d'étiquettes correspondant, et de les annoncer dans un nouvel Update, sans supprimer les annonces précédentes.

Si la configuration du PE-a est changée pour supprimer l'identifiant de VE V du VPLS foo, alors le PE-a DOIT supprimer toutes ses annonces pour VPLS foo qui contiennent l'identifiant de VE V. Si toutes les liaisons du PE-a à ses CE dans le VPLS foo sont closes, alors le PE-a DEVRAIT soit supprimer toutes ses NLRI pour le VPLS foo, soit faire savoir d'une certaine manière aux autres PE dans le VPLS foo que le PE-a n'est plus connecté à ses CE.

3.4 VPLS multi AS

Comme dans [VPN-Tunnel] et la [RFC4364], les fonctions d'auto découverte et de signalisation ci-dessus sont normalement annoncées via I-BGP. Cela suppose que tous les sites dans un VPLS sont connectés aux PE dans un seul système autonome (AS, *Autonomous System*).

Cependant, les sites dans un VPLS peuvent se connecter à des PE dans différents AS. Cela conduit à deux problèmes : 1) il ne va pas y avoir de connexion I-BGP entre ces PE , donc des moyens de signalisation à travers les AS sont nécessaires ; et 2) il peut ne pas y avoir de tunnels de PE à PE entre les AS.
Un problème similaire est résolu à la Section 10 de la [RFC4364]. Trois méthodes sont suggérées pour traiter le problème (1) ; toutes ces méthodes ont leurs analogues dans le VPLS multi AS.

Voici un diagramme de référence :

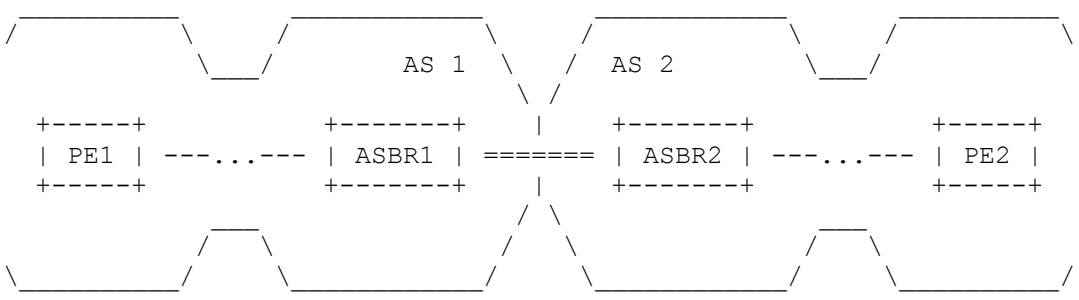

Figure 5 : VPLS inter-AS

Comme dans la référence ci-dessus, on donne trois méthodes pour la signalisation inter fournisseur de VPLS ; elles sont présentées dans l'ordre d'adaptabilité croissante. La méthode (a) est la plus facile à comprendre conceptuellement, et la plus facile à déployer ; cependant, elle exige une interconnexion Ethernet entre les AS, et l'état de plan de contrôle et de plan des données du VPLS sur les routeurs de bordure d'AS (ASBR, *AS Border Router*). La méthode (b) exige l'état du plan de contrôle du VPLS sur les ASBR et de MPLS sur l'interconnexion AS-AS (qui n'a pas besoin d'être Ethernet). La méthode

(c) exige MPLS sur l'interconnexion AS-AS, mais aucune sorte d'état de VPLS sur les ASBR.

3.4.1 Méthode (a) : connexions de VPLS à VPLS aux ASBR

Dans cette méthode, un routeur de bordure d'AS (ASBR1) agit comme un PE pour tous les VPLS qui s'étendent sur l'AS1 et un AS auquel l'ASBR1 est connecté, comme ici l'AS2. L'ASBR sur l'AS voisin (ASBR2) est vu par l'ASBR1 comme un CE pour les VPLS qui s'étendent sur AS1 et AS2 ; de façon similaire, ASBR2 agit comme un PE pour ce VPLS du point de vue de AS2, et voit l'ASBR1 comme un CE.

Cette méthode n'exige pas MPLS sur la liaison ASBR1-ASBR2, mais exige que cette liaison porte du trafic Ethernet et qu'il y ait une sous interface de VLAN séparée pour chaque VPLS qui traverse cette liaison. Elle exige de plus que ASBR1 fasse les opérations de PE (découverte, signalisation, apprentissage de l'adresse MAC, arrosage, encapsulation, etc.) pour tous les VPLS qui traversent ASBR1. Cela impose une charge significative sur ASBR1, à la fois sur le plan de contrôle et sur le plan des données, ce qui limite le nombre de VPLS multi AS.

Noter qu'en général, il va y avoir plusieurs connexions entre une paire d'AS, pour la redondance. Dans ce cas, le protocole d'arborescence d'expansion (STP, *Spanning Tree Protocol*) [802.1D], ou quelque autre moyen de détection et prévention de boucle, doit fonctionner sur chaque VPLS qui s'étend sur ces AS, afin qu'une topologie sans boucle puisse être construite dans chaque VPLS. Cela impose une charge supplémentaire sur les ASBR et le PE qui participent à ces VPLS, car ces appareils vont avoir besoin d'utiliser un algorithme de détection de boucle pour chacun de ces VPLS. La façon dont ce peut être réalisé sort du domaine d'application du présent document.

3.4.2 Méthode (b) : redistribution par EBGP des informations de VPLS entre les ASBR

Cette méthode exige des échanges de trafic I-BGP entre les PE dans AS1 et ASBR1 dans AS1 (peut-être via des réflecteurs de chemin) un échange de trafic E-BGP entre ASBR1 et ASBR2 dans AS2, et des échanges de trafic I-BGP entre ASBR2 et les PE dans AS2. Dans l'exemple ci-dessus, PE1 envoie des NLRI VPLS à ASBR1 avec un bloc d'étiquettes et lui-même comme prochain bond BGP ; ASBR1 envoie les NLRI à ASBR2 avec de nouvelles étiquettes et lui-même comme prochain bond BGP ; et ASBR2 envoie les NLRI à PE2 avec de nouvelles étiquettes et lui-même comme prochain bond. Trois tunnels leur correspondent : T1 de PE1 à ASBR1, T2 de ASBR1 à ASBR2, et T3 de ASBR2 à PE2. Au sein de chaque tunnel, l'étiquette de VPLS à utiliser est déterminée par l'appareil receveur ; par exemple, l'étiquette de VPLS au sein de T1 est une étiquette provenant du bloc d'étiquettes que ASBR1 a envoyé à PE1. Les ASBR sont responsables de la réception des paquets VPLS encapsulés dans un tunnel et d'effectuer les opérations appropriées d'échange d'étiquettes décrites ensuite afin que le prochain appareil receveur puisse correctement identifier et transmettre le paquet.

Les NLRI de VPLS que ASBR1 envoie à ASBR2 (et les NLRI que ASBR2 envoie à PE2) sont identiques aux NLRI de VPLS que PE1 envoie à ASBR1, excepté le bloc d'étiquettes. Pour être précis, les champs Longueur, Différenciateur de chemin, Identifiant de VE, Décalage de bloc de VE, et Taille de bloc de VE DOIVENT être les mêmes ; la base d'étiquette peut être différente. De plus, ASBR1 doit aussi mettre à jour son chemin de transmission comme suit : si la base d'étiquette LB envoyée par PE1 est L1, la taille de bloc d'étiquette est N, la base d'étiquette envoyée par ASBR1 est L2, et l'étiquette de tunnel de ASBR1 à PE1 est T, alors ASBR1 doit installer ce qui suit dans le chemin de transmission :

échanger L2 avec L1 et pousser T,
échanger L2+1 avec L1+1 et pousser T, ...
échanger L2+N-1 avec L1+N-1 et pousser T.

ASBR2 doit agir de la même façon, sauf qu'il peut n'avoir pas besoin d'une étiquette de tunnel si il est directement connecté à ASBR1.

Quand PE2 vent envoyer un paquet VPLS à PE1, PE2 utilise son identifiant de VE pour obtenir la bonne étiquette de VPLS provenant du bloc d'étiquettes de ASBR2 pour PE1, et utilise une étiquette de tunnel pour atteindre ASBR2. ASBR2 échange l'étiquette de VPLS avec l'étiquette provenant de ASBR1 ; ASBR1 échange alors l'étiquette de VPLS avec l'étiquette provenant de PE1, et pousse une étiquette de tunnel pour atteindre PE1.

Dans cette méthode, on a besoin de MPLS sur l'interface ASBR1-ASBR2, mais il n'est pas exigé que la couche de liaison soit Ethernet. De plus, les ASBR prennent part à la distribution des informations de VPLS. Cependant, les exigences de plan des données des ASBR sont beaucoup plus simples que dans la méthode (a), étant limitées aux opérations d'étiquettes. Finalement, la construction de topologies VPLS sans boucle est faite par des décisions d'acheminement, à savoir de chemin BGP et de choix de prochain bond, de sorte qu'il n'y a pas besoin du protocole d'arborescence d'expansion VPLS par VPLS. Donc, cette méthode est considérablement plus adaptable que la méthode (a).

3.4.3 Méthode (c) : redistribution EBGP multi bonds des informations de VPLS entre les AS

Dans cette méthode, il y a un échange de trafic multi bonds E-BGP entre les PE (ou de préférence, un réflecteur de chemin) dans AS1 et les PE (ou réflecteur de chemin) dans AS2. PE1 envoie des NLRI VPLS avec les étiquettes et lui-même comme prochain bond à PE2 ; si c'est via des réflecteurs de chemin, le prochain bond BGP n'est pas changé. Ceci exige qu'il y ait un tunnel LSP de PE1 à PE2. Ce tunnel LSP peut être créé exactement comme dans la [RFC4364], Section 10 (c), par exemple en utilisant E-BGP pour échanger des chemins étiquetés IPv4 pour les bouclages de PE.

Quand PE1 veut envoyer un paquet VPLS à PE2, il pousse l'étiquette de VPLS correspondant à son propre identifiant de VE sur le paquet. Il pousse alors la ou les étiquettes de tunnel pour atteindre PE2.

Cette méthode n'exige aucune information de VPLS (ni dans le plan de contrôle ni dans le plan des données) sur les ASBR. Les ASBR ont seulement besoin d'établir des tunnels LSP de PE à PE dans le plan de contrôle, et de faire les opérations d'étiquettes dans le plan des données. Là encore, comme dans le cas de la méthode (b), la construction de topologies de VPLS sans boucle est faite par des décisions d'acheminement, c'est-à-dire, chemin BGP et choix du prochain bond, de sorte qu'il n'est pas besoin du protocole d'arborescence d'expansion VPLS par VPLS. Cette option est probablement la plus adaptable des trois méthodes présentées ici.

3.4.4 Allocation des identifiants de VE sur plusieurs AS

Afin de faciliter l'allocation des identifiants de VE pour un VPLS qui s'étend sur plusieurs AS, on peut allouer des gammes pour chaque AS. Par exemple, AS1 utilise les identifiants de VE dans la gamme 1 à 100, AS2 de 101 à 200, etc. Si il y a dix sites rattachés à AS1 et 20 à AS2, les identifiants de VE alloués pourraient être 1 à 10 et 101 à 120. Cela minimise le nombre de NLRI VPLS qui sont échangées tout en assurant que les identifiants de VE restent uniques.

Dans l'exemple ci-dessus, si AS1 avait besoin de plus de 100 sites, une autre gamme pourrait alors être allouée à AS1. La seule précaution à prendre est qu'il n'y ait pas de chevauchement entre les gammes d'identifiant de VE parmi les AS. L'exception à cette règle est le multi rattachements, qui est traité dans le paragraphe suivant.

3.5 Multi rattachements et choix de chemin

Il est souvent désiré de faire un multi rattachement d'un site VPLS, c'est-à-dire, de le connecter à plusieurs PE, peut-être même dans des AS différents. Dans ce cas, les PE connectés au même site peuvent être configurés soit avec le même identifiant de VE, soit avec des identifiants différents. Dans ce dernier cas, il est obligatoire de faire fonctionner STP sur l'appareil CE, et éventuellement sur les PE, pour construire une topologie de VPLS sans boucle. Comment cela peut être accompli sort du domaine d'application du présent document ; cependant, le reste de cette section va décrire avec un peu de détails le premier cas. Noter que les multi rattachements par le SP et STP sur les CE peuvent coexister; donc, il est recommandé que le consommateur de VPLS fasse fonctionner STP si les CE en sont capables.

Dans le cas où les PE connectés au même site ont le même identifiant de VE alloué, une topologie sans boucle est construite par les mécanismes d'acheminement, en particulier, par le choix de chemin BGP. Quand un locuteur BGP reçoit deux NLRI équivalentes (voir la définition ci-dessous) il applique les critères standard de choix de chemin comme la préférence locale et la longueur de chemin d'AS pour déterminer quelles NLRI choisir ; il DOIT en prendre une seule. Si les NLRI choisies sont ultérieurement retirées, le locuteur BGP applique le choix de chemin aux NLRI VPLS équivalentes restantes pour en prendre d'autres ; si il n'en reste aucune, les informations de transmission associées à ces NLRI sont supprimées.

Deux NLRI VPLS sont considérées équivalente du point de vue du choix de chemin si le différenciateur de chemin, l'identifiant de VE, et le décalage de bloc de VE sont les mêmes. Si deux PE ont le même identifiant de VE alloué dans un certain VPLS, ils DOIVENT utiliser le même différenciateur de chemin, et ils DEVRAIENT annoncer la même taille de bloc de VE pour un décalage de VE donné.

3.6 VPLS BGP hiérarchique

Ce paragraphe discute comment on peut adapter le plan de contrôle VPLS en utilisant BGP. Il y a au moins trois aspects d'adaptation du plan de contrôle :

1. adoucir l'exigence de connexité de maillage complet entre les locuteurs BGP du VPLS ;

2. limiter le passage de messages BGP VPLS à juste les locuteurs intéressés plutôt qu'à tous les locuteurs BGP ; et
3. simplifier l'ajout et la suppression des locuteurs BGP, pour VPLS ou pour les autres applications.

Heureusement, l'utilisation de BGP pour l'acheminement Internet ainsi que pour les VPN IP a donné plusieurs bonnes solutions pour tous ces problèmes. La technique de base est hiérarchique, en utilisant les réflecteurs de chemin (RR, *Route Reflector*) [RFC4456] de BGP. L'idée est de désigner un petit ensemble de RR qui sont eux-mêmes pleinement maillés, et ensuite d'établir une session BGP entre chaque locuteur BGP et un ou plusieurs RR. De cette façon, il n'est pas besoin d'une connexité de maillage complet directe entre tous les locuteurs BGP. Si les besoins particuliers d'adaptabilité d'un fournisseur exigent un grand nombre de RR, alors cette technique peut être appliquée de façon récurrente : la connexité de maillage complet entre les RR peut être relayée par encore un autre niveau de RR. L'utilisation des RR résout les problèmes 1 et 3 ci-dessus.

Il est important de noter que les RR, tels qu'utilisés pour VPLS et les VPN, sont purement une technique de plan de contrôle. L'utilisation des RR n'introduit pas d'état au plan des données et pas d'exigence de transmission au plan des données sur les RR, et ne change en aucune façon le chemin de transmission du trafic VPLS. Ceci est différent de la technique du VPLS hiérarchique définie dans la [RFC4762].

Une autre conséquence de cette approche est qu'il n'est pas exigé qu'un ensemble de RR traite tous les messages BGP, ou qu'un RR particulier traite tous les messages provenant d'un certain PE. On peut définir plusieurs ensembles de RR, par exemple, un ensemble pour traiter VPLS, un autre pour traiter les VPN IP, et un autre pour l'acheminement Internet. Un autre partage pourrait être d'avoir un certain sous ensemble des VPLS et VPN IP traité par un ensemble de RR, et un autre sous ensemble de VPLS et VPN IP traité par un autre ensemble de RR ; l'utilisation du filtrage de chemins cibles (RTF, *Route Target Filtering*) décrite dans la [RFC4684], peut rendre ceci plus simple et plus efficace.

Finalement, le problème 2 (celui de limiter le passage de messages VPLS BGP aux seuls locuteurs BGP intéressés) est réglé par l'utilisation de RTF. Cette technique est orthogonale à l'utilisation des RR, mais fonctionne bien en conjonction avec les RR. RTF est aussi très efficace dans le VPLS inter AS ; plus de détails sur la façon dont RTF fonctionne et ses avantages sont fournis dans la [RFC4684].

Il paraît utile de mentionner un aspect du plan de contrôle qui est souvent une source de confusion. Aucune adresse MAC n'est échangée via BGP. Tout l'apprentissage d'adresses MAC et leur vieillissement est fait dans le plan des données individuellement par chaque PE. La seule tâche de l'échange de message VPLS de BGP est l'auto découverte et l'échange d'étiquettes.

Donc, le traitement de BGP pour les VPLS se produit quand
1. un PE se joint ou quitte un VPLS ; ou
2. une défaillance se produit dans le réseau, détruisant un tunnel de PE à PE ou une liaison de PE à CE.

Ces événements sont relativement rares, et normalement, chacun de ces événements cause la génération d'une mise à jour BGP. Couplées avec l'efficacité de la messagerie de BGP quand elle est utilisée pour la signalisation de VPLS, ces observations mènent à la conclusion que BGP comme plan de contrôle pour VPLS va bien s'adapter en termes de traitement et d'exigence de mémoire.

4. Plan des données

Cette Section discute deux aspects du plan des données pour les PE et u-PE qui mettent en œuvre VPLS : l'encapsulation et la transmission.

4.1 Encapsulation

Les trames Ethernet reçues des appareils CE sont encapsulées pour la transmission sur le réseau à commutation de paquets qui connecte les PE. L'encapsulation est comme décrit dans la [RFC4448].

4.2 Transmission

Les paquets VPLS sont classés comme appartenant à une instance de service donnée et à un tableau de transmission associé fondé sur l'interface sur laquelle le paquet est reçu. Les paquets sont transmis dans le contexte de l'instance de service sur la base de l'adresse MAC de destination. La première transposition est déterminée par la configuration. La dernière est l'objet

de cette section.

4.2.1 Acquisition d'adresse MAC

Comme mentionné précédemment, la caractéristique clé distinctive de VPLS est qu'il est un service en multipoints. Cela signifie que le réseau entier de fournisseur de services devrait apparaître comme un seul pont d'apprentissage logique pour chaque VPLS que le réseau de fournisseur de service prend en charge. Les accès logiques pour le "pont" de fournisseur de services sont les accès de consommateur ainsi que les pseudo-filaires sur un VE. Tout comme un pont d'apprentissage apprend les adresses MAC sur ses accès, le pont de SP doit apprendre les adresses MAC à ses VE.

L'apprentissage consiste en l'association des adresses MAC de source des paquets avec les accès (logiques) sur lesquels ils arrivent ; cette association est la base de données d'informations de transmission (FIB, *Forwarding Information Base*). La FIB est utilisée pour transmettre les paquets. Par exemple, supposons que le pont reçoive un paquet avec l'adresse MAC de source S sur l'accès (logique) P. Si ultérieurement, le pont reçoit un paquet avec l'adresse MAC de destination S, il sait qu'il devrait envoyer le paquet sur l'accès P.

Si un VE apprend une adresse de source MAC S sur l'accès logique P, puis voit ensuite S sur un accès différent P', alors le VE DOIT mettre à jour sa FIB pour refléter le nouvel accès P'. Un VE PEUT mettre en œuvre un mécanisme pour atténuer les oscillations d'accès de source pour une adresse MAC donnée.

4.2.2 Vieillissement

Les PE VPLS DEVRAIENT avoir un mécanisme de vieillissement pour supprimer une adresse MAC associée à un accès logique, un peu le même que le font les ponts apprenants.Ceci est exigé pour qu'une adresse MAC puisse être apprise de nouveau si elle "bouge" d'un accès logique à un autre, soit parce que la station à laquelle appartient cette adresse a réellement bougé, soit parce qu'un changement topologique dans le LAN cause l'arrivée de cette adresse MAC sur un nouvel accès. De plus, le vieillissement réduit la taille d'un tableau MAC de VPLS à seulement les adresses MAC actives, plutôt que toutes les adresses MAC dans ce VPLS.

L'"âge" d'une adresse MAC de source S sur un accès logique P est le temps écoulé depuis la dernière fois qu'elle a été vue comme source MAC sur l'accès P. Si l'âge excède le temps de vieillissement T, S DOIT être purgé de la FIB. Ceci signifie bien sûr que chaque fois que S est vu comme adresse de source MAC sur l'accès P, l'âge de S est réinitialisé.

Une mise en œuvre DEVRAIT fournir un bouton configurable pour régler le temps de vieillissement T par VPLS. De plus, une mise en œuvre PEUT accélérer le vieillissement de toutes les adresses MAC dans un VPLS si elle détecte certaines situations, comme un changement de topologie d'arborescence d'expansion dans ce VPLS.

4.2.3 Arrosage

Quand un pont reçoit un paquet pour une destination qui n'est pas dans sa FIB, il arrose le paquet sur tous les autres accès. De façon similaire, un VE va arroser les paquets de destination inconnue à tous les autres VE dans le VPLS.

Dans la Figure 1 ci-dessus, si CE2 envoie une trame Ethernet à PE2, et si l'adresse MAC de destination sur la trame n'est pas dans la FIB de PE2 (pour ce VPLS) alors PE2 va être responsable de l'arrosage de cette trame à tous les autres PE dans le même VPLS. À réception de cette trame, PE1 va être responsable de la poursuite de l'arrosage de la trame à CE1 et CE5 (sauf si PE1 sait quel CE "possède" cette adresse MAC).

Par ailleurs, si PE3 a reçu la trame, il pourrait déléguer la poursuite de l'arrosage de la trame à ses u-PE. Si PE3 est connecté à deux u-PE, il annoncerait cela à ses deux u-PE. PE3 pourrait soit annoncer qu'il est incapable d'arroser, auquel cas il recevrait deux trames, une pour chaque u-PE, soit il pourrait annoncer qu'il est capable d’arroser, et dans ce cas il recevrait une copie de la trame, qu'il enverrait alors aux deux u-PE.

4.2.4 Diffusion et diffusion groupée

Il y a une adresse MAC de diffusion bien connue. Une trame Ethernet dont l'adresse MAC de destination es l'adresse MAC de diffusion doit être envoyée à toutes les stations dans ce VPLS. Cela peut être accompli par les mêmes moyens qu'utilisés pour l'arrosage.

Il y a aussi un ensemble facilement reconnaissable d'adresses MAC de "diffusion groupée". Les trames Ethernet avec une adresse MAC de destination de diffusion groupée PEUVENT être diffusées à toutes les stations ; un VE PEUT aussi utiliser certaines techniques pour restreindre la transmission des trames de diffusion groupée à un plus petit ensemble de receveurs, ceux qui ont indiqué leur intérêt pour le groupe correspondant de diffusion groupée. La discussion de ceci sort du domaine d'application du présent document.

4.2.5 Transmission en "horizon partagé"

Quand un PE capable d'arrosage (disons le PEx) reçoit une trame de diffusion Ethernet, ou qui a une adresse MAC de destination inconnue, il doit arroser la trame. Si la trame est arrivée d'un CE rattaché, le PEx doit envoyer une copie de la trame à tous les autres CE rattachés, ainsi qu'à tous les autres PE qui participent au VPLS. Si, par ailleurs, la trame est arrivée d'un autre PE (disons PEy) PEx doit envoyer une copie du paquet aux seuls CE rattachés. Le PEx NE DOIT PAS envoyer la trame aux autres PE, car le PEy l'aurait déjà fait. Cette notion a été appelée la transmission en "horizon partagé" et est une conséquence de ce que les PE sont logiquement pleinement maillés pour VPLS.

Les règles de la transmission en horizon partagé s'appliquent aux paquets en diffusion et en diffusion groupée, ainsi qu'aux paquets pour une adresse MAC inconnue.

4.2.6 Acquisition qualifiée et non qualifiée

La clé de l'apprentissage normal de MAC Ethernet est généralement juste les 6 octets de l'adresse MAC. C'est appelé "apprentissage non qualifié". Cependant, il est aussi possible que la clé de l'apprentissage inclut l'étiquette de VLAN quand elle est présente ; c'est appelé "apprentissage qualifié".

Dans le cas de VPLS, l'apprentissage est fait dans le contexte d'une instance de VPLS, qui correspond normalement à un consommateur. Si le consommateur utilise des étiquettes de VLAN, on peut faire les mêmes distinctions d'apprentissage qualifié et non qualifié. Si la clé de l'apprentissage au sein d'un VPLS est juste l'adresse MAC, alors ce VPLS fonctionne en apprentissage non qualifié. Si la clé de l'apprentissage est (étiquette de VLAN de consommateur + adresse MAC) ce VPLS fonctionne alors en apprentissage qualifié.

Choisir entre apprentissage qualifié et non qualifié implique plusieurs facteurs, dont le plus important est si on veut un seul domaine de diffusion global (non qualifié) ou un domaine de diffusion par VLAN (qualifié). Ce dernier rend l'arrosage et la diffusion plus efficaces, mais exige de plus grands tableaux de MAC. Ces considérations s'appliquent également à la transmission normale Ethernet et aux VPLS.

4.2.7 Classe de service

Afin d'offrir différentes classes de service au sein d'un VPLS, une mise en œuvre PEUT choisir de transposer les bits 802.1p dans une trame Ethernet de consommateur avec une étiquette de VLAN en un réglage approprié des bits EXP dans le pseudo-filaire et/ou étiquette de tunnel, permettant un traitement différentiel des trames VPLS dans le réseau à commutation de paquets.

Pour être utile, une mise en œuvre DEVRAIT permettre que cette fonction de transposition soit différente pour chaque VPLS, car chaque consommateur VPLS peut avoir sa propre vue du comportement requis pour un certain réglage des bits 802.1p.

5. Options de déploiement

En déployant un réseau qui prend en charge VPLS, le SP doit décider quelles fonctions prend en charge l'appareil à capacité VPLS le plus proche du consommateur (le VE). Le cas par défaut décrit dans ce document est que le VE est un PE. Cependant, il y a un certain nombre de raisons pour que le VE puisse être un appareil qui assure toutes les fonctions de couche 2 (comme l'apprentissage d'adresse MAC et l'arrosage) et un ensemble limité de fonctions de couche 3 (comme de communiquer avec son PE) mais, par exemple, ne fasse pas complètement la découverte et la signalisation de PE à PE. Un tel appareil est appelé un "u-PE".

Comme ces deux cas ont des avantages, on aimerait être capable de "mixer et faire correspondre" ces scénarios. Le mécanisme de signalisation présenté ici le permet. Par exemple, dans le réseau d'un certain fournisseur, un PE peut être

directement connecté aux appareils CE, un autre peut être connecté aux u-PE qui sont connecté aux CE, et un troisième peut être connecté directement à un consommateur sur certaines interfaces et aux u-PE sur les autres. Tous ces PE effectuent la découverte et la signalisation de la même manière. Comment ils font l'apprentissage et la transmission dépend de si il y a ou non un u-PE ; cependant, ceci est une affaire locale, et n'est pas signalé. Cependant, les détails du fonctionnement d'un u-PE et ses interactions avec les PE et autres u-PE sortent du domaine d'application de ce document.

6. Considérations sur la sécurité

Le point central du service de LAN privé virtuel est la confidentialité des données, c'est-à-dire, que les données dans un VPLS ne sont distribuées qu'aux autre nœuds dans ce VPLS et à aucun agent externe ou autre VPLS. Noter que le VPLS n'offre pas la confidentialité, l'intégrité, ou l'authentification : les paquets VPLS sont envoyés en clair dans le réseau à commutation de paquets, et un interposé peut les espionner, et peut être capable d'injecter des paquets dans le flux des données. Si la sécurité est désirée, les tunnels de PE à PE peuvent être des tunnels IPsec. Pour plus de sécurité, les systèmes d'extrémité dans les sites de VPLS peuvent utiliser des moyens appropriés de chiffrement pour sécuriser leurs données même avant qu'elles entrent dans le réseau du fournisseur de services.

Il y a deux aspects pour réaliser la confidentialité des données dans un VPLS : sécuriser le plan de contrôle et protéger le chemin de transmission. Compromettre le plan de contrôle pourrait résulter en ce qu'un PE envoie des données appartenant à un VPLS à un autre VPLS, ou d'envoyer les données du VPLS dans un trou noir, ou même de les envoyer à un espion ; aucune de ces situations n'est acceptable du point de vue de la confidentialité des données. Comme tous les échanges du plan de contrôle sont via BGP, des techniques comme celles de la [RFC2385] aident à authentifier les messages BGP, rendant plus difficile de falsifier les mises à jour (qui peuvent être utilisées pour détourner le trafic VPLS sur un mauvais VPLS) ou le supprimer (attaques de déni de service). Dans les méthodes multi AS (b) et (c) décrites à la Section 3, cela signifie aussi de protéger les sessions BGP inter AS, entre les ASBR, les PE, ou les réflecteurs de chemins. On peut aussi utiliser les techniques décrites à la Section 10 (b) et (c) de la [RFC4364], à la fois pour le plan de contrôle et le plan des données. Noter que la [RFC2385] ne va pas aider à garder la confidentialité des étiquettes VPLS – en connaissant les étiquettes, on peut espionner le trafic VPLS. Cependant, ceci exige l'accès au chemin des données au sein d'un réseau de fournisseur de services.

Il peut aussi y avoir des mauvaises configurations conduisant à la connexion involontaire de CE dans des VPLS différents. Ceci peut être causé, par exemple, par l'association de la mauvaise route cible avec une instance de VPLS. Ce problème, partagé par la [RFC4364], fera l'objet de futures études.

La protection du plan des données exige de s'assurer que les tunnels de PE à PE se comportent bien (ceci sort du domaine d'application du présent document) et que les étiquettes VPLS ne sont acceptées que d'interfaces valides. Pour un PE, les interfaces valides comprennent les liaisons provenant des routeurs P. Pour un ASBR, une interface valide est une liaison provenant d'un ASBR dans un AS qui fait partie d'un certain VPLS. Il est particulièrement important dans le cas des VPLS multi AS qu'on n'accepte que les paquets VPLS provenant d'interfaces valides.

Le tunnelage MPLS dans IP et MPLS dans GRE sont spécifiés dans la [RFC4023]. Si on désire utiliser de tels tunnels pour porter les paquets VPLS, alors les considérations sur la sécurité décrites dans la Section 8 de de document doivent être bien comprises. Toute mise en œuvre de VPLS qui permet que les paquets VPLS soient tunnelés comme décrit dans ce document DOIT contenir une mise en œuvre de IPsec qui puisse être utilisée comme il y est décrit. Si le tunnel n'est pas sécurisé par IPsec, alors la technique du filtrage d'adresse IP aux routeurs de bordure, décrite au paragraphe 8.2 de ce document, est le seul moyen de s'assurer qu'un paquet qui sort du tunnel à un PE de sortie particulier a réellement été placé dans le tunnel par le nœud de tête de tunnel approprié (c'est-à-dire, que le paquet n'a pas une adresse de source falsifiée). Comme les routeurs de bordure filtrent fréquemment seulement les adresses de source, le filtrage de paquet peut n'être pas efficace sauf si le PE de sortie peut vérifier l'adresse IP de source de tout paquet tunnelé qu'il reçoit, et la comparer à une liste d'adresses IP qui sont des adresses de tête de tunnel valides. Toute mise en œuvre qui permet le tunnelage MPLS dans IP et/ou MPLS dans GRE sans IPsec DOIT permettre au PE de sortie de valider de cette manière l'adresse IP de source de tout paquet tunnelé qu'elle reçoit.

7. Considérations relatives à l'IANA

L'IANA a alloué la valeur (25) pour l'AFI d'informations de L2VPN. Ce devrait être la même que l'AFI demandé par [RFC5195].

L'IANA a alloué une valeur de communauté étendue de (0x800a) pour la communauté étendue d'informations de couche 2.

8. Références

8.1 Références normatives

[RFC2119] S. Bradner, "Mots clés à utiliser dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. *(MàJ par* RFC8174)

[RFC2385] A. Heffernan, "Protection des sessions de BGP via l'option de signature MD5 de TCP", août 1998. *(P.S. ; MàJ par la* RFC6691) *; remplacée par* RFC5925)

[RFC4023] T. Worster et autres, "Encapsulation de MPLS dans IP ou encapsulation d'acheminement générique (GRE)", mars 2005. (*MàJ par* RFC5332) *(P.S.)*

[RFC4360] S. Sangli et autres, "Attribut BGP-4 Communauté étendue", février 2006. *(P.S.)*

[RFC4364] E. Rosen et Y. Rekhter, "Réseaux privés virtuels IP BGP/MPLS", février 2006. (*P.S., MàJ par* RFC4577, RFC4684)

[RFC4448] L. Martini et autres, "Méthodes d'encapsulation pour le transport d'Ethernet sur des réseaux MPLS", avril 2006. *(P.S. ; MàJ par* RFC8469)

[RFC4760] T. Bates, R. Chandra, D. Katz et Y. Rekhter, "Extensions multi protocoles pour BGP-4", janvier 2007.

8.2 Références pour information

[802.1D] Institute of Electrical and Electronics Engineers, "Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Common specifications - Part 3: Media Access Control (MAC) Bridges: Revision." C'est une révision de la norme ISO/CEI 10038: 1993, 802.1j- 1992 et 802.6k-1992. Elle incorpore P802.11c, P802.1p et P802.12e. ISO/CEI 15802-3: 1998., IEEE Standard 802.1D, juillet 1998.

[RFC4447] L. Martini et autres, "Établissement et maintenance de pseudo filaires avec le protocole de distribution d'étiquettes", avril 2006. (*MàJ par la* RFC6723) *(P.S. ; Remplacé par* RFC8077 STD 84)

[RFC4456] T. Bates, E. Chen, R. Chandra, "Réflexion de chemin BGP : une solution de remplacement au BGP interne à maillage complet (IBGP)", avril 2006. (*Remplace* RFC2796, RFC1966) *(D.S.)*

[RFC4664] L. Andersson et E. Rosen, éd., "Cadre pour les réseaux virtuels privés de couche 2 (L2VPN)", septembre 2006. *(Info.)*

[RFC4762] M. Lasserre et V. Kompella, éditeurs, "Service de LAN privé virtuel (VPLS) utilisant la signalisation du protocole de distribution d'étiquette (LDP)", janvier 2007. *(P.S.)*

[RFC5195] H. Ould-Brahim et autres, "Auto découverte fondée sur BGP pour VPN de couche 1", juin 2008. *(P.S.)*

[RFC4684] P. Marques et autres, "Distribution de chemin contraint pour réseaux privés virtuels (VPN) au protocole Internet selon le protocole de routeur frontière/commutation d'étiquettes multiprotocoles (BGP/MPLS)", novembre 2006. *(P.S.)*

[VPN-Tunnel] Kompella, K., "Layer 2 VPNs Over Tunnels", Travail en cours, janvier 2006.

Appendice A. Contributeurs

Les personnes suivantes ont contribué à ce document : Javier Achirica, Telefonica ; Loa Andersson, Acreo ; Giles Heron,

Tellabs ; Sunil Khandekar, Alcatel-Lucent ; Chaitanya Kodeboyina, Nuova Systems : Vach Kompella, Alcatel-Lucent ; Marc Lasserre, Alcatel-Lucent ; Pierre Lin ; Pascal Menezes ; Ashwin Moranganti, Appian ; Hamid Ould-Brahim, Nortel ; Seo Yeong-il, Korea Tel.

Appendice B. Remerciements

Merci à Joe Regan et Alfred Nothaft de leurs contributions. Un grand merci aussi à Eric Ji, Chaitanya Kodeboyina, Mike Loomis, et Elwyn Davies pour leur relecture détaillée.

Adresse des éditeurs

Kireeti Kompella
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US
mél : kireeti@juniper.net

Yakov Rekhter
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US
mél : yakov@juniper.net

Déclaration complète de droits de reproduction

Remerciement
Le financement de la fonction d'édition des RFC est fourni par l'activité de soutien administratif (IASA) de l'IETF.