Groupe de travail Réseau
Request for Comments : 4951
Catégorie : Sur la voie de la normalisation

V. Jain, éd., Riverstone Networks
août 2007
Traduction Claude Brière de L'Isle

Extensions de reprise sur défaillance pour le protocole de tunnelage de couche2 (L2TP)

Statut du présent mémoire

Le présent document spécifie un protocole de l'Internet sur la voie de la normalisation pour la communauté de l'Internet, et appelle à des discussions et suggestions pour son amélioration. Prière de se référer à l'édition en cours des "Protocoles officiels de l'Internet" (STD 1) pour voir l'état de normalisation et le statut de ce protocole. La distribution du présent mémoire n'est soumise à aucune restriction.

Résumé

Le protocole de tunnelage de couche 2 (L2TP, *Layer 2 Tunneling Protocol*) est un protocole en mode connexion qui a un état partagé entre des points d'extrémité actifs. Une partie de cet état partagé est vitale pour le fonctionnement, mais peut être de nature volatile, comme des numéros de séquence de paquets utilisés sur la connexion de contrôle L2TP. Quand se produit une défaillance sur un côté d'une connexion de contrôle, une nouvelle connexion de contrôle est créée et est associée à l'ancienne connexion par l'échange d'informations sur l'ancienne connexion. Ce mécanisme n'est pas destiné à remplacer une reprise sur défaillance active avec des états de connexion reflétés, mais comme une aide pour les paramètres qu'il est particulièrement difficile de rendre immédiatement disponibles. Les extensions de protocole L2TP définies dans le présent document sont destinées à faciliter la récupération d'état, fournissant une résilience supplémentaire dans le réseau L2TP, et améliorant la connexité de couche 2 d'un système distant.

Table des matières

1. Introduction

Le but du présent document est d'aider à la résilience globale d'un point d'extrémité L2TP en introduisant des extensions aux [RFC2661] et [RFC3931] qui vont minimiser le temps de récupération de la couche L2TP après une reprise sur défaillance, tout en minimisant l'impact sur ses performances. Donc, il est supposé que l'architecture globale du point d'extrémité est aussi un support de l'effort de résilience.

Pour assurer un fonctionnement approprié d'un point d'extrémité L2TP après une reprise sur défaillance, les informations associées à la connexion de contrôle et aux sessions entre elles doivent être correctes et cohérentes. Cela inclut les informations configurées et dynamiques. Les informations configurées sont supposées être correctes et cohérentes après une reprise sur défaillance, autrement les tunnels et sessions n'auraient pas été établis en premier lieu.

Les informations dynamiques, qui sont aussi appelées les informations à état plein, changent avec le traitement des paquets de contrôle et de données du tunnel. Actuellement, les seules informations de cette sorte qui soient essentielles au fonctionnement du tunnel sont ses numéros de séquence. Pour le canal de contrôle du tunnel, les incohérences dans les numéros de séquence peuvent résulter en la terminaison du tunnel entier. Pour les sessions de tunnel, l'incohérence de leurs numéros de séquence, quand ils sont utilisés, peut causer des pertes de données significatuves, ce qui fait percevoir une "perte de service" à l'utilisateur final.

Donc, une architecture résiliente optimale qui vise à minimiser la "perte de service" après une reprise sur défaillance, doit prendre des dispositions pour les informations d'état plein essentielles du tunnel, c'est-à-dire, ses numéros de séquence. Actuellement, deux options sont disponibles : la première est de s'assurer que le point d'extrémité de récupération est complètement synchronisé avec le point d'extrémité actif, par rapport aux numéros de séquence des sessions de contrôle et de données. La seconde option est de rétablir tous les tunnels et leurs sessions après une reprise sur défaillance. L'inconvénient de la première option est qu'elle ajoute un impact significatif sur les performances et la complexité de l'architecture du point d'extrémité, en particulier lorsque augmente l'agrégation de tunnels et de sessions. L'inconvénient de la seconde option est qu'elle augmente le temps de "perte de service", en particulier avec l'extension de l'architecture.

Pour alléger les inconvénients sus-mentionnés des options actuelles, le présent document introduit un mécanisme pour amener les informations dynamiques d'état plein d'un tunnel à un état correct et cohérent après une défaillance. Le mécanisme proposé définit la récupération des tunnels et sessions qui étaient dans un état établi avant la défaillance.

1.1 Terminologie

Point d'extrémité : point d'extrémité de connexion de contrôle L2TP, c'est-à-dire, un LAC ou LNS (aussi appelé LCCE dans la [RFC3931]).

Point d'extrémité actif : point d'extrémité qui fournit actuellement le service.

Point d'extrémité de sauvegarde : point d'extrémité redondant pour le point d'extrémité actif qui a sa base de données de tunnels et sessions actifs synchronisée avec son point d'extrémité actif.

Point d'extrémité défaillant : le point d'extrémité qui était le point d'extrémité actif au moment de la défaillance.

Point d'extrémité de récupération : point d'extrémité qui initie le protocole de reprise sur défaillance pour récupérer de la défaillance d'un point d'extrémité actif.

Point d'extrémité distant : point d'extrémité qui échange du trafic avec le point d'extrémité actif avant la défaillance et avec le point d'extrémité de récupération après la défaillance.

Reprise sur défaillance : action d'un point d'extrémité de sauvegarde sur le service d'un point d'extrémité actif. Ce pourrait être dû à une action administrative ou à la défaillance du point d'extrémité actif.

Ancien tunnel : connexion de contrôle qui existait avant la défaillance et soumise à récupération lors de la reprise sur défaillance.

Tunnel de récupération : nouvelle connexion de contrôle établie seulement pour récupérer un ancien tunnel.

Tunnel récupéré : après que la connexion de contrôle et les sessions d'un ancien tunnel sont restaurées en utilisant le

mécanisme décrit dans le présent document, on l'appelle un tunnel récupéré.

Défaillance de canal de contrôle : défaillance du composant responsable de l'établissement/maintenance des tunnels et sessions à un point d'extrémité.

Défaillance de canal de données : défaillance du composant responsable de la transmission des données encapsulées dans L2TP.

1.2 Abréviations

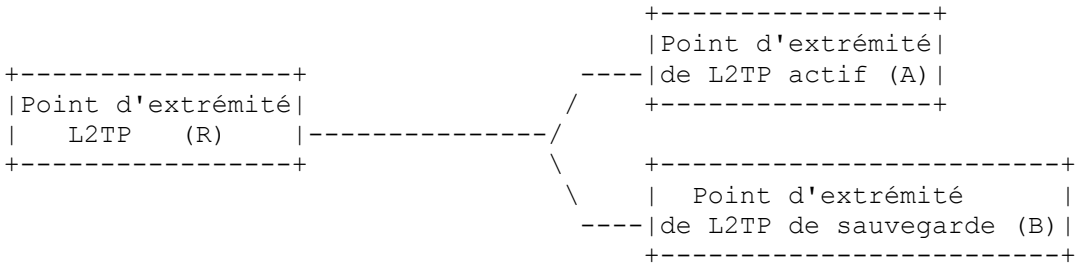
LAC *(L2TP Access Concentrator)* : concentrateur d'accès L2TP
LNS *(L2TP Network Server)* : serveur réseau L2TP
LCCE *(L2TP Control Connection Endpoint)* : point d'extrémité de connexion de contrôle L2TP
AVP *(Attribute Value Pair)* : paire attribut/valeur
SCCRQ *(Start-Control-Connection-Request)* : demande de début de connexion de contrôle
SCCRP *(Start-Control-Connection-Reply)* : réponse de début de connexion de contrôle
ZLB *(Zero Length Body Message)* : message de longueur de corps de zéro

1.3 Notation des exigences

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. Vue d'ensemble

Le diagramme suivant décrit l'architecture de redondance et les entités pertinentes utilisées pour décrir le protocole de reprise sur défaillance :



Les points d'extrémité actif et de sauvegarde peuvent résider sur le même appareil, cependant, ils ne sont pas obligés de l'être. Par ailleurs, certains appareils peuvent n'avoir pas de module d'attente avec eux, et dans ce cas, le point d'extrémité défaillant, après le rétablissement, peut devenir le point d'extrémité de récupération pour récupérer de sa défaillance antérieure.

Donc, dans le diagramme ci-dessus, lors d'une défaillance de A (le point d'extrémité actif) :
- le point d'extrémité A va être appelé le point d'extrémité défaillant,
- si B est présent, il va alors devenir le point d'extrémité de récupération et aussi un point d'extrémité actif,
- si B n'est pas présent, alors A pourrait devenir le point d'extrémité de récupération après son redémarrage, pourvu qu'il ait sauvegardé les informations sur les tunnels/sessions actifs dans une mémorisation persistente,
- R n'initie pas le protocole de reprise sur défaillance ; il attend plutôt une indication de défaillance de la part du point d'extrémité de récupération.

Le présent document suppose que la détection réelle d'une défaillance dans le point d'extrémité de sauvegarde est faite d'une façon spécifique de la mise en œuvre. Il suppose aussi que la détection de défaillance par le point d'extrémité de sauvegarde est plus rapide que la temporisation du canal de contrôle L2TP entre les points d'extrémité actif et distant. Normalement, les points d'extrémité actif et de sauvegarde résident sur le même appareil physique, permettant une détection de défaillance rapide et fiable sans qu'il soit besoin de communiquer sur le réseau entre ces points d'extrémité.

De même, un point d'extrémité actif qui agit aussi comme point d'extrémité de sauvegarde peut aisément commencer la récupération une fois qu'il s'est réamorcé. Cependant, quand les points d'extrémité actif et de sauvegarde résident dans des appareils séparés, il y a besoin de communication entre eux afin dedétecter les défaillances. Comme solution pour de telles situations, des protocoles de détection de défaillance supplémentaires, par exemple, de la [RFC5883], peuvent être nécessaires.

Un appareil pourrait avoir trois sortes de défaillances :
i) défaillance du canal de contrôle
ii) défaillance du canal de données
iii) défaillance du canal de contrôle et du canal de données

Le protocole décrit dans le présent document spécifie la récupération dans les conditions i) et iii). Il est estimé qu'il ne pourrait pas être récupéré beaucoup (d'informations d'état plein) via un échange de protocole de contrôle dans le cas ii).

Le protocole de reprise sur défaillance comporte trois phases :

1) Négociation de capacité de reprise sur défaillance : un point d'extrémité actif et un point d'extrémité distant échangent les capacités et attributs de reprise sur défaillance à utiliser durant le processus de récupération.

2) Récupération de reprise sur défaillance : un point d'extrémité de récupération établit une nouvelle connexion de contrôle L2TP (appelée tunnel de récupération) pour chaque ancien tunnel qu'il souhaite récupérer. Le tunnel de récupération sert à trois objets :
   - Il identifie l'ancien tunnel qui est à récupérer,
   - il fournit un moyen d'authentification et une prise de contact en trois phases pour s'assurer que les deux extrémités sont d'accord pour la reprise sur défaillance pour l'ancien tunnel spécifié,
   - il pourrait échanger les valeurs Ns et Nr à utiliser dans le tunnel récupéré.

   À l'établissement du tunnel de récupération, les deux points d'extrémité rétablissent les canaux de contrôle et de données sur le tunnel récupéré en utilisant les procédures décrites respectivement aux paragraphes 3.2.2 et 3.2.3. Le tunnel de récupération pourrait être supprimé après cela, ainsi que les sessions établies pour reprendre le trafic.

3) Synchronisation d'état de session : le processus de synchronisation d'état de session survient à la récupération de l'ancien tunnel et permet aux deux points d'extrémité de s'accorder sur l'état des diverses sessions dans le tunnel après la reprise sur défaillance. Les incohérences, qui pourraient survenir dues à la défaillance, sont traitées de la manière suivante : d'abord, les deux points d'extrémité éliminent en silence les sessions qui n'étaient pas dans l'état établi. Ensuite, elles utilisent l'interrogation de session de reprise sur défaillance (FSQ, *Failover Session Query*) et la réponse de session de reprise sur défaillance (FSR, *Failover Session Response*) sur le tunnel récupéré pour obtenir l'état des sessions connu par le point d'extrémité homologue et supprimer les sessions en conséquence.

3. Protocole de reprise sur défaillance

Le protocole consiste en les trois étapes décrivant les spécifications durant la vie d'une connexion de contrôle - avant et après la reprise sur défaillance.

3.1 Négociation de capacité de reprise sur défaillance

Les points d'extrémité actif et distant échangent la paire d'attribut-valeur (AVP, *attribute-value pair*) Capacité de reprise sur défaillance dans les messages Demande de début de connexion de contrôle (SCCRQ, *Start-Control-Connection-Request*) et Réponse de début de connexion de contrôle (SCCRP, *Start-Control-Connection-Reply*) durant l'établissement de la connexion de contrôle au titre du fonctionnement normal (avant la reprise sur défaillance). L'AVP Capacité de reprise sur défaillance, définie au paragraphe 5.1, permet à un point d'extrémité de spécifier si il est capable de reprise sur défaillance de canal de contrôle et/ou de données et le délai permis pour la récupération du tunnel.

3.2 Procédure de récupération de reprise sur défaillance

La procédure de récupération de reprise sur défaillance décrite dans ce paragraphe n'est effectuée que si il y avait une défaillance du canal de contrôle. Le choix des tunnels à récupérer est spécifique de la mise en œuvree.

La procédure de récupération de reprise sur défaillance consiste en les trois étapes suivantes, qui sont décrites en détails

dans les paragraphes qui suivent :
- établissement du tunnel de récupération
- rétablissement du canal de contrôle
- rétablissement du canal de données

3.2.1 Établissement de tunnel de récupération

Le point d'extrémité de récupération établit une nouvelle connexion de contrôle, appelée tunnel de récupération, pour chaque ancien tunnel qu'il souhaite récupérer. L'objet du tunnel de récupération est seulement de récupérer l'ancien tunnel correspondant. Il y a une relation biunivoque entre tunnel de récupération et ancien tunnel récupéré.

Considérations sur l'établissement du tunnel de récupération :
- Un LCCE DOIT suivre les procédures décrites dans la [RFC2661] ou la [RFC3931] pour établir le tunnel de récupération.
- Le tunnel de récupération DOIT utiliser la même version de L2TP (et de procédures d'établissement) que celle utilisée pour l'ancien tunnel.
- La SCCRQ pour le tunnel de récupération DOIT inclure l'AVP Récupération de tunnel, définie au paragraphe 5.2, pour identifier l'ancien tunnel qui est à récupérer.
- Le tunnel de récupération NE DOIT PAS inclure l'AVP Capacité de reprise sur défaillance dans ses messages SCCRQ ou SCCRP.
- Un point d'extrémité NE DEVRAIT PAS envoyer de message autre que les suivants sur le tunnel de récupération : SCCRQ, SCCRP, SCCCN, StopCCN, HELLO, ZLB, et ACK [RFC3931].
- Un point d'extrémité NE DOIT PAS utiliser d'ancien identifiant de tunnel pour le tunnel de récupération. Les anciens tunnels DOIVENT être valides jusqu'à la conclusion du processus de récupération.
- Un point d'extrémité DOIT utiliser l'AVP Départage *(Tie Breaker)* (paragraphe 4.4.3 de la [RFC2661]) ou l'AVP Départage de connexion de contrôle (paragraphe 5.4.3 de la [RFC3931]) dans l'établissement du tunnel de récupération pour s'assurer qu'un seul tunnel de récupération (quand les deux points d'extrémité ont une reprise sur défaillance simultanée) est établi pour récupérer un ancien tunnel. Le tunnel qui gagne le départage est utilisé pour décider des valeurs de Ns et Nr suggérées sur le tunnel récupéré. Donc, le point d'extrémité qui perd le départage devrait rétablir les valeurs de Ns et Nr (paragraphe 3.2.2) comme si il était un point d'extrémité distant. L'Appendice B illustre le scénario de double reprise sur défaillance.
- Traitement de l'AVP Départage : la portée de l'action d'une AVP Départage doit être indépendante pour les tunnels de récupération et les tunnels non de récupération, et elle est définie comme suit :
  - o Quand l'AVP Départage est utilisée dans un tunnel non de récupération, la portée de l'action de l'AVP Départage DOIT seulement être au sein des tunnels non de récupération. Donc, perdre un départage contre un tunnel non de récupération NE DOIT PAS résulter en la terminaison d'un tunnel de récupération.
  - o Quand une AVP Départage est utilisée dans un tunnel de récupération, la portée de l'action de l'AVP Départage est de plus restreinte aux tunnels de récupération pour un seul tunnel à récupérer. Donc, une mise en œuvre DOIT appliquer le départage reçu dans un tunnel de récupération aux seuls tunnels qui sont a) des tunnels de récupération, et b) associés au même tunnel à récupérer. Cela NE DOIT PAS impacter le fonctionnement des tunnels de non récupération et des tunnels de récupération associés à d'autres anciens tunnels à récupérer.

Quand il obtient un SCCRQ avec une AVP Récupération de tunnel, un point d'extrémité valide l'identifiant de tunnel récupéré et l'identifiant de tunnel récupéré distant et répond par un SCCRP. Il DOIT terminer le tunnel de récupération si :
- l'identifiant de tunnel récupéré ou l'identifiant de tunnel récupéré distant est inconnu,
- le point d'extrémité actif ou distant (avant la reprise sur défaillance) n'avait pas indiqué qu'il était capable de reprise sur défaillance,
- la version L2TP du tunnel de récupération est différente de la version utilisée dans l'ancien tunnel.

Si le point d'extrémité distant accepte le SCCRQ, il DEVRAIT inclure l'AVP Séquence de contrôle suggérée, définie au paragraphe 5.3, dans le message SCCRP.

Considérations d'authentification :
- Pour authentifier un point d'extrémité homologue durant l'établissement de tunnel de récupération, un point d'extrémité DOIT suivre la procédure décrite au paragraphe 5.1.1 de la [RFC2661] ou au paragraphe 4.3 de la [RFC3931]. Il DOIT utiliser le même secret qu'utilisé pour authentifier l'ancien tunnel.
- Ne pas être capable de s'authentifier pourrait être une raison de terminer le tunnel de récupération.
- Pour les tunnels L2TPv3, un tunnel de récupération DOIT utiliser l'authentification de message de contrôle (c'est-à-dire, un échange de valeurs de noms occasionnels) comme décrit au paragraphe 4.3 de la [RFC3931], si l'ancien tunnel était configuré à faire l'authentification de message de contrôle. Un tunnel L2TPv3 récupéré DOIT remettre ses valeurs de nom occasionnel (les deux points d'extrémité) aux valeurs de nom occasionnel échangées dans le tunnel de récupération.

Pour quelque raison que ce soit, si le point d'extrémité de récupération ne peut pas établir le tunnel de récupération, il DOIT alors supprimer en silence l'ancien tunnel et les sessions qu'il contient, concluant à l'échec du processus de récupération.

Tout paquet de contrôle reçu sur le tunnel récupéré avant le rétablissement du canal de contrôle (paragraphe 3.2.2) DOIT être éliminé en silence.

3.2.2 Rétablissement du canal de contrôle

Le rétablissement du canal de contrôle permet que de nouveaux messages de contrôle soient envoyés et reçus sur le tunnel récupéré.

Procédure de rétablissement du canal de contrôle :

- Un point d'extrémité DEVRAIT purger la fenêtre d'émission/réception et rétablir les numéros de séquence du canal de contrôle (c'est-à-dire, les valeurs Ns et Nr) sur le tunnel récupéré. Le canal de contrôle sur le point d'extrémité de récupération est rétabli quand on obtient un SCCRP valide sur le tunnel de récupération, tandis que le canal de contrôle sur le point d'extrémité distant est rétabli quand on obtient un SCCCN valide sur le tunnel de récupération. Si le point d'extrémité de récupération n'a pas reçu l'AVP Séquence de contrôle suggérée (SCS, *Suggested Control Sequence*) dans le SCCRP, il DOIT alors rétablir les valeurs Ns et Nr à zéro. Si le point d'extrémité distant a opté pour ne pas envoyer l'AVP SCS, il DOIT alors remettre les valeurs Ns et Nr à zéro. L'un et l'autre point d'extrémité peut supprimer le tunnel de récupération après l'achèvement de la procédure de rétablissement du canal de contrôle.

- Un point d'extrémité DOIT empêcher l'établissement de nouvelles sessions jusqu'à ce qu'il ait supprimé (ou marquées pour suppression) les sessions qui n'étaient pas dans un état établi, c'est-à-dire, jusqu'à ce que l'étape I du paragraphe 3.3 soit achevée.

3.2.3 Rétablissement du canal de données

La procédure de rétablissement du canal de données n'est applicable que pour les sessions qui utilisent des numéros de séquence. Pour le canal de données L2TPv3, les termes Nr et Ns dans le présent document sont utilisés pour signifier respectivement "numéro de séquence attendu" et "numéro de séquence".

Procédure de rétablissement du canal de données :

- Le point d'extrémité de récupération règle la valeur Ns à zéro.

- Le point d'extrémité distant (homologue du point d'extrémité de récupération) continue d'utiliser les valeurs de Ns qu'il utilisait précédemment.

- Pour rétablir les valeurs de Nr durant une reprise sur défaillance, si un point d'extrémité reçoit 'n' paquets en désordre mais en séquence, il DOIT alors régler la valeur de Nr sur la base de la valeur de Ns des paquets entrants, comme suggéré dans l'Appendice C de la [RFC3931]. La valeur de 'n' DEVRAIT être configurable.

- Si un des points d'extrémité n'a pas la capacité (indiquée dans le bit 'D' de l'AVP Capacité de reprise sur défaillance) de rétablir la valeur de Nr, alors les canaux de données qui utilisent des numéros de séquence sont considérés comme non récupérables. Ces sessions DEVRAIENT être supprimées par le point d'extrémité de récupération en envoyant un message Notification de déconnexion d'appel (CDN, *Call-Disconnect-Notify*).

- Pour une défaillance du seul canal de données, les deux points d'extrémité PEUVENT utiliser le mécanisme d'interrogation/réponse d'état de session sur le canal de contrôle pour synchroniser l'état des sessions comme décrit au paragraphe 3.3.

3.3. Synchronisation d'état de session

Si une défaillance de canal de contrôle se produit quand une session est en cours d'établissement ou de suppression, il est alors possible qu'un point d'extrémité considère qu'une session est dans un état établi alors que son homologue considère que la même session n'existe pas. Deux de ces situations se présentent quand une défaillance se produit sur un point d'extrémité immédiatement après l'envoi :

- d'un message CDN qui n'est jamais arrivé à l'homologue,
- d'un message ICCN qui n'est jamais arrivé à l'homologue.

Le mécanisme suivant DOIT être utilisé pour identifier et supprimer les sessions qui existent sur un point d'extrémité, mais pas sur son homologue :

Étape I : pour une défaillance du canal de contrôle, après l'établissement du tunnel de récupération, les sessions qui n'étaient pas dans un état établi DOIVENT être supprimées en silence (c'est-à-dire, sans envoi de message CDN) par chaque point d'extrémité.

Étape II : les deux points d'extrémité PEUVENT identifier les sessions qui pourraient avoir été dans des états incohérents, peut-être sur la base de l'inactivité du canal de données. Les messages FSQ et FSR ont été introduits pour synchroniser l'état de session à tout moment pendant la vie d'une session entre deux points d'extrémité. Ces messages sont utilisés quand un point d'extrémité détermine ou suspecte d'une manière spécifique de la mise en œuvre que son état de session pourrait être incohérent avec celui de son homologue.

Étape III : un point d'extrémité envoie un message Interrogation de session de reprise sur défaillance (FSQ, *Failover Session Query*) pour interroger sur l'état des sessions tel que connu de son homologue. Un message FSQ contient une AVP État de session de reprise sur défaillance (FSS, *Failover Session State*) définie au paragraphe 5.4, pour chaque session sur laquelle il souhaite interroger. Plusieurs AVP FSS pourraient être incluses dans un message FSQ. Un message FSQ DOIT inclure au moins une AVP FSS. Un point d'extrémité PEUT envoyer un autre message FSQ avant d'obtenir une réponse à ses FSQ précédents.

Une incohérence sur l'existence d'une session durant une reprise sur défaillance pourrait résulter en ce qu'un point d'extrémité choisisse le même identifiant de session pour une nouvelle session. Dans une telle situation, il enverrait un ICRQ pour une session déjà établie. Donc, avant que toutes les sessions soient synchronisées en utilisant un mécanisme FSQ/FSR, si le point d'extrémité reçoit une ICRQ pour une session dans un état établi, il DOIT alors répondre à cet ICRQ par un CDN. Le message CDN doit régler une AVP Identifiant de session allouée/local ([RFC2661] paragraphe 4.4.4, [RFC3931] paragraphe 5.4.4) à son identifiant de session locale et supprimer la session qu'il considérait comme établie. L'utilisation de l'identifiant de session le plus récenmment utilisé pour les nouvelles sessions pourrait aider à réduire ce symptôme durant une reprise sur défaillance.

Quand un point d'extrémité reçoit un message FSQ, il DOIT s'assurer que pour chaque AVP FSS dans le message FSQ, il inclut une AVP FSS dans le message Réponse de session de reprise sur défaillance (FSR, *Failover Session Response*). Un point d'extrémité pourrait répondre à plusieurs FSQ en utilisant un message FSR, ou il pourrait répondre à une FSQ avec plusieurs FSR. Les FSS ne sont pas obligées d'avoir des réponses dans le même ordre que celui de leur réception. Pour chaque AVP FSS reçue dans des messages FSQ, un point d'extrémité DOIT valider l'identifiant de session distante et déterminer si il est apparié à l'identifiant de session spécifié dans le message. Si une AVP FSS n'est pas valide (c'est-à-dire, si la session n'existe pas ou est appariée à un identifiant de session différent) alors le champ Identifiant de session dans l'AVP FSS dans le FSR DOIT être réglé à zéro. Quand il se trouve que la session est appariée à un identifiant de session discordant, la session locale NE DOIT pas être supprimée, mais plutôt marquée comme périmée, pour être interrogée ultérieurement en utilisant un message FSQ. L'Appendice C présente un exemple de dialogue entre deux points d'extrémité avec des identifiants de session discordants.

Quand on répond à un FSQ par un message FSR, l'identifiant de session distant dans l'AVP FSS du message FSR est toujours réglé à la valeur reçue de l'identifiant de session dans l'AVP FSS du message FSQ.

Quand un point d'extrémité reçoit un message FSR, il DOIT utiliser pour chaque AVP FSS le champ Identifiant de session distant pour identifier la session locale et supprimer en silence (sans envoyer de message CDN) la session si l'identifiant de session dans l'AVP était à zéro. Autrement, il DOIT considérer que la session est dans un état établi et récupérée.

4. Nouveaux messages de commandes

Le présent document introduit deux nouveaux messages qui pourraient être envoyés sur une connexion de contrôle établie/récupérée.

4.1 Interrogation de session de reprise sur défaillance

Le message de contrôle Interrogation de session de reprise sur défaillance (SQ, *Failover Session Query*) est utilisé par un point d'extrémité durant le processus de récupération pour interroger l'état des diverses sessions. Il déclenche une réponse de l'homologue, qui contient l'état demandé des diverses sessions.

Ce message de contrôle est codé comme suit :

Identifiant de fabricant = 0 (IETF)
Type d'attribut = 21

Les AVP suivantes DOIVENT être présentes dans le message de contrôle FSQ :
Type de message
État de session de reprise sur défaillance

Les AVP suivantes PEUVENT être présentes dans le message de contrôle FSQ :
Valeur aléatoire
Résumé de message ([RFC3931] seulement pour les tunnels)

D'autres AVP NE DOIVENT PAS être envoyées dans ce message de contrôle et DEVRAIENT être ignorées à réception.

Le bit M sur l'AVP Type de message pour ce message de contrôle DOIT être réglé à 0.

4.2 Réponse de session de reprise sur défaillance

Le message de contrôle Réponse de session de reprise sur défaillance (FSR, *Failover Session Response*) est utilisé par un point d'extrémité durant le processus de récupération pour répondre avec l'état local des diverses sessions. Il est envoyé en réponse à un message FSQ. Un point d'extrémité PEUT choisir de répondre à un message FSQ avec plusieurs messages FSR.

Ce message de contrôle est codé comme suit :

Identifiant de fabricant = 0 (IETF)
Type d'attribut = 22

Les AVP suivantes DOIVENT être présentes dans le message de contrôle FSR ,:
Type de message
État de session de reprise sur défaillance

Les AVP suivantes PEUVENT être présentes dans le message de contrôle FSR :
Valeur aléatoire
Résumé de message ([RFC3931] seulement pour les tunnels)

D'autres AVP NE DOIVENT PAS être envoyées dans ce message de contrôle et DEVRAIENT être ignorées à réception.

Le bit M sur l'AVP Type de message pour ce message de contrôle DOIT être réglé à 0.

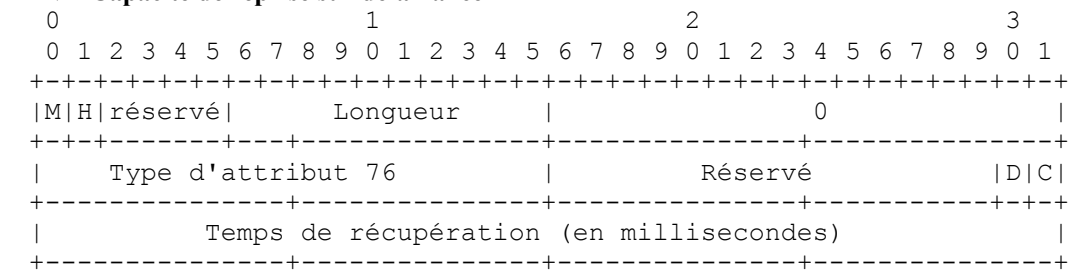
5. Nouvelles paires Attribut-Valeur

Les paragraphes qui suivent contiennent une liste de nouvelles AVP L2TP définies dans ce document.

5.1 AVP Capacité de reprise sur défaillance

L'AVP Capacité de reprise sur défaillance, type d'attribut 76, indique les capacités d'un point d'extrémité exigées pour le processus de récupération. Le format d'AVP est défini comme suit :

AVP Capacité de reprise sur défaillance



L'AVP PEUT être cachée (le bit H réglé à 0 ou 1). L'AVP n'est pas obligatoire (le bit M DOIT être réglé à 0).

Le bit C gouverne la capacité de reprise sur défaillance pour le canal de contrôle. Quand le bit C est établi, il indique que le point d'extrémité peut récupérer d'une défaillance du canal de contrôle en utilisant la procédure décrite au paragraphe 3.2.2. Quand le bit C n'est pas établi, il indique que le point d'extrémité ne peut pas récupérer d'une reprise sur défaillance du canal de contrôle. Dans ce cas, le bit D DOIT être établi. Noter qu'une reprise sur défaillance d'un canal de contrôle dans ce cas va être fatale pour le tunnel et tous les canaux de données associés.

Le bit D gouverne la capacité de reprise sur défaillance pour les canaux de données qui utilisent des numéros de séquence. Les canaux de données qui n'utilisent pas de numéros de séquence n'ont pas besoin d'aide pour récupérer d'une défaillance d'un canal de données.

Quand le bit D est établi, il indique que le point d'extrémité est capable de rétablir la valeur Nr des canaux de données en utilisant la procédure décrite au paragraphe 3.2.3 Rétablissement du canal de données.

Quand le bit D n'est pas établi, il indique que le point d'extrémité ne peut pas récupérer les canaux de données qui utilisent des numéros de séquence. En cas d'échec, ces canaux de données seront perdus.

L'AVP Capacité de reprise sur défaillance NE DOIT PAS être envoyée avec les bits C et D à zéro.

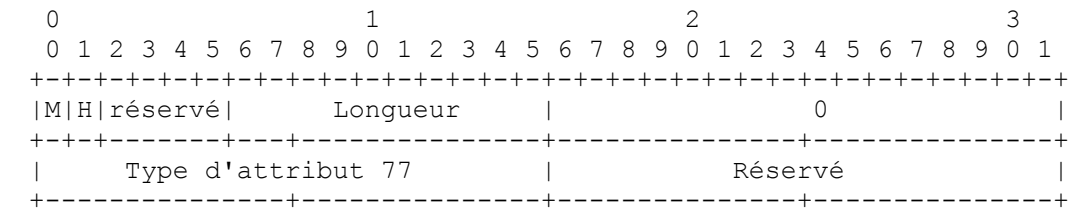
L'heure de réccupération, applicable seulement quand le bit C est établi, est la durée en millisecondes qu'un point d'extrémité demande à son homologue d'attendre avant de suposer que le processus de récupération a échoué. Ce temporisateur débute quand commence la temporisation de canal de contrôle d'un point d'extrémité (paragraphe 5.8 de la [RFC2661], paragraphe 4.2 de la [RFC3931]) et n'est pas arrêté (avant expiration) jusqu'à ce qu'un point d'extrémité réussisse à authentifier son homologue durant la récupération. Une valeur de zéro ne signifie pas qu'une reprise sur défaillance ne va pas se produire, cela signifie qu'aucun délai supplémentaire n'est demandé à l'homologue. Le temporisateur est aussi arrêté si l'homologue accuse réception d'un message de canal de contrôle dans la situation où il n'y a pas de reprise sur défaillance, mais où la perte du message de canal de contrôle était un phénomène temporaire.

Cette AVP NE DOIT PAS être incluse dans un message de contrôle autre que des messages SCCRQ et SCCRP.

5.2 AVP Récupération de tunnel

L'AVP Récupération de tunnel, type d'attribut 77, indique qu'un envoyeur voudrait récupérer le tunnel identifié dans cette AVP à cause d'une défaillance. Le format de l'AVP est défini comme suit :

AVP Récupération de tunnel pour tunnels L2TPv3 :



```
|                 Identifiant de tunnel de récupération         |
+---------------+---------------+---------------+---------------+
|            Identifiant de tunnel de récupération distant      |
+---------------+---------------+---------------+---------------+
```

AVP Récupération de tunnel pour tunnels L2TPv2 :

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|M|H|réservé|      Longueur     |               0               |
+-+-+-------+---+---------------+---------------+---------------+
|       Type d'attribut 77      |           Réservé             |
+---------------+---------------+---------------+---------------+
|         Réservé               |  ID de tunnel de récupération |
+---------------+---------------+---------------+---------------+
|         Réservé               | ID de tunnel de récup distant |
+---------------+---------------+---------------+---------------+
```

Cette AVP NE DOIT PAS être cachée (le bit H est réglé à 0). L'AVP est obligatoire (le bit M est établi à 1).

L'identifiant de tunnel de récupération code l'identifiant du tunnel local qu'un point d'extrémité veut récupérer. L'identifiant de tunnel de récupération distant code l'identifiant de tunnel distant qui correspond à l'ancien tunnel.

Cette AVP NE DOIT PAS être incluse dans un message de contrôle autre que SCCRQ lors de l'établissement d'un tunnel de récupération.

5.3 AVP Séquence de contrôle suggérée

L'AVP Séquence de contrôle suggérée (SCS, *Suggested Control Sequence*) type d'attribut 78, spécifie les valeurs de Ns et Nr pour le tunnel récupéré. Cette AVP est incluse dans un message SCCRP d'un tunnel de récupération par le point d'extrémité distant. Le format de l'AVP est défini comme suit :

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|M|H|réservé|      Longueur     |               0               |
+-+-+-------+---+---------------+---------------+---------------+
|       Type d'attribut 78      |            Réservé            |
+---------------+---------------+---------------+---------------+
|        Ns suggérée            |          Nr suggérée          |
+---------------+---------------+---------------+---------------+
```

Cette AVP PEUT être cachée (le bit H réglé à 0 ou 1) ; elle n'est pas obligatoire (le bit M est réglé à 0).

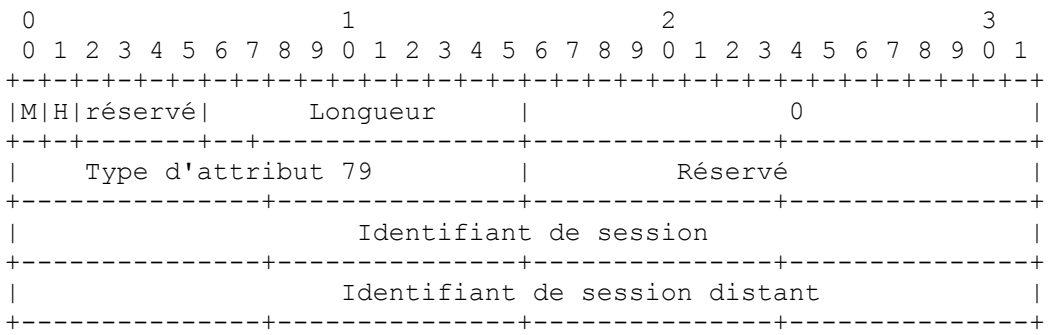
C'est une AVP facultative, qui suggère les valeurs de Ns et Nr à utiliser par le point d'extrémité de récupération. Si cette AVP est présente dans un message SCCRP durant l'établissement du tunnel de récupération, le point d'extrémité de récupération DOIT régler les valeurs Ns et Nr du tunnel récupéré aux valeurs suggérées respectives. Quand cette AVP n'est pas envoyée dans un SCCRP ou pas présente dans un SCCRP entrant, les valeurs Ns et Nr pour le tunnel récupéré sont réglées à zéro. L'utilisation de cette AVP aide à éviter l'interférence dans le canal de contrôle du tunnel récupéré avec les anciens paquets de contrôle.

Cette AVP NE DOIT PAS être incluse dans un autre message de contrôle que SCCRP lors de l'établissement d'un tunnel de récupération.

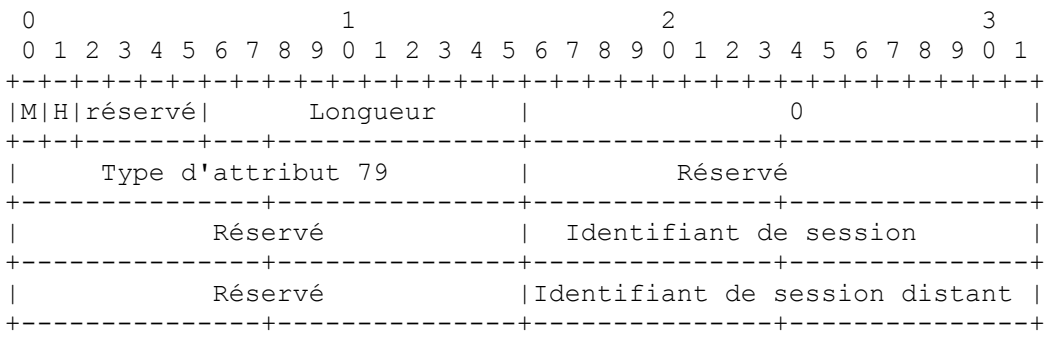
5.4 AVP État de session de reprise sur défaillance

L'AVP État de session de reprise sur défaillance (FSS, *Failover Session State* (FSS) type d'attribut 79, est utiliseé pour interroger l'homologue sur l'état d'une session afin de supprimer les sessions qui autrement resteraient dans un état indéfini après la reprise sur défaillance. Le format de l'AVP est défini comme suit :

Format de l'AVP FSS pour les sessions L2TPv3 :



Format de l'AVP FSS pour les sessions L2TPv3 :



Cette AVP PEUT être cachée (le bit H réglé à 0 ou 1). L'AVP est obligatoire (le bit M est réglé à 1).

L'identifiant de session identifie l'identifiant de session locale que l'envoyeur a alloué, pour laquelle il voudrait interroger l'état chez son homologue. Un identifiant de session distant est l'identifiant de session distant pour la même session.

Une AVP FSS NE DOIT PAS être utilisée dans un message autre que les messages FSQ et FSR.

6. Paramètres de configuration

Un point d'extrémité L2TP PEUT exposer les paramètres de configuration suivants pour les spécifier pour les connexions de contrôle :
- Capacité de reprise sur défaillance de canal de contrôle : AVP Capacité de reprise sur défaillance (paragraphe 5.1) bit C.
- Capacité de reprise sur défaillance de canal de données : AVP Capacité de reprise sur défaillance (paragraphe 5.1) bit D.
- Temps de récupéraation : AVP Capacité de reprise sur défaillance (paragraphe 5.1).

La MIB L2TP définie dans la [RFC3371] et la [L2TPv3-MIB], définit un certain nombre d'objets qui peuvent être utilisés pour surveiller l'état des nœuds L2TP, mais est rarement utilisée pour les besoins de configuration. Il est prévu que les paramètrers susmentionnés seront configurés en utilisant une interface de ligne de commande (CLI, *Command Line Interface*) ou autre mécanisme propriétaire.

Des notifications asynchrones pour les événements de reprise sur défaillance et de récupération peuvent être envoyés par les nœuds L2TP aux applications de gestion de réseau, mais la spécification du protocole et du format à utiliser pour ces notifications sort du domaine d'application de ce document.

7. Considérations relatives à l'IANA

Le présent document définit les valeurs suivantes allouées par l'IANA.

- Quatre paires de valeurs d'attribut de message de contrôle (paragraphe 10.1 de la [RFC3931]) :
Capacité de reprise sur défaillance : 76
Récupération de tunnel : 77
Séquence de contrôle suggérée : 78
État de session de reprise sur défaillance : 79

- Deux valeurs de type de message (Type d'attribut 0) (paragraphe 10.2 de la [RFC3931]):
Interrogation de session de reprise sur défaillance : 21
Réponse de session de reprise sur défaillance : 22

8. Considérations sur la sécurité

Une demande usurpée de reprise sur défaillance (SCCRQ avec AVP Récupération de tunnel) au nom d'un point d'extrémité pourrait causer la terminaison d'un canal de contrôle si les mesures d'authentification mentionnées au paragraphe 3.2.1 ne sont pas utilisées.

Même si les mesures d'authentification (décrites au paragraphe 3.2.1) étaient utilisées, il est encore possible d'apprendre une identité d'un tunnel opérationnel d'un point d'extrémité en produisant des demandes usurpées de reprise sur défaillance qui échouent à la procédure d'authentification. La probabilité de réussite avec une demande usurpée de reprise sur défaillance est de 1 sur (2^16 - 1) pour la [RFC2661] et de 1 sur (2^32 - 1) pour la [RFC3931]. L'identité découverte d'un tunnel opérationnel pourrait alors être utilisée pour envoyer des messages de contrôle pour une possible entrave à la connexion de contrôle. Normalement, les messages de contrôle qui sont en dehors de la fenêtre de réception du point d'extrémité sont éliminés. Cependant, si l'AVP Séquence de contrôle suggérée (paragraphe 5.3) n'est pas utilisée durant le processus réel de reprise sur défaillance, les numéros de séquence pourraient être remis à zéro, rendant ainsi la fenêtre de réception prévisible. Pour améliorer la sécurité dans de telles circonstances, un point d'extrémité peut être configuré avec l'ensemble des points d'extrémité de récupération possibles qui pourraient récupérer un tunnel, et utiliser l'AVP Séquence de contrôle suggérée lors de la récupération d'un tunnel.

9. Remerciements

Leo Huber a fourni des suggestions aidant à définir le concept de reprise sur défaillance. Mark Townsley, Carlos Pignataro, et Ignacio Goyret ont relu ce document et fourni de précieuses suggestions.

10. Contributeurs

Paul Howard, Juniper Networks
Vipin Jain, Riverstone Networks
Sam Henderson, Cisco Systems
Keyur Parikh, Harris Corporations

11. Références

11.1 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)

[RFC2661] W. Townsley, A. Valencia, A. Rubens, G. Pall, G. Zorn et B. Palter, "Protocole de tunnelage de couche 2 "L2TP"", *(P.S.)*

[RFC3931] J. Lau et autres, "Protocole de tunnelage de couche deux - version 3 (L2TPv3)", mars 2005. *(P.S.)*

11.2 Références pour information

[L2TPv3-MIB] Nadeau, T. et K. Koushik, "Layer Two Tunneling Protocol (version 3) Management Information Base", Travail en cours, août 2006.

[RFC3371] E. Caves, P. Calhoun, R. Wheeler, "Base de données d'informations de gestion du protocole de tunnelage de couche deux (L2TP)", août 2002. *(P.S.)*

[RFC5883] D. Katz, D. Ward, "Détection de transmission bidirectionnelle (BFD) pour chemins à plusieurs bonds", juin 2010. *(P.S.)*

Appendice A

La description ci-dessous montre le fonctionnement du protocole de reprise sur défaillance pour un exemple de tunnel. Le protocole de reprise sur défaillance n'empêche pas un point d'extrémité de récupérer plusieurs tunnels en parallèle. Il permet aussi à un point d'extrémité d'envoyer plusieurs FSQ, chacune incluant plusieurs AVP FSS, pour récupérer rapidement.

Négociation de capacité de récupération (paragraphe 3.1) :

Point d'extrémité **Homologue**

(tid alloué = x, capable de reprise sur défaillance)
SCCRQ ----------> validation de SCCRQ

(tid alloué = y, capable de reprise sur défaillance)
valider <---------- envoi de SCCRP, SCCRP, etc.

.... <après la création du tunnel, les sessions sont établies> ....

< Ce nœud échoue >

Le point d'extrémité de récupération établit le tunnel de récupération (paragraphe 3.2.1). Il initie l'établissement du tunnel de récupération pour l'ancien tunnel 'x' :

Point d'extrémité de récupération **Homologue**

(tid alloué = z, AVP Récupération)
SCCRQ ----------> Détecte la reprise sur défaillance
(tid de récupération = x, tid de récupération distant = y) valide la SCCRQ

(AVP Séquence de contrôle suggérée, Ns/Nr suggérées = 3/100)
valide <---------- SCCRP envoyée
SCCRP (tid de récupération = y, tid de récupération distant = x)
remet Ns = 3, Nr = 100 sur le tunnel récupéré

SCCCN ----------> valide et remet Ns = 100, Nr = 3 sur le tunnel récupéré
Termine le tunnel de récupération
tid = 'z'
StopCCN ----------> Nettoie 'w'

Les états de session sont synchronisés aux deux points d'extrémité qui peuvent envoyer des FSQ et supprimer les sessions périmées (paragraphe 3.3)

(AVP FSS pour les sessions s1, s2, s3..)
envoie FSQ ----------> calcule l'état des sessions dans FSQ

(AVP FSS pour les sessions s1, s2, s3...)
supprime les sessions périmées, si il en est <---------- envoie FSR

(AVP FSS pour les sessions s7, s8, s9...)

calcule l'état des sessions dans FSQ <---------- envoie FSQ

(FSS AVP pour les sessions s7, s8, s9...)
envoie FSR ----------> supprime les sessions périmées, si il en est

Appendice B

Cette section montre un exemple de dialogue pour illustrer un double échec de récupération. La différence notable, décrite au paragraphe 3.2.1, dans la procédure par rapport au scénario d'une seule reprise sur défaillance est l'utilisation d'un départage par un des points d'extrémité de récupération pour utiliser le tunnel de récupération établi par son homologue (aussi un point d'extrémité de récupération) comme un tunnel de récupération.

Point d'extrémité de récupération **Point d'extrémité de récupération**

(on suppose que l'ancien tid = A) (on suppose que l'ancien tid = B)

```
 AVP  Récupération = (A, B)
      SCCRQ     ----------------------------------+
  (avec AVP Départage (tunnel de récupération 'C') |
                                                   |
               AVP Récupération = (B, A)           |
   +-  SCCRQ valide <------------------------------ Envoi de SCCRQ
   |                 (tunnel de récupération 'D') |    (avec AVP Départage)
   |  Ce point d'extrémité perd                   |
   |  le départage ;                              |
   |  Élimine tunnel 'C'                          +--> SCCRQ valide
   |                                      Ce point d'extrémité gagne le départage ;
   |                                       il élimine le SCCRQ.
   |
   |              (peut inclure l'AVP SCS)
   +->Envoi de SCCRP ------------------------> Valide le SCCRP
                                        Rétablit 'B' ;
                                  Règle les valeurs Ns, Nr --+
                                                             |
                                                             |
                                                             |
      Valide le SCCN <---------------------- Envoi de SCCN -------+
      Rétablit 'A' ;
      Règle les valeurs Ns, Nr
```

Les FSQ et FSR pour l'ancien tunnel (A, B) sont échangées sur le tunnel récupéré par les deux points d'extrémité.

Appendice C

Une discordance d'identifiant de session pourrait n'être pas le résultat d'une défaillance d'un des points d'extrémité. Cependant, la procédure de récupération de session de reprise sur défaillance pourrait exacerber la situation, résultant en une discordance permanente des identifiants de session entre deux points d'extrémité. Le dialogue ci-dessous montre le comportement décrit au paragraphe 3.3, étape III pour traiter en douceur de telles situations.

Point d'extrémité de récupération **Point d'extrémité de récupération**
(on suppose une discordance) (on suppose une discordance)
Sid = A, Sid distant = B Sid = B, Sid distant = C
Sid = C, Sid distant = D

AVP FSS (A, B)
Envoi de FSQ ----------> Aucune paire (B, A) n'existe ; il existe plutôt (B, C).
Si il supprime B l'homologue nesait pas si C est par ailleurs périmé.

Si il marque que B est périmé et interroge l'état de session via FSQ, C va être supprimé de l'autre côté.

```
                                AVP FSS (0, A)
Supprime A <---------------------------------------------------------- envoi d'une FSR

                            ... quelques temps plus tard ...

                                AVP FSS (B, C)
Pas de (C,B) <------------------------------------------------------------------- envoi d'une FSQ
Marque C périmé

                                AVP FSS (0, B)
Envoi de FSR ----------------------------------------------------------> supprime B
```

Informations sur les auteurs

Vipin Jain
Riverstone Networks
5200 Great America Parkway
Santa Clara, CA 95054
mél : vipinietf@yahoo.com

Paul W. Howard
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
mél : phoward@juniper.net

Sam Henderson
Cisco Systems
7025 Kit Creek Rd.
PO Box 14987
Research Triangle Park, NC 27709
mél : samh@cisco.com

Keyur Parikh
Harris Corporation
4393 Digitalway
Mason, OH 45040
mél : kparikh@harris.com

Déclaration complète de droits de reproduction

Remerciement
Le financement de la fonction d'édition des RFC est actuellement fourni par la Internet Society.