Groupe de travail Réseau
Request for Comments : 5042
Catégorie : Sur la voie de la normalisation
Traduction Claude Brière de L'Isle

J. Pinkerton, Microsoft Corporation
E. Deleganes, indépendante
octobre 2007

Sécurité du protocole de placement direct des données (DDP) / protocole d'accès direct à une mémoire distante (RDMAP)

Résumé
Le présent document analyse les problèmes de sécurité autour de la mise en œuvre et l'utilisation du protocole de placement direct de données (DDP, *Direct Data Placement Protocol*) et du protocole d'accès direct à une mémoire distante (RDMAP, *Remote Direct Memory Access Protocol*). Il définit d'abord un modèle architectural pour une carte d'interface réseau RDMA (RNIC, *RDMA Network Interface Card*) qui peut mettre en œuvre DDP ou RDMAP et DDP. Le document passe en revue diverses attaques contre les ressources définies dans le modèle architectural et les contre-mesures qui peuvent être utilisées pour protéger le système. Les attaques sont groupées selon qu'elles peuvent être atténuées en utilisant des canaux de communication sûrs à travers le réseau, les attaques provenant d'homologues distants, et les attaques provenant des homologues locaux. Les catégories d'attaques incluent l'usurpation d'identité, l'altération, la divulgation d'informations, le déni de service, et l'élévation de privilège.

Table des Matières

1. Introduction

RDMA permet de nouveaux niveaux de souplesse dans la communication entre deux parties comparé à la pratique courante du réseautage conventionnel (par exemple, un modèle fondé sur le flux ou un modèle de datagrammes). Cette souplesse amène de nouveaux problèmes de sécurité qui doivent être bien compris quand on conçoit des protocoles de couche supérieure (ULP, *Upper Layer Protocol*) qui utilisent RDMA et quand on met en œuvre des contrôleurs d'interface réseau à capacité RDMA (RNIC). Noter que pour les besoins de cette analyse de sécurité, un RNIC peut mettre en œuvre RDMAP [RFC5040] et DDP [RFC5041], ou juste DDP. Aussi, un ULP peut être une application ou une bibliothèque de logiciels.

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]. Additionally, the security terminology defined in [RFC4949] is utilisé in this specification.

Le document développe d'abord un modèle architectural qui est pertinent pour l'analyse de la sécurité. La Section 2 détaille les composants, ressources, et propriétés de système qui peuvent être attaquées. Le document utilise le terme d'homologue local pour représenter la mise en œuvre de protocole RDMA/DDP sur l'extrémité locale d'un flux (mis en œuvre avec un protocole de transport, tel que de la [RFC0793] ou de la [RFC4960]). Le protocole de couche supérieure (ULP, *Upper-Layer-Protocol*) local est utilisé pour représenter la couche d'application ou de logiciel médiateur au dessus de l'homologue local. Le document ne tente pas de différencier entre un homologue distant et un ULP distant (une mise en œuvre de protocole RDMA/DDP sur l'extrémité distante d'un flux par opposition à l'application sur l'extrémité distante) pour plusieurs raisons : souvent, la source de l'attaque est difficile à connaître avec certitude et, sans considération de la source, les atténuations requises de l'homologue local ou de l'ULP local sont les mêmes. Donc, le document se réfère de façon générique à un homologue distant plutôt que d'essayer de mieux préciser l'attaquant. Le document définit ensuite quelles ressources un ULP local peut partager à travers les flux et quelles ressources l'ULP local peut partager avec l'homologue distant à travers les flux à la Section 3.

Le partage intentionnel de ressources entre plusieurs flux peut impliquer un certain niveau de confiance entre les flux. Cependant, certains types de partage de ressources subissent des attaques non mitigées contre la sécurité, ce qui obligerait à ne pas partager un type spécifique de ressources sauf si il y a un certain niveau de confiance entre les flux qui partagent les ressources.

Le présent document définit un nouveau terme, "confiance mutuelle partielle", pour traiter ce concept :
Confiance mutuelle partielle : une collection de flux RDMAP/DDP, représentant les points d'extrémité local et distant du flux qui sont d'accord pour supposer que les flux de la collection ne vont pas effectuer d'attaques malveillantes contre un des autres flux de la collection.

Les ULP ont le contrôle explicite de quelle collection de points d'extrémité est dans une collection de confiance mutuelle partielle grâce aux outils discutés dans l'Appendice C, "Taxonomie de la confiance partielle".

Une relation d'homologues sans confiance est appropriée quand un ULP souhaite s'assurer qu'il va être robuste et non compromis même en face d'une attaque délibérée par l'homologue. Par exemple, un seul ULP qui prend concurramment en charge plusieurs flux sans relation (par exemple, un serveur) va probablement traiter chacun de ses homologues comme un homologue non de confiance. Pour une collection de flux qui partagent une confiance mutuelle partielle, l'hypothèse est que tout flux qui n'est pas dans la collection n'est pas de confiance. Pour l'homologue qui n'est pas de confiance, une brève liste de capacités est donnée à la Section 4.

Le reste du document se concentre sur l'analyse des attaques et la recommandation d'atténuations spécifiques des attaques. Les attaques sont catégorisées en attaques atténuées par la sécurité de bout en bout, les attaques initiées par les homologues distants, et les attaques initiées par les homologues locaux. Pour chaque attaque, les contre-mesures possibles sont examinées.

Les ULP au sein d'un hôte sont divisés en deux catégories - privilégiés et non privilégiés. Les deux types d'ULP peuvent envoyer et recevoir des données et demander des ressources. Les différences clés entre les deux sont :
L'ULP privilégié est de confiance pour le système local pour ne pas attaquer par malveillance l'environnement de fonctionnement, mais il n'est pas de confiance pour optimiser globalement l'allocation de ressources. Par exemple, l'ULP privilégié pourrait être un ULP du noyau ; donc, le noyau a probablement effectué d'une certaine façon un contrôle de

sécurité sur l'ULP avant de lui permettre de s'exécuter. Les capacités d'un ULP non privilégié sont un sous ensemble logique de celles de l'ULP privilégié. Il est supposé par le système local qu'un ULP non privilégié n'est pas de confiance. Toutes les interactions d'un ULP non privilégié avec le moteur RNIC qui pourraient affecter les autres ULP doivent être faites à travers un intermédiaire de confiance qui peut vérifier les demandes de l'ULP non privilégié.

Les appendices fournissent des résumés ciblés de la présente spécification. L'Appendice A, "Problèmes d'ULP pour les protocoles client/serveur RDDP", se concentre sur la mise en œuvre des protocoles client/serveur traditionnels. L'Appendice B, "Résumé des exigences de mise en œuvre de RNIC et d'ULP", récapitule toutes les exigences normatives de la présente spécification. L'Appendice C, "Taxonomie de la confiance partielle", fournit un modèle abstrait pour catégoriser les limites de confiance.

Si une mise en œuvre de protocole RDMAP/DDP utilise les atténuations recommandées dans le présent document, cette mise en œuvre ne devrait pas subir de vulnérabilités de sécurité supplémentaires, au delà de celles d'une mise en œuvre du protocole de transport (c'est-à-dire, TCP ou SCTP) et des protocoles derrière lui (par exemple, IP) sans RDMAP/DDP.

2. Modèle architectural

Cette Section décrit un modèle de référence architectural de RDMA qui est utilisé lorsque les questions de sécurité sont examinées. Elle introduit les composants du modèle, les ressources qui peuvent être attaquées, les types d'interactions possibles entre composants et ressources, et les propriétés du système qui doivent être préservées.

La Figure 1 montre les composants de l'architecture et les interfaces où de potentielles attaques contre la sécurité pourraient être lancées. Les attaques externes peuvent être injectées dans le système à partir d'un ULP qui se tient au dessus de l'interface de RNIC ou du réseau. L'intention ici est de décrire les composants et capacités de haut niveau qui affectent l'analyse des menaces, et non de se concentrer sur des options spécifiques de la mise en œuvre. Noter aussi que le modèle architectural est une abstraction, et une mise en œuvre réelle peut choisir de subdiviser ses composants selon des lignes de frontière différentes de celles définies ici. Par exemple, le gestionnaire de ressource privilégié peut être partiellement ou complètement encapsulé dans l'ULP privilégié. Néanmoins, on suppose que l'analyse de sécurité des menaces et contre-mesures potentielles s'applique. Noter que le modèle ci-dessous est dérivé de plusieurs mises en œuvre spécifiques de RDMA. Parmi elles, [VERBS-RDMAC], [VERBS-RDMAC-Overview], et [INFINIBAND].

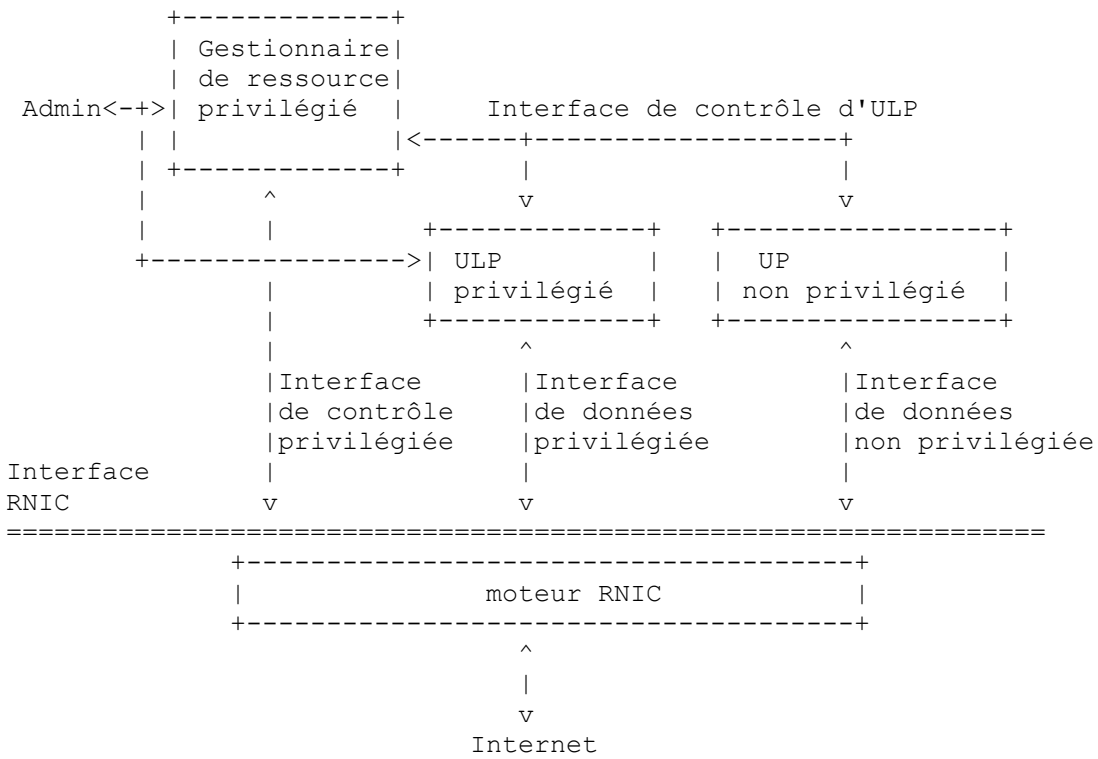

Figure 1 : Modèle de sécurité RDMA

2.1 Composants

Les composants montrés à la Figure 1 : Modèle de sécurité RDMA sont :

* Moteur de contrôleur d'interface réseau RDMA (RNIC) : c'est le composant qui met en œuvre le protocole RDMA et/ou le protocole DDP.

* Gestionnaire de ressource privilégié : c'est le composant responsable de la gestion et de l'allocation des ressources associées au moteur RNIC. Le gestionnaire de ressources n'envoie ni ne reçoit de données. Noter que si le gestionnaire de ressources est un composant indépendant, si il fait partie du RNIC, ou si il fait partie de l'ULP dépend de la mise en œuvre.

* ULP privilégié : voir à la Section 1, "Introduction", la définition de l'ULP privilégié. L'infrastructure de l'hôte local peut permettre à l'ULP privilégié de transposer une mémoire tampon de données directement du moteur RNIC à l'hôte à travers l'interface de RNIC, mais elle ne permet pas à l'ULP privilégié de consommer directement les ressources du moteur RNIC.

* ULP non privilégié : voir à la Section 1, "Introduction", la définition d'un ULP non privilégié. Un objectif de conception des protocoles DDP et RDMAP est de permettre, sous certaines conditions, aux ULP non privilégiés d'envoyer et recevoir des données directement de/vers le moteur RDMA sans intervention du gestionnaire de ressource privilégié, tout en s'assurant que l'hôte reste sûr. Donc, un des buts principaux du présent document est d'analyser ce modèle d'usage pour l'application qui est exigée dans le moteur RNIC pour s'assurer que le système reste sûr.

DDP fournit deux mécanismes pour transférer les données :
* transfert de données non étiqueté - la charge utile entrante consomme simplement la première mémoire tampon dans une file d'attente de mémoires tampon qui sont dans l'ordre spécifié par l'homologue receveur (couramment appelée la file d'attente de réception), et
* transfert de données étiqueté - l'homologue qui transmet la charge utile déclare explicitement quelle mémoire tampon de destination est ciblée, par l'utilisation d'une STag. Les transferts fondés sur la STag permettent à l'ULP receveur d'être indifférent à l'ordre (ou aux messages) dans lequel l'homologue opposé a envoyé les données, ou dans quel ordre les paquets sont reçus.

Les deux mécanismes de transfert de données sont aussi activés par RDMAP, avec une sémantique de contrôle supplémentaire. Normalement, le transfert de données étiqueté peut être utilisé pour le transfert de charge utile, tandis que le transfert de données non étiqueté est mieux utilisé pour les messages de contrôle. Cependant, chaque protocole de couche supérieure peut déterminer l'utilisation optimale des messages étiquetés et non étiquetés pour eux-mêmes. Voir dans la [RFC5045] plus d'informations sur l'applicabilité pour les deux mécanismes de transfert.

Pour DDP, les deux formes correspondent respectivement aux messages DDP étiqueté et non étiqueté. Pour RDMAP, les deux formes correspondent aux messages de type Send et aux messages RDMA (RDMA Read ou RDMA Write), respectivement.

Les interfaces d'hôte qui pourrait être concernées incluent :

* interface de contrôle privilégiée - un gestionnaire de ressource privilégié utilise l'interface de RNIC pour allouer et gérer les ressources du moteur RNIC, contrôler l'état au sein du moteur RNIC, et surveiller divers événements provenant du moteur RNIC. Il utilise aussi cette interface pour agir comme mandataire pour certaines opérations qu'un ULP non privilégié peut exiger (après avoir effectué les contre-mesures appropriées).

* interface de contrôle d'ULP - un ULP utilise cette interface avec le gestionnaire de ressource privilégié pour allouer les ressources au moteur RNIC. Le gestionnaire de ressource privilégié met en œuvre des contre-mesures pour s'assurer que, si l'ULP non privilégié lance une attaque, il peut empêcher l'attaque d'affecter les autres ULP.

* interface de transfert de données non privilégiée - un ULP non privilégié utilise cette interface pour initier et vérifier l'état des opérations de transfert des données.

* interface de transfert de données privilégiée - un sur-ensemble de la fonction fournie par l'interface de transfert de données non privilégiée. Il est permis à l'ULP de manipuler directement les ressources de transposition du moteur RNIC pour transposer une STag en une mémoire tampon de données d'ULP.

Si des messages de contrôle Internet, comme ICMP, ARP, RIPv4, etc. sont traités par le moteur RNIC, les analyses de

menaces pour ces protocoles sont aussi applicables, mais sortent du domaine d'application de ce document.

2.2 Ressources

Ce paragraphe décrit les ressources principales du moteur RNIC qui pourraient être affectées par une attaque. Pour RDMAP, toutes les ressources définies s'appliquent. Pour DDP, toutes les ressources s'appliquent sauf la file d'attente de RDMA Read.

2.2.1 Mémoire de contexte de flux

Les informations d'état pour chaque flux sont conservées en mémoire, qui pourrait être située dans un certain nombre d'endroits - sur le NIC, dans la RAM rattachée au NIC, dans la mémoire de l'hôte, ou dans une combinaison des trois, selon la mise en œuvre.

La mémoire de contexte de flux inclut l'état associé aux mémoires tampon de données. Pour les mémoires tampon étiquetées, cela inclut comment s'inter relatent les noms de STag, les mémoires tampon de données, et les tableaux de traduction de page (voir au paragraphe 2.2.3). Elle inclut aussi la liste des mémoires tampon de données non étiquetées envoyée pour la réception des messages non étiquetés (couramment appelée la file d'attente de réception) et une liste des opérations à effectuer pour envoyer des données (couramment appelée la file d'attente d'envoi).

2.2.2 Mémoires tampon de données

Comme mentionné précédemment, il y a deux façons différentes d'exposer les mémoires tampon de données d'un ULP local pour le transfert des données : le transfert de données non étiqueté, où une mémoire tampon peut être exposée pour recevoir des messages RDMAP de type Send (autrement dit, des messages DDP non étiquetés) sur une file d'attente DDP zéro, ou le transfert de données étiqueté, où la mémoire tampon peut être exposée pour l'accès à distance par des STag (autrement dit, des messages DDP étiquetés). Cette distinction est importante parce que les attaques et les contre-mesures utilisées pour protéger contre l'attaque sont différentes selon la méthode pour exposer la mémoire tampon au réseau.

Pour les besoins de la discussion de la sécurité, pour le transfert de données étiqueté, une seule mémoire tampon de données logique est exposée avec une seule STag sur un flux donné. Les mises en œuvre réelles peuvent prendre en charge des capacités de dispersion/rassemblement pour permettre que plusieurs mémoires tampon de données physiques soient accédées avec une seule STag, mais du point de vue de l'analyse des menaces, on suppose qu'une seule STag permet l'accès à une seule mémoire tampon de données logique.

Dans tous les cas, il est de la responsabilité du gestionnaire de ressource privilégié de s'assurer qu'aucune STag ne peut être créée qui expose une mémoire que le consommateur n'avait pas autorité à exposer.

Une mémoire tampon de données a des droits d'accès spécifiques. L'ULP local peut contrôler si une mémoire tampon de données est exposée seulement en local, ou en accès local et à distance, et allouer des privilèges d'accès spécifiques (lecture, écriture, lecture et écriture) flux par flux.

Pour DDP, quand une STag est annoncée, l'homologue distant va probablement donner des droits d'accès en écriture aux données (autrement, l'annonce n'aurait pas grand sens). Pour RDMAP, quand un ULP annonce une STag, il peut activer les droits d'accès "écriture seule", "lecture seule", ou "écriture et lecture".

De façon similaire, certains ULP peuvent souhaiter fournir une seule mémoire tampon avec différents droits d'accès sur la base du flux. Par exemple, certains flux peuvent avoir l'accès en lecture seule, certains peuvent avoir l'accès en lecture et écriture à distance, tandis que pour d'autres flux, l'accès est seulement permis à l'ULP/homologue local.

2.2.3 Tableaux de traduction de page

Les tableaux de traduction de page sont les structures utilisées par le RNIC pour être capable d'accéder à la mémoire d'ULP pour des opérations de transfert des données. Bien que ces structures soient appelées des tableaux de traduction de "page", elles peuvent ne pas référencer du tout de page - conceptuellement, elles sont utilisées pour transposer une représentation d'espace d'adresses d'ULP (par exemple, une adresse virtuelle) d'une mémoire tampon en les adresses physiques qui sont utilisées par le moteur RNIC pour déplacer les données. Si, sur un système spécifique, une transposition n'est pas utilisée, alors un sous ensemble des attaques examinées peut être approprié. Noter que le tableau de traduction de page peut ou non

être une ressource partagée.

2.2.4 Domaine de protection (PD)

Un domaine de protection (PD) est une construction locale de la mise en œuvre de RDMA, et qui n'est jamais visible sur le réseau. Les domaines de protection sont alloués aux trois ressources concernées - mémoire de contexte de flux, STag associée aux entrées de tableau de traduction de page, et mémoires tampon de données. Une mise en œuvre correcte d'un domaine de protection exige que les ressources qui appartiennent à un certain domaine de protection ne puissent pas être utilisées sur une ressource appartenant à un autre domaine de protection, parce que l'appartenance à un domaine de protection est vérifiée par le RNIC avant d'effectuer toute action impliquant une telle ressource. Les domaines de protection sont donc utilisés pour s'assurer qu'une STag peut seulement être utilisée pour accéder à une mémoire tampon de données associée sur un ou plusieurs flux qui sont associés au même domaine de protection que la STag spécifique.

Si une mise en œuvre choisit de ne pas partager les ressources entre les flux, il est recommandé que chaque flux soit associé à son propre domaine de protection unique. Si une mise en œuvre choisit de permettre le partage de ressources, il est recommandé que le domaine de protection soit limité à la collection de flux qui ont une confiance mutuelle partielle avec chaque autre.

Noter qu'un ULP (privilégié ou non privilégié) peut éventuellement avoir plusieurs domaines de protection. Cela pourrait être utilisé, par exemple, pour s'assurer que plusieurs clients d'un serveur n'ont pas la capacité de corrompre les autres. Le serveur va allouer un domaine de protection par client pour assurer que les ressources couvertes par le domaine de protection ne pourront pas être utilisées par un autre client (qui n'est pas de confiance).

2.2.5 Espace de noms et portée de STag

La spécification DDP définit un espace de noms de 32 bits pour la STag. Les mises en œuvre peuvent varier en termes de nombre réel de STag prises en charge. Dans tous les cas, c'est une ressource imitée qui peut être attaquée. Selon les algorithmes d'allocation d'espace de noms de STag, l'espace de noms réel à attaquer peut être significativement moindre que 2^32.

La portée d'une STag est l'ensemble de flux DDP/RDMAP sur lesquels la STag est valide. Si une STag est valide sur un flux DDP/RDMAP particulier, alors ce flux peut modifier la mémoire tampon, sous réserve des droits d'accès que le flux a pour la STag (voir au paragraphe 2.2.2, "Mémoires tampon de données", des informations supplémentaires).

L'analyse présentée dans ce document suppose deux mécanismes pour limiter la portée des flux pour lesquels la STag est valide :

* portée de domaine de protection : la STag est valide si elle est utilisée sur tout flux au sein d'un domaine de protection spécifique, et invalide si elle est utilisée sur tout flux non membre du domaine de protection.
* portée d'un seul flux : la STag est valide sur un seul flux, sans considération de ce qu'est l'association du flux à un domaine de protection. Si elle est utilisée sur un autre flux, elle est invalide.

2.2.6 Files d'attente d'achèvement

Les files d'attente d'achèvement (CQ, *Completion Queue*) sont utilisées dans ce document pour représenter conceptuellement comment le moteur RNIC notifie à l'ULP l'achèvement de la transmission des données, ou l'achèvement de la réception de données à travers l'interface de transfert de données (spécifiquement pour le transfert de données non étiquetées ; le transfert de données étiqueté ne peut pas causer la survenance d'un achèvement). Parce qu'il pourrait y avoir de nombreuses transmissions ou réceptions en cours à tout moment, les achèvements sont modélisés comme une file d'attente plutôt que comme un seul événement. Une mise en œuvre peut aussi utiliser la file d'attente d'achèvement pour notifier à l'ULP d'autres activités ; par exemple, l'achèvement d'une transposition d'une STag en une mémoire tampon d'ULP spécifique. Les files d'attente d'achèvement peuvent être partagées par un groupe de flux, ou peuvent être conçues pour traiter un trafic de flux spécifique. Limiter l'association de files d'attente d'achèvement à un, ou à un petit nombre, de flux RDMAP/DDP peut prévenir plusieurs formes d'attaques en limitant fortement la portée de l'effet de l'attaque.

Certaines mises en œuvre peuvent permettre que cette file d'attente soit manipulée directement par les ULP non privilégiés et privilégiés.

2.2.7 File d'attente d'événement asynchrones

La file d'attente d'événement asynchrone est une file d'attente du RNIC au gestionnaire de ressource privilégié de taille limitée. Elle est utilisée par le RNIC pour notifier à l'hôte les divers événements qui pourraient exiger une action de gestion, incluant des violations de protocole, des changements d'état de flux, des erreurs de fonctionnement local, des marques de faible niveau sur les files d'attente de réception, et d'autres événements possibles.

La file d'attente d'événement asynchrone est une ressource qui peut être attaquée parce que des homologues locaux ou distants et/ou les ULP peuvent causer la survenance d'événements qui ont un potentiel de débordement de la file d'attente.

Noter qu'une mise en œuvre est libre d'utiliser les fonctions de la file d'attente d'événement asynchrone de diverses façons, incluant plusieurs files d'attente ou même de simples rappels. Toutes les vulnérabilités identifiées sont destinées à s'appliquer, sans considération de la mise en œuvre de la file d'attente d'événement asynchrone. Par exemple, une fonction de rappel peut être vue simplement comme une très courte file d'attente.

2.2.8 File d'attente de demandes RDMA Read

La file d'attente de demandes RDMA Read est la mémoire qui contient les informations d'état pour un ou plusieurs messages de demande RDMA Read qui sont arrivés, mais pour lesquels les messages de réponse RDMA Read n'ont pas encore été complètement envoyés. Parce que potentiellement plus d'une demande RDMA Read peut être en instance à un moment donné, la mémoire est modélisée comme une file d'attente de taille limitée. Certaines mises en œuvre peuvent activer le partage d'une seule file d'attente de demandes RDMA Read à travers plusieurs flux.

2.3 Interactions avec le RNIC

Avec les ressources et interfaces de RNIC définies, il est maintenant possible d'examiner les interactions prises en charge par les interfaces fonctionnelles génériques de RNIC à travers chacune des trois interfaces : interface de contrôle privilégiée, interface de données priviligiée, et interface de données non privilégiée. Comme mentionné au paragraphe 2.1, "Composants", il y a deux mécanismes de transfert de donnés à examiner, le transfert de données non étiqueté et le transfert de données étiqueté.

2.3.1 Sémantique d'interface de contrôle privilégiée

De façon générale, l'interface de contrôle privilégiée contrôle l'allocation, la désallocation, et l'initialisation des ressources globales de RNIC. Cela inclut l'allocation et la désallocation de la mémoire de contexte de flux, les tableaux de traduction de page, les noms de STag, les files d'attente d'achèvement, les files d'attente de demandes RDMA Read, et les files d'attente d'événement asynchrone.

L'interface de contrôle privilégiée est aussi normalement utilisée pour gérer les ressources d'ULP non privilégié pour l'ULP non privilégié (et éventuelement aussi pour l'ULP privilégié). Cela inclut l'initialisation et la suppression de ressource de tableau de traduction de page, et la gestion d'événements de RNIC (éventuellement de gérer tous les événements pour la file d'attente d'événement asynchrone).

2.3.2 Sémantique d'interface de contrôle non privilégiée

L'interface de données non privilégiée permet le transfert de données (émission et réception) mais ne permet pas l'initialisation des ressources de tableau de traduction de page. Cependant, une fois que les ressources de tableau de traduction de page ont été initialisées, l'interface peut permettre qu'une transposition de STag spécifique soit activée et désactivée en communiquant directement avec le RNIC, ou créer une transposition de STag pour une mémoire tampon qui a été précédemment initialisée dans le RNIC.

Pour RDMAP, les données d'ULP peuvent être envoyées par un des mécanismes de transfert de données précédemment décrits : transfert de données non étiqueté ou transfert de données étiqueté. Deux mécanismes RDMAP de transfert de données sont définis, un en utilisant le transfert de données non étiqueté (messages de type Send) et un en utilisant le transfert de données étiqueté (réponses RDMA Read et RDMA Write). La réception des données d'ULP à travers RDMAP peut être faite en recevant les messages de type Send dans des mémoires tampon qui ont été postées dans la file d'attente de réception ou une file d'attente de réception partagée. Donc, une file d'attente de réception ou une file d'attente de réception partagée peut seulement être affectée par un transfert de données non étiqueté. La réception des données peut aussi être faite en recevant des messages de réponse RDMA Write et RDMA Read dans des mémoires tampon qui ont été

précédemment exposées pour un accès en écriture externe par l'annonce d'une STag (c'est-à-dire, un transfert de données étiqueté). De plus, pour causer le tirage (lecture) des données d'ULP à travers le réseau, RDMAP utilise un message de demande RDMA Read (qui contient seulement des informations de contrôle RDMAP nécessaires pour accéder à la mémoire tampon d'ULP à lire) pour causer la génération d'un message de réponse RDMA Read contenant les données d'ULP.

Pour DDP, transmettre des données signifie envoyer des messages DDP étiquetés ou non étiquetés. Pour la réception des données, DDP peut recevoir des messages non étiquetés dans des mémoires tampon qui ont été postées sur la file d'attente de réception ou file d'attente de réception partagée. Il peut aussi recevoir des messages DDP étiquetés dans des mémoires tampon qui ont été précédemment exposées pour un accès en écriture externe par l'annonce d'une STag.

L'achèvement de la transmission ou réception de données entraine généralement d'informer l'ULP de l'achèvement du travail en plaçant les informations d'achèvement sur la file d'attente d'achèvement. Pour la réception des données, seulement un transfert de données non étiquetées peut causer la mise des informations d'achèvement dans la file d'attente d'achèvement.

2.3.3 Sémantique d'interface de contrôle privilégiée

La sémantique d'interface de données privilégiée est un sur ensemble de la sémantique de transfert de données non privilégiées. L'interface peut faire tout ce qui est défini au paragraphe précédent, aussi bien que créer/supprimer la mémoire tampon directement en transposition de STag. Cela entraine généralement une initialisation ou une suppression de l'état de tableau de traduction de page dans le RNIC.

2.3.4 Initialisation de structures de données de RNIC pour un transfert de données

L'initialisation de la transposition entre une STag et une mémoire tampon de données peut être vue abstraitement comme deux opérations séparées :

a. l'initialisation des entrées de tableau de traduction de page allouées avec la localisation de la mémoire tampon de données, et
b. l'initialisation d'une transposition d'un nom de STag alloué en un ensemble d'entrées ou entrées partielles de tableau de traduction de page.

Noter qu'une mise en œuvre peut ne pas avoir de tableau de traduction de page (c'est-à-dire, elle peut prendre en charge une transposition directe entre une STag et une mémoire tampon de données). Si il n'y a pas de tableau de traduction de page, les attaques fondées sur le changement de son contenu ou sur l'épuisement de ses ressources ne sont alors pas possibles.

L'initialisation du contenu du tableau de traduction de page peut être faite par l'ULP privilégié ou par le gestionnaire de ressource privilégié en tant que mandataire de l'ULP non privilégié. Par définition, l'ULP non privilégié n'est pas de confiance pour manipuler directement le tableau de traduction de page. En général, le problème est que l'ULP non privilégié peut essayer d'initialiser malicieusement le tableau de traduction de page pour accéder à une mémoire tampon pour laquelle il n'a pas de permission.

L'algorithme exact d'allocation de ressource pour le tableau de traduction de page sort du domaine d'application du présent document. Elles peuvent être allouées pour une mémoire tampon de données spécifique, ou comme un réservoir de ressources à consommer par potentiellement plusieurs mémoires tampon de données, ou être gérées d'une autre façon. Le présent document tente de s'abstraire des problèmes dépendants de la mise en œuvre, et les groupe en problèmes de sécurité de niveau supérieur, comme la privation de ressources et le partage de ressources entre les flux.

Le problème suivant est comment un nom de STag est associé à une mémoire tampon de données. Pour le cas d'une mémoire tampon de données non étiquetée (c'est-à-dire, un transfert de données non étiqueté) il n'y a pas de transposition visible au niveau du réseau entre une STag et la mémoire tampon de données. Noter qu'il peut, en fait, y avoir une STag qui représente la mémoire tampon, si une mise en œuvre choisit de représenter en interne la mémoire tampon de données non étiquetée en utilisant des STag. Cependant, parce que, par définition, la STag n'est pas visible du réseau, c'est un problème d'hôte local, spécifique de la mise en œuvre qui devrait être analysé dans le contexte d'une analyse de sécurité de la mise en œuvre spécifique de l'hôte local, et donc, cela sort du domaine d'application du présent document.

Pour une mémoire tampon de données étiquetée (c'est-à-dire, un transfert de données étiqueté) soit l'ULP privilégié, soit le gestionnaire de ressource privilégié agissant au nom de l'ULP non privilégié, peut initialiser une transposition d'une STag en un tableau de traduction de page, ou peut avoir la capacité de simplement activer/désactiver une transposition de STag

existante en tableau de traduction de page. Il peut aussi y avoir plusieurs noms de STag qui se transposent en un groupe spécifique d'entrées (ou sous entrées) de tableau de traduction de page. Les questions de sécurité spécifiques de ce niveau de flexibilité sont examinées au paragraphe 6.2.3, "Plusieurs STag pour accéder à la même mémoire tampon".

Il y a diverses options de mise en œuvre pour l'initialisation des entrées de tableau de traduction de page et la transposition d'une STag en un groupe d'entrées de tableau de traduction de page qui ont des répercussions sur la sécurité. Cela inclut de prendre en charge la séparation de la transposition d'une STag et de la transposition d'un ensemble d'entrées de tableau de traduction de page, et de prendre en charge les ULP qui manipulent directement les transpositions de STag en entrées de tableau de traduction de page (plutôt que d'exiger l'accès par le gestionnaire de ressource privilégié).

2.3.5 Interactions de transfert de données au RNIC

Les opérations de transfert de données au RNIC peuvent être subdivisée en opérations d'envoi et de réception.

Pour les opérations d 'envoi, il y a normalement une file d'attente qui active l'ULP pour poster plusieurs demandes d'opération d'envoi de données (appelée la file d'attente d'envoi). Selon la mise en œuvre, les mémoires tampon de données utilisées dans les opérations peuvent ou non avoir des entrées de tableau de traduction de page associées, et peuvent ou non avoir des STag associées. Parce que ceci est un problème de mise en œuvre spécifique de l'hôte local plutôt qu'un problème de protocole, l'analyse de sécurité de menaces et de leurs atténuations est laissée à la mise en œuvre d'hôte.

Les opérations de réception sont différentes pour les mémoires tampon de données étiquetées et les mémoires tampon de données non étiquetées (c'est-à-dire, de transfert de données étiquetées opposé au transfert de données non étiquetées). Pour le transfert de données non étiquetées, si plus d'une mémoire tampon de données non étiquetées peuvent être postées par l'ULP, la spécification de DDP exige qu'elles soient consommées en ordre séquentiel (la spécification de RDMAP l'exige aussi). Donc, la mise en œuvre la plus générale est qu'il y a une file d'attente séquentielle des mémoires tampon de réception de données non étiquetées (file d'attente de réception). Certaines mises en œuvre peuvent aussi prendre en charge le partage de la file d'attente séquentielle entre plusieurs flux. Dans ce cas, définir "séquentiel" devient non trivial - en général, les mémoires tampon pour un seul flux sont consommées depuis la file d'attente dans l'ordre dans lequel elles ont été placées dans la file d'attente, mais il n'y a pas de garantie d'ordre de consommation entre les flux.

Pour la réception de transfert de données étiquetées (c'est-à-dire, de mémoires tampon de données étiquetées, de mémoire tampon RDMA Write, ou RDMA Read) à un certain moment avant le transfert de données, la transposition de la STag en entrées de tableau de traduction de page spécifiques (si il en est) et la transposition des entrées de tableau de traduction de page en la mémoire tampon de données doivent avoir été initialisées (voir au paragraphe 2.3.4 les détails d'interaction).

3. Confiance et partage de ressources

On suppose que, en général, les homologues, local et distant, ne sont pas de confiance, et donc des attaques par l'un ou l'autre devraient prévoir des atténuations.

Un problème distinct, mais en relation est le partage de ressources entre plusieurs flux. Si les ressources locales ne sont pas partagées, les ressources sont dédiées flux par flux. Les ressources sont définies au paragraphe 2.2, "Resources". L'avantage de ne pas partager de ressources entre les flux est que cela réduit les types d'attaques possibles. L'inconvénient de ne pas partager les ressources est que les ULP pourraient se trouver à court de ressources. Donc, ce peut être une forte incitation à partager les ressources, si les problèmes de sécurité associés au partage de ressources peuvent être atténués.

On suppose dans ce document que le composant qui met en œuvre le mécanisme pour contrôler le partage des ressources du moteur RNIC est le gestionnaire de ressource privilégié. Le moteur RNIC expose ses ressources à travers l'interface de RNIC au gestionnaire de ressource privilégié. Toutes les ressources de demande d'ULP privilégié et non privilégié provenant du gestionnaire de ressources (noter que par définition les deux applications non privilégiée et privilégiée peuvent essayer de consommer avidement les ressources, créant donc une potentielle attaque de déni de service (DoS)). Le gestionnaire de ressources met en œuvre des politiques de gestion de ressources pour s'assurer d'un accès équitable aux ressources. Le gestionnaire de ressources devrait être conçu pour prendre en compte les attaques contre la sécurité détaillées dans ce document. Noter que pour certains systèmes, le gestionnaire de ressource privilégié peut être mis en œuvre au sein de l'ULP privilégié.

Toutes les interactions d'ULP non privilégié avec le moteur RNIC qui pourraient affecter d'autres ULP DOIVENT être faites en utilisant le gestionnaire de ressource privilégié comme mandataire. Toutes les demandes d'allocation de ressources

d'ULP pour des ressources rares DOIVENT aussi être faites en utilisant un gestionnaire de ressources privilégié.

Le partage de ressources à travers les flux devrait être sous le contrôle de l'ULP, à la fois en termes de modèle de confiance que l'ULP souhaite utiliser, aussi bien que de niveau de partage de ressources que l'ULP souhaite donner au traitement local. Pour une discussion des types de modèles de confiance qui combinent confiance partielle et partage de ressources, voir l'Appendixe C, "Taxonomie de la confiance partielle".

Le gestionnaire de ressource privilégié NE DOIT PAS supposer que des flux différents partagent une confiance mutuelle partielle sauf si il y a un mécanisme pour assurer que les flux partagent bien une confiance mutuelle partielle. Ceci peut être fait de plusieurs façons, incluant une notification explicite de la part de l'ULP qui possède les flux.

4. Capacités de l'attaquant

Les capacités d'un attaquant délimitent les types d'attaques que l'attaquant est capable de lancer. RDMAP et DDP exigent que le flux initial de LLP (et la connexion) soient établis avant de transférer des messages RDMAP/DDP. Ceci exige au moins qu'une prise de contact d'un aller-retour se produise.

Si l'attaquant n'est pas l'homologue distant qui a créé la connexion initiale, les capacités de l'attaquant peuvent être segmentées en capacités d'envoi seul ou en capacités d'envoi et de réception. Attaquer avec des capacités d'envoi seul exige que l'attaquant devine d'abord les paramètres de flux de LLP courants avant qu'il puisse attaquer les ressources de RNIC (par exemple, le numéro de séquence TCP). Si cette classe d'attaquant a aussi des capacités de réception et la capacité de se faire passer pour le receveur à l'envoyeur et l'envoyeur au receveur, elle est normalement désignée comme un attaquant "interposé" [RFC3552]. Un attaquant interposé a une bien plus grande capacité à attaquer les ressources de RNIC. La portée de l'attaque est essentiellement la même que celle d'un homologue distant attaquant (c'est-à-dire, l'homologue distant qui établit le flux initial de LLP).

5. Attaques qui peuvent être atténuées avec la sécurité de bout en bout

Cette Section décrit les attaques contre RDMAP/DDP où la seule solution est de mettre en œuvre une forme de sécurité de bout en bout. L'analyse inclut une description détaillée de chaque attaque, de ce qui est attaqué, et une description des contre-mesures qui peuvent être prises pour déjouer l'attaque.

Certaines formes d'attaque impliquent de modifier la charge utile RDMAP ou DDP par un attaquant appuyé sur le réseau ou impliquent de surveiller le trafic pour découvrir des informations confidentielles. Un outil efficace pour s'assurer de la confidentialité est de chiffrer le flux de données par des mécanismes comme le chiffrement IPsec. De plus, des protocoles d'authentification, comme l'authentification IPsec, sont des outils efficaces pour s'assurer que l'entité distante est qui elle prétend être, aussi bien que s'assurer que la charge utile n'esr pas modifiée quand elle traverse le réseau.

Noter que l'établisement et la suppression de connexion sont supposés être faits en mode flux (c'est-à-dire, pas d'encapsulation RDMA de la charge utile) de sorte qu'il n'y ait pas de nouvelles attaques relatives à l'établissement/suppression de la connexion au delà de ce qui est déjà présent dans le LLP (par exemple, TCP ou SCTP). Noter cependant, que les paramètres RDMAP/DDP peuvent être échangés en mode flux, et si ils sont corrompus par un attaquant, des conséquences imprévues vont en résulter. Donc, toutes les atténuations existantes pour l'usurpation de LLP, altération, répudiation, divulgation d'informations, déni de service, ou élévation de privilège, continuent de s'appliquer (et sortent du domaine d'application de ce document). Donc, l'analyse de cette Section se concentre sur les attaques qui sont présentes, sans considération du type de flux de LLP.

L'altération est toute modification du trafic légitime (interne à la machine ou au réseau). L'attaque d'usurpation d'identité est un cas particulier d'altération où l'attaquant falsifie une identité de l'homologue distant (l'identité peut être une adresse IP, un nom de machine, une identité de niveau ULP, etc.).

5.1 Usurpation d'identité

Les attaques en usurpation d'identité peuvent être lancées par l'homologue distant, ou par un attaquant dans le réseau. Une attaque en usurpation d'identité fondée sur le réseau s'applique à tous les homologues distants. Ce paragraphe analyse les divers types d'attaques en usurpation d'identité applicables à RDMAP et DDP.

5.1.1 Se faire passer pour quelqu'un d'autre

Un attaquant fondé sur le réseau peut se faire passer pour un homologue RDMAP/DDP légal (en usurpant une adresse IP légale). Ceci peut être fait comme une attaque aveugle (voir la [RFC3552]) ou en établissanr un flux RDMAP/fDDP avec la victime. Parce que un flux RDMAP/fDDP exige qu'un flux de LLP soit pleinement initialisé (par exemple, pour la [RFC0793], il est dans l'état ÉTABLI) les mécanismes existants de protection de la couche transport contre les attaques aveugles restent en place.

Pour qu'une attaque aveugle réussisse, il faut que l'attaquant injecte un segment valide de couche transport (par exemple, pour TCP, il doit correspondre au moins au quadruplet et aussi deviner un numéro de séquence au sein de la fenêtre) tout en devinant aussi des paramètres RDMAP ou DDP valides. Il y a de nombreuses façons d'attaquer le protocole RDMAP/DDP si le protocole de transport est supposé être vulnérable. Par exemple, pour les messages étiquetés, cela entraîne de deviner les valeurs de STag et de TO. Si l'attaquant souhaite simplement terminer la connexion, il peut le faire en devinant correctement les valeurs de couche transport et réseau, et en fournissant une STag invalide. Selon la spécification DDP, si une STag invalide est reçue, le flux est supprimé et une erreur est notifiée à l'homologue distant. Si un attaquant souhaite écraser une mémoire tampon annoncée, il doit réussir à deviner les STag et TO correctes. Étant donné que le TO va souvent commencer à zéro, c'est facile. La valeur de STag devrait être choisie au hasard, comme discuté au paragraphe 6.1.1, en utilisant une STag sur un flux différent. Pour les messages non étiquetés, si le MSN est invalide, la connexion peut alors être supprimée. Si il est valide, les mémoires tampon de réception peuvent être corrompues.

L'authentification de bout en bout (par exemple, IPsec ou d'ULP) fournit une protection contre l'attaque aveugle ou l'attaque connectée.

5.1.2 Capture de flux

La capture de flux se produit quand un attaquant fondé sur le réseau espionne la connexion de LLP à travers la phase d'établissement de flux, et attend que la phase d'authentification (si elle existe) soit achevée avec succès. L'attaquant usurpe alors l'adresse IP et redirige le flux de la victime à sa propre machine. Par exemple, un attaquant peut attendre que l'authentification iSCSI soit achevée avec succès, et capturer alors le flux iSCSI.

La meilleure protection contre cette forme d'attaque est la protection de l'intégrité de bout en bout et l'authentfiication, comme par IPsec, pour empêcher l'usurpation d'identité. Une autre option est de fournir un réseau physiquement séparé pour la sécurité. La discussion de la sécurité physique sort du domaine d'application de ce document.

Parce que la connexion et/ou le flux lui-même est établi par le LLP, certains LLP sont plus difficiles à capturer que d'autres. Voir la documentation de LLP pertinente sur les questions de sécurité autour de la capture de connexion et/ou de flux.

5.1.3 Attaque par interposition

Si un attaquant fondé sur le réseau a la capacité de supprimer ou modifier des paquets qui vont quand même être acceptés par le LLP (par exemple, le numéro de séquence TCP est correct) alors le flux peut être exposé à une attaque par interposition. Un style d'attaque par interposition est d'envoyer des messages étiquetés (RDMAP ou DDP). Si il peut découvrir une mémoire tampon qui a été exposée pour un accès activé par une STag, l'interposé peut utiliser une opération RDMA Read pour lire le contenu de la mémoire tampon de données associée, effectuer une opération RDMA Write pour modifier le contenu de la mémoire tampon de données associée, ou invalider la STag pour désactiver tout autre accès à la mémoire tampon.

La meilleure protection contre cette forme d'attaque est la protection d'intégrité de bout en bout et l'authentification, comme avec IPsec, pour empêcher l'usurpation d'identité ou l'altération. Si l'authentification et la protection de l'intégrité ne sont pas utilisées, la protection physique doit alors être employée pour empêcher des attaques par interposition.

Parce que la connexion et/ou le flux lui-même sont établis par le LLP, certains LLP sont plus exposés que d'autres à l'attaque par interposition. Voir la documentation de LLP pertinente sur les questions de sécurité autour de la capture de connexion et/ou de flux.

Une autre approche est de restreindre l'accès seulement au sous-réseau/liaison local, et de fournir des mécanismes pour limiter l'accès, comme la sécurité physique ou 802.1.x. Ce modèle est un scénario de déploiement extrêmement limité, et ne sera pas examiné plus avant.

5.2 Altération - modification appuyée sur le réseau du contenu de mémoire tampon

Ceci est en fait une attaque par interposition, mais seulement sur le contenu de la mémoire tampon, par opposition à l'attaque par interposition présentée ci-dessus, où la signalisation et le contenu peuvent tous deux être modifiés. Voir au paragraphe 5.1.3, "Attaque par interposition".

5.3 Divulgation d'informations - espionnage appuyé sur le réseau

Un attaquant qui est capable d'espionner sur le réseau peut lire le contenu de tous les accès en lecture et écriture dans les mémoires tampon d'un homologue. Pour empêcher la divulgation d'informations, les données en lecture/écriture doivent être chiffrées. Voir aussi au paragraphe 5.1.3, "Attaque par interposition". Le chiffrement peut être fait soit par l'ULP, soit par un protocole qui peut fournir des services de sécurité à RDMAP et DDP (par exemple, IPsec).

5.4 Exigences spécifiques pour les services de sécurité

Généralement parlant, la confidentialité du flux protège contre l'espionnage. L'authentification et la protection d'intégrité du flux et/ou session sont une contre-mesure contre diverses attaques d'usurpation d'identité et d'altération. L'efficacité de l'authentification et de la protection de l'intégrité contre une attaque spécifique dépend de si l'authentification est au niveau machine (comme avec IPsec) ou au niveau de l'ULP.

5.4.1 Introduction aux options de sécurité

Les services de sécurité suivants peuvent être appliqués à un flux RDMAP/DDP :

1. Confidentialité de session : protège contre l'espionnage (paragraphe 5.3).

2. Authentification de la source des données par paquet : protège contre les attaques d'usurpation d'identité suivantes : se faire passer pour quelqu'un d'autre dans le réseau (paragraphe 5.1.1) et capture de flux (paragraphe 5.1.2).

3. Intégrité par paquet : protège contre l'altération faite par une modification du contenu de la mémoire tampon fondée sur le réseau (paragraphe 5.2) et quand elle est combinée avec l'authentification, protège aussi contre les attaques par interposition (paragraphe 5.1.3).

4. Séquençage de paquet : protège contre les attaques en répétition, qui sont un cas particulier de l'attaque d'altération ci-dessus.

Si un flux RDMAP/DDP peut être soumis à des attaques où on se fait passer pour un autre, ou des attaques de capture de flux, il est recommandé que le flux soit authentifié, protégé en intégrité, et protégé des attaques en répétition ; on peut utiliser la protection de la confidentialité pour protéger contre l'espionnage (dans le cas où le flux RDMAP/DDP traverse un réseau public).

IPsec est une suite de protocoles utilisée pour sécuriser la communication à la couche réseau entre deux homologues. La suite de protocoles IPsec est spécifiée au sein des documents d'architecture de sécurité IP [RFC2401], IKE [RFC2409], en-tête d'authentification IPsec (AH) [RFC2402], et charge utile de sécurité encapsulante IPsec (ESP) [RFC2406]. IKE est le protocole de gestion de clés, tandis que AH et ESP sont utilisés pour protéger le trafic. Voir dans ces RFC une description complète des protocoles respectifs.

IPsec est capable de fournir les services de sécurité ci-dessus pour le trafic respectivement IP et TCP. Les protocoles de couche supérieure sont capables de fournir seulement une partie de ces services de sécurité.

5.4.2 TLS n'est pas approprié pour la sécurité de DDP/RDMAP

TLS [RFC4346] fournit l'authentification, l'intégrité et la confidentialité des flux pour les ULP fondés sur TCP. TLS prend en charge l'authentification fondée sur les certificats unidirectionnelle (seulement du serveur) ou mutuelle.

Si TLS est mis en couche en dessous de RDMAP, l''orientation de la connexion TLS rend TLS inapproprié pour la sécurité de DDP/RDMA. Si un chiffrement de flux ou de bloc en mode CBC est utilisé pour un chiffrement en vrac, un paquet peut être déchiffré seulement après que tous les paquets précédents sont déjà arrivés. Si TLS est utilisé pour protéger le trafic DDP/RDMAP, TCP doit alors rassembler tous les paquets déclassés avant que TLS puisse les déchiffcer. C'est seulement

après cela que RDMAP/DDP peut les placer dans la mémoire tampon d'ULP. Donc, une des principales caractéristiques de DDP/RDMAP - permettre aux mises en œuvre d'avoir une architecture de flux avec peu ou pas de mise en mémoire tampon - ne peut pas être réalisée si TLS est utilisé pour protéger le flux de données.

Si TLS est mis en couche par dessus RDMAP ou DDP, TLS ne protège pas les en-têtes RDMAP et/ou DDP. Donc, une attaque par interposition peut se produire en modifiant l'en-tête RDMAP/DDP pour placer les données dans une mauvaise mémoire tampon, corrompant donc effectivement le flux de données.

Pour ces raisons, il n'est pas RECOMMANDÉ que TLS soit mis en couche par dessus RDMAP ou DDP.

5.4.3 DTLS et RDDP

DTLS [RFC4347] fournit des services de sécurité pour les protocoles de datagrammes, incluant des protocoles de datagrammes non fiables. Ces services incluent l'anti-répétition fondée sur un mécanisme adapté de IPsec qui est destiné à fonctionner sur les paquets lorsque ils sont reçus du réseau. Pour cette raison et d'autres, DTLS est mieux appliqué à RDDP en employant DTLS en dessous de TCP, donnant une mise en couches de RDDP sur TCP sur DTLS sur UDP/IP. Une telle mise en couches insère DTLS en gros au même niveau que IPsec dans la pile de protocoles, faisant des services de sécurité de DTLS une solution de remplacement aux services de IPsec du point de vue de RDDP.

Pour RDDP, IPsec est le meilleur choix de cadre de sécurité , et donc est de mise en œuvre obligatoire (comme spécifié ailleurs dans le présent document). Un facteur contributif important de la spécification de IPsec plutôt que de DTLS est que les versions non RDDP des deux déploiements initiaux de RDDP (iSCSI [RFC3720], [RFC5046] et NFSv4 [RFC3530], [RFC5661]) sont compatibles avec IPsec mais aucun de ces protocoles n'utilise actuellement ni TLS ni DTLS. Pour le cas spécifique de iSCSI, IPsec est la base des services de sécurité de mise en œuvre obligatoire [RFC3723]. Donc, le présent document et les spécifications du protocole RDDP contiennent des exigences de mise en œuvre obligatoires pour IPsec plutôt que pour DTLS.

5.4.4 ULP qui fournissent la sécurité

Les ULP qui fournissent une sécurité intégrée mais souhaitent développer une sécurité de protocole de couche inférieure, devraient être conscients des problèmes de sécurité qui se rapportent à la corrélation de mécanismes de sécurité d'un canal spécifique à l'authentification effectuée par l'ULP. Voir dans la [RFC5056] des informations supplémentaires sur une approche prometteuse appelée "lien de canal". D'après la [RFC5056] : "Le concept de lien de canal permet aux applications de prouver que les points d'extrémité de deux canaux sécurisés à des couches réseau différentes sont les mêmes en liant l'authentification à un canal à la protection de session à l'autre canal. L'utilisation des liens de canaux permet aux applications de déléguer la protection de sessions aux couches inférieures, ce qui peut significativement améliorer les performances pour certaines applications."

5.4.5 Exigences pour l'encapsulation IPsec de DDP

Le groupe de travail "IP Storage" a passé un temps et des efforts significatifs pour définir les exigences normatives de IPsec pour la mémorisation IP [RFC3723]. Des portions de cette spécification sont applicables à une grande variété de protocoles, incluant la suite de protocoles RDDP. Afin de ne pas répéter cet effort, une mise en œuvre de RNIC DOIT suivre les exigences définies dans la RFC 3723, paragraphe 2.3 et Section 5, incluant les références normatives associées pour ces paragraphes. Noter que cela signifie que la prise en charge du mode ESP d'IPsec est normative.

De plus, comme le matériel d'accélération IPsec peut seulement être capable de traiter un nombre limité de SA actives IKE phase 2, les messages Delete de phase 2 peuvent être envoyés pour des SA inactives comme moyen de garder le nombre de SA actives de phase 2 au minimum. La réception d'un message Delete IKE phase 2 NE DOIT PAS être interprétée comme une raison pour supprimer un flux DDP/RDMA. Il est préférable de laisser le flux actif, et si du trafic supplémentaire y est envoyé, d'établir une autre SA IKE de phase 2 pour le protéger. Cela évite de continuellement établir et supprimer les flux.

Noter qu'il y a de sérieux problèmes de sécurité si IPsec n'est pas mis en œuvre de bout en bout. Par exemple, si IPsec est mis en œuvre comme un tunnel au milieu du réseau, tous les hôtes entre l'homologue et l'appareil de tunnelage IPsec peuvent librement attaquer le flux non protégé.

Les exigences de IPsec pour RDDP se fondent sur la version de IPsec spécifiée dans la [RFC2401] et les RFC en rapport, comme le profil de la [RFC3723], en dépit de l'existence d'une nouvelle version de IPsec spécifiée dans la [RFC4301] et les RFC qui s'y rapportent. Une des importantes premières applications des protocoles RDDP est leur utilisation avec iSCSI

[RFC5046] ; les exigences de IPsec pour RDDP suivent celles de IPsec afin de faciliter leur usage en permettant qu'un profil commun de IPsec soit utilisé avec iSCSI et les protocoles RDDP. À l'avenir, la RFC 3723 pourrait être mise à jour avec la plus récente version de IPsec ; les exigences de sécurité de IPsec d'une telle mise à jour devraient s'appliquer uniformément à iSCSI et aux protocoles RDDP.

6. Attaques provenant d'homologues distants

Cette Section décrit les attaques à distance qui sont possibles contre le système RDMA défini à la Figure 1 "Modèle de sécurité RDMA" et les ressources de moteur RNIC définies au paragraphe 2.2. L'analyse inclut une description détaillée de chaque attaque, de ce qui est attaqué, et une description des contre-mesures qui peuvent être prises pour déjouer l'attaque.

Les attaques sont classées en cinq catégories : usurpation d'identité, altération, divulgation d'informations, déni de service (DoS) et élévation de privilège. Comme mentionné précédemment, l'altération est toute modification du trafic légitime (interne à la machine ou au réseau). Une attaque d'usurpation d'identité est un cas particulier d'altération où l'attaquant falsifie une identité de l'homologue distant (l'identité peut être une adresse IP, un nom de machine, une identité au niveau de l'ULP, etc.).

6.1 Usurpation d'identité

Ce paragraphe analyse les divers types d'attaques d'usurpation d'identité applicables à RDMAP et DDP. Les attaques d'usurpation d'identité peuvent être lancées par l'homologue distant ou un attaquant dans le réseau. Pour les contre-mesures contre un attaquant dans le réseau, voir la Section 5, "Attaques qui peuvent être atténuées avec la sécurité de bout en bout".

6.1.1 Utilisation d'une STag sur un flux différent

Un style d'attaque par l'homologue distant est de tenter d'utiliser des valeurs de STag qui ne lui sont pas autorisées. Noter que si l'homologue distant envoie une STag invalide à l'homologue local, selon les spécifications de DDP et RDMAP, le flux doit être supprimé. Donc, la menace existe si une STag a été activée pour l'accès à distance sur un flux et qu'un homologue distant est capable de l'utiliser sur un flux sans rapport avec lui. Si l'attaque réussit, l'attaquant pourrait éventuellement être capable d'effectuer des opérations RDMA Read pour lire le contenu de la mémoire tampon de données associée, effectuer des opérations RDMA Write pour modifier le contenu de la mémoire tampon de données associée, ou invalider la STag pour empêcher d'autre accès à la mémoire tampon.

Une tentative par un homologue distant d'accéder à une mémoire tampon avec une STag sur un flux différent dans le même domaine de protection peut ou non être une attaque, si on a l'intention de partager les ressources (c'est-à-dire, si les flux partagent la confiance mutuelle partielle). Pour certains ULP, utiliser une STag sur plusieurs flux au sein du même domaine de protection pourrait être le comportement désiré. Pour d'autres ULP, tenter d'utiliser une STag sur un flux différent pourrait être considéré comme une attaque. Comme cela varie selon l'ULP, un ULP va normalement avoir besoin d'être capable de contrôler la portée de la STag.

Dans le cas où une mise en œuvre ne partage pas de ressources entre les flux (incluant les STag) cette attaque peut être combattue en allouant chaque flux à un domaine de protection différent. Avant de permettre l'accès à distance à la mémoire tampon, le domaine de protection du flux où la tentative d'accès a été faite est confronté au domaine de protection de la STag. Si les domaines de protection ne correspondent pas, l'accès à la mémoire tampon est refusé, une erreur est générée, et le flux RDMAP associé au flux attaquant est terminé.

Pour les mises en œuvre qui partagent des ressources entre plusieurs flux, il peut n'être pas pratique de séparer chaque flux en son propre domaine de protection. Dans ce cas, l'ULP peut quand même limiter la portée de toutes les STag à un seul flux (si il l'active pour l'accès à distance). Si la portée de la STag a été limité à un seul flux, toute tentative d'utiliser cette STag sur un flux différent va résulter en une erreur, et le flux RDMAP sera terminé.

Donc, pour les mises en œuvre qui ne partagent pas les STag entre les flux, soit chaque flux DOIT être dans un domaine de protection séparé, soit la portée d'une STag DOIT être limité à un seul flux.

Un RNIC DOIT s'assurer qu'un flux spécifique dans un domaine de protection spécifique ne peut pas accéder à une STag dans un domaine de protection différent.

Un RNIC DOIT s'assurer que, si une STag est limitée en portée à un seul flux, aucun autre flux ne peut utiliser cette STag.

Un problème supplémentaire peut être le partage involontaire de STag (c'est-à-dire, une faute dans l'ULP) ou une faute dans l'homologue distant qui cause l'utilisation d'une STag avec un décalage de un. Pour une protection supplémentaire, une mise en œuvre devrait allouer des STag d'une façon telle qu'il soit difficile de prédire le numéro de la prochaine STag allouée, et aussi de s'assurer que les STag sont réutilisées le moins souvent possible. Toute méthode d'allocation qui conduirait à une réutilisation intentionnelle ou non d'une STag par l'homologue devrait être évitée (par exemple, une méthode qui commencerait toujours par une certaine STag et une augmentation monotone de numéro pour chaque nouvelle allocation, ou une méthode qui utiliserait toujours la même STag pour chaque opération).

6.2 Altération

Un attaquant homologue distant ou fondé sur le réseau peut tenter d'altérer le contenu des mémoires tampon de données sur un homologue local qui a été activé pour l'accès en écriture distant. Les types d'attaques d'altération d'un homologue distant sont mentionnés dans les paragraphes qui suivent. Pour les contre-mesures contre un attaquant fondé sur le réseau, voir à la Sectione 5, "Attaques qui peuvent être atténuées avec la sécurité de bout en bout.

6.2.1 Débordement de mémoire tampon - réponse RDMA Write ou Read

Cette attaque est une tentative de l'homologue distant d'effectuer une réponse RDMA Write ou Read sur une mémoire en dehors de la gamme de longueurs valides de la mémoire tampon de données activée pour l'accès en écriture distant. Cette attaque peut survenir même quand aucune ressource n'est partagée entre les flux. Ce problème peut aussi survenir si l'ULP a une faute.

La contre-mesure pour ce type d'attaque doit être dans la mise en œuvre de RNIC, en s'apuyant sur la STag. Quand l'ULP local spécifie au RNIC l'adresse de base et le nombre d'octets dans la mémoire tampon qu'il souhaite rendre accessibles, le RNIC doit s'assurer que les vérifications de base et de limites sont appliquées à tout accès à la mémoire tampon référencée par la STag avant que la STag soit activée pour l'accès. Quand une opération de transfert des données RDMA (qui inclut une STag) arrive sur un flux, une vérification de base et de limites d'accès de granularité d'octet doit être effectuée pour s'assurer que l'opération accède seulement au sein de la mémoire tampon aux localisations de mémoire décrites par cette STag.

Donc une mise en œuvre de RNIC DOIT s'assurer qu'un homologue distant n'est pas capable d'accéder à une mémoire en dehors de la mémoire tampon spécifiée quand la STag a été activée pour l'accès à distance.

6.2.2 Modification d'une mémoire tampon après l'indication

Cette attaque peut se produire si un homologue distant tente de modifier le contenu d'une mémoire tampon référencée par une STag en effectuant une réponse RDMA Write ou Read après que l'homologue distant a indiqué à l'homologue local ou l'ULP local (par divers moyens) que le contenu de la STag de mémoire tampon de données est prêt à l'emploi. Cette attaque peut survenir même quand aucune ressources n'est partagée entre les flux. Noter qu'une faute dans l'homologue distant, ou une altération fondée sur le réseau, pourrait aussi résulter en ce problème.

Par exemple, si on suppose que la mémoire tampon référencée par la STag contient des informations de contrôle d'ULP aussi bien qu'une charge utile d'ULP, et si la séquence d'opération de l'ULP est de d'abord valider les informations de contrôle et ensuite d'effectuer des opérations sur les informations de contrôle. Si l'homologue distant peut effectuer une réponse RDMA Write ou Read supplémentaire (donc, de changer la mémoire tampon) après que les vérifications de validité ont été achevées mais avant que les données de contrôle aient été traitées, l'homologue distant pourrait forcer l'ULP à opérer sur des chemins qui n'ont jamais été prévus.

L'ULP local peut se protéger contre ce type d'attaque en révoquant l'accès à distance quand le transfert de données original a été achevé et avant qu'il valide le contenu de la mémoire tampon. L'ULP local peut faire cela soit en révoquant explicitement les droits d'accès distants pour la STag quand l'homologue distant indique que l'opération est achevée, ou en vérifiant que l'homologue distant a invalidé la STag par la capacité d'invalidation de RDMAP à distance. Si l'homologue distant n'a pas invalidé la STag, l'ULP local révoque alors explicitement les droits d'accès distants de la STag. (Voir au paragraphe 6.4.5, "Invalidation à distance d'une STag partagée sur plusieurs flux" pour une définition de l'invalidation à distance.)

L'ULP local DEVRAIT suivre la procédure ci-dessus pour protéger la mémoire tampon avant de valider le contenu de la mémoire tampon (ou utiliser la mémoire tampon de quelque façon que ce soit).

Un RNIC DOIT s'assurer que les paquets du réseau qui utilisent la STag pour une mémoire tampon annoncée précédemment ne peuvent plus modifier la mémoire tampon après que l'ULP a révoqué les droits d'accès distants pour la STag spécifiée.

6.2.3 Plusieurs STag pour accéder à la même mémoire tampon

Voir cette analyse au paragraphe 6.3.6, "Utilisation de plusieurs STag qui se transposent en la même mémoire tampon".

6.3 Divulgation d'informations

La principale source potentielle de divulgation d'informations est par une mémoire tampon locale qui a été activée pour l'accès à distance. Si la mémoire tampon peut être sondée par un homologue distant sur un autre flux, il y a une divulgation potentielle d'informations.

Les attaques potentielles qui pourraient résuler en une divulgation involontaire d'informatione et les contre-mesures sont détaillées dans les paragraphes qui suivent.

6.3.1 Sondage de mémoire en dehors des limites de la mémoire tampon

C'est essentiellement la même attaque que décrite au paragraphe 6.2.1, "Débordement de mémoire tampon - réponse RDMA Write ou Read", sauf qu'une demande RDMA Read est utilisée pour monter l'attaque. La même contre-mesure s'applique.

6.3.2 Utilisation de RDMA Read pour accéder à des données périmées

Si une mémoire tampon est utilisé pour une combinaison de lectures et d'écritures (distantes ou locales) et est exposée à un homologue distant avec au moins des droits d'accès en lecture distante avant qu'elle soit initialisée avec les données correctes, il y a une condition potentielle de concurrence où l'homologue distant peut voir le contenu précédent de la mémoire tampon. Cela devient un problème de sécurité si le contenu précédent de la mémoire tampon n'était pas destiné à être partagé avec l'homologue distant.

Pour éliminer cette condition de concurrence, l'ULP local DEVRAIT s'assurer qu'aucune donnée périmée n'est contenue dans la mémoire tampon avant que des droits d'accès en lecture distante soient accordés (ceci peut être fait, par exemple, en mettant à zéro le contenu de la mémoire). Cela assure que l'homologue distant ne peut pas accéder à la mémoire tampon jusqu'à ce que les données périmées aient été supprimées.

6.3.3 Accès à une mémoire tampon après le transfert

Si l'homologue distant a l'accès en lecture distante à une mémoire tampon et, par un mécanisme quelconque, dit à l'ULP local que le transfert a été achevé, mais si l'ULP local ne désactive pas l'accès à distance à la mémoire tampon avant de modifier les données, il est possible à l'homologue distant de restituer les nouvelles données.

Ceci est similaire à l'attaque définie au paragraphe 6.2.2, "Modifier une mémoire tampon après indication". Les mêmes contre-mesures s'appliquent. De plus, l'ULP local DEVRAIT accorder des droits d'accès en lecture distante seulement pendant la durée nécessaire pour restituer les données.

6.3.4 Accès à des données non prévues avec une STag valide

Si l'ULP active l'accès à distance à une mémoire tampon en utilisant une STag qui fait référence à la mémoire tampon entière, mais a seulement l'intention de donner l'accès à une portion de la mémoire tampon, il est alors possible à l'homologue distant d'accéder aux autres parties de la mémoire tampon.

Pour empêcher cette attaque, l'ULP DEVRAIT établir la base et les limites de la mémoire tampon quand la STag est initialisée pour exposer seulement les données à restituer.

6.3.5 RDMA Read dans une mémoire tampon RDMA Write

Une forme de divulgation peut survenir si les droits d'accès à la mémoire tampon ont activé la lecture à distance, quand seulement l'accès en écriture distant était prévu. Si la mémoire tampon contenait des données d'ULP, ou des données provenant d'un transfert sur un flux sans relation, l'homologue distant pourrait restituer les données par une opération RDMA Read. Noter qu'une mise en œuvre de RNIC n'est pas obligée de prendre en charge des STag qui aient l'accès à la fois en lecture et en écriture.

La contre-mesure la plus évidente pour cette attaque est de ne pas accorder l'accès en lecture distante si la mémoire tampon est destinée à être seulement en écriture. L'homologue distant ne sera alors pas capable de restituer les données associées à la mémoire tampon. Une tentative pour le faire résulterait en une erreur et le flux RDMAP associé au flux serait terminé.

Donc, si un ULP a l'intention qu'une mémoire tampon soit seulement exposée pour l'accès en écriture distante, il DOIT régler les droits d'accès à la mémoire tampon à seulement activer l'accès en écriture distante. Noter que cette exigence n'est pas destinée à restreindre l'utilisation de RDMA Read de longueur zéro. Les RDMA Read de longueur zéro n'exposent pas de données d'ULP. Parce que ils sont destinés à être utilisés comme un mécanisme pour assurer que tous les RDMA Write ont été reçus, et n'exigent même pas une STag valide, leur utilisation est permise même si une mémoire tampon a été seulement activée pour l'accès en écriture.

6.3.6 Utilisation de plusieurs STag qui se transposent en la même mémoire tampon

Plusieurs STag qui se transposent en la même mémoire tampon au même moment peut résulter en une divulgation d'informations involontaire si les STag sont utilisées par des homologues distants différents, entre lesquels n'existe pas de confiance mutuelle. Ce modèle s'applique spécifiquement à la communication client/serveur, où le serveur communique avec plusieurs clients, dont aucun n'a de confiance mutuelle avec chaque autre.

Si seulement l'accès en écriture est activé, alors l'ULP local a un contrôle complet sur la divulgation d'informations. Donc, un serveur qui avait l'intention d'exposer les mêmes données (c'est-à-dire, une mémoire tampon) à plusieurs clients en utilisant plusieurs STag pour la même mémoire tampon ne crée aucun nouveau problème de sécurité au delà de ce qui a déjà été décrit dans le présent document. Noter que si le serveur n'avait pas l'intention d'exposer les mêmes données aux clients, il devrait utiliser des mémoires tampon (et STag) séparées pour chaque client.

Quand une STag a l'accès en lecture distante activé et qu'une STag différente a l'accès en écriture distant activé pour la même mémoire tampon, il est possible à un homologue distant de voir le contenu qui a été écrit par un autre homologue distant.

Si les deux STag ont l'accès en écriture distant activé et si les deux homologues distants ne partagent pas de confiance mutuelle, il est possible à un homologue distant d'écraser les contenus qui ont été écrits par l'autre homologue distant.

Donc, un ULP avec plusieurs homologues distants qui ne partagent pas de confiance mutuelle partielle NE DOIVENT PAS accorder l'accès en écriture à la même mémoire tampon par différentes STag. Une mémoire tampon devrait être exposée à seulement un homologue distant non de confiance à la fois pour assurer qu'aucune divulgation ou altération d'informations ne se produit entre les homologues.

6.4 Déni de service (DoS)

Une attaque de DoS est un des principaux risques pour la sécurité de RDMAP. C'est parce que les ressources de RNIC sont rares et précieuses, et que de nombreux environnements d'ULP exigent une communication avec des homologues distants qui ne sont pas de confiance. Si l'homologue distant peut être authentifié ou si la charge utile d'ULP peut être chiffrée, il est clair que le profil de DoS peut être réduit. Pour les besoins de cette analyse, on suppose que le RNIC doit être capable d'opérer dans des environnements qui ne sont pas de confiance, qui sont ouverts aux attaques de style DoS.

Les attaques de déni de service contre les ressources de RNIC ne sont pas le bombardement typique par un tiers inconnu de paquets sur un hôte aléatoire (comme une attaque de TCP SYN). Parce que la connexion/flux doit être pleinement établie (par exemple, une prise de contact de trois messages de couche transport a eu lieu) l'attaquant doit être capable d'envoyer et recevoir des messages sur cette connexion/flux, ou être capable de deviner un paquet valide sur un flux RDMAP existant.

Cette section souligne les attaques potentielles et les contre-mesures disponibles pour traiter chaque attaque.

6.4.1 Épuisement des ressources du RNIC

Ce paragraphe traite des attaques qui tombent dans la catégorie générale de l'ULP local qui tente d'allouer inéquitablement des ressources rares (c'est-à-dire, limitées) de RNIC. L'ULP local peut tenter d'allouer des ressources en son nom propre, ou au nom d'un homologue distant. Les ressources qui entrent dans cette catégorie incluent des domaines de protection, de la mémoire de contexte de flux, des tableaux de traduction et de protection, et de l'espace de noms de STag. Cela peut être dû à des attaques par des ULP locaux actuellement actifs ou qui ont alloué des ressources antérieurement mais sont maintenant inactifs.

Ce type d'attaque peut survenir sans considération de si les ressources sont partagées à travers les flux.

L'allocation de ressources rares DOIT être placée sons le contrôle d'un gestionnaire de ressources privilégié. Cela permet au gestionnaire de ressource privilégié :

* d'empêcher un ULP local d'allouer plus que sa juste part des ressources ;
* de détecter si un homologue distant tente de lancer une attaque de DoS en tentant de créer un nombre excessif de flux (avec les ressources associées) et de prendre une action corrective (comme de refuser la demande ou d'appliquer des filtres de couche réseau à l'homologue distant).

Cette analyse suppose que le gestionnaire de ressources est responsable la distribution des domaines de protection, et que les mises en œuvre de RNIC vont fournir assez de domaines de protection pour permettre au gestionnaire de ressources d'être capable d'allouer un unique domaine de protection à chaque ULP local sans rapport, non de confiance (pour un nombre limité raisonnable d'ULP locaux). Cette analyse suppose de plus que le gestionnaire de ressources met en œuvre des politiques pour s'assurer que les ULP locaux non de confiance ne sont pas capables de consommer tous les domaines de protection par une attaque de DoS. Noter que la consommation de domaines de protection ne peut pas résulter d'une attaque de DoS lancée par un homologue distant, sauf si un ULP local agit au nom de l'homologue distant.

6.4.2 Consommation des ressources par des ULP inactives

La plus simple forme d'attaque de DoS, étant donnée une quantité de ressources fixée, est que l'homologue distant crée un flux RDMAP vers un homologue local, demande des ressources dédiées, et ne fasse ensuite aucun travail réel. Cela permet à l'homologue distant d'être très léger (c'est-à-dire, de seulement négocier des ressources, mais de ne pas transférer de données) et de consommer une quantité disproportionnée de ressources chez l'homologue local.

Une contre-mesure générale pour ce style d'attaque est de surveiller les flux RDMAP actifs et, si les ressources baissent, de récolter les ressources des flux RDMAP qui ne transfèrent pas de données et éventuellement de terminer le flux. Cela va probablement être sous contrôle administratif.

Voir au paragraphe 6.4.1 l'analyse et les contre-mesures pour ce style d'attaque sur les ressources de RNIC suivantes : mémoire de contexte de flux, tableaux de traduction de page, et espace de noms de STag.

Noter que certaines ressources de RNIC ne courent pas de risque de ce type d'attaque de la part d'un homologue distant parce que une attaque exige que l'homologue distant envoie des messages afin de consommer les ressources. Les mémoires tampon de réception de données, les ressources de file d'attente d'achèvement, et les ressources de file d'attente de demandes RDMA Read en sont des exemples. Ces ressources courent cependant le rique qu'un ULP local qui tente d'allouer des ressources devienne alors inactif. Cela pourrait aussi être créé si l'ULP négocie des niveaux de ressource avec l'homologue distant, ce qui cause une consommation de ressources chez l'homologue local ; cependant, l'homologue distant n'envoie jamais de données pour les consommer. La contre-mesure générale décrite dans ce paragraphe peut être utilisée pour libérer les ressources allouées par un homologue local inactif.

6.4.3 Consommation des ressources par des ULP actives

Ce paragraphe décrit les attaques de DoS provenant des homologues locaux et distants qui échangent activement des messages. Les attaques sur chaque ressource de NIC RDMA sont examinées et les contre-mesures spécifiques sont identifiées. Noter que les attaques sur la mémoire de contexte de flux, les tableaux de traduction de page, et l'espace de noms de STag sont traitées au paragraphe 6.4.1, "Consommation des ressources de RNIC", de sorte qu'elles ne sont pas incluses ici.

6.4.3.1 Plusieurs flux partagent des mémoires tampon de réception

L'homologue distant peut tenter de consommer plus que sa juste part des mémoires tampon de réception de données (c'est-à-dire, des mémoires tampon non étiquetées pour des messages DDP ou de type Send pour RDMAP) si les mémoires tampon de réception sont partagées à travers plusieurs flux.

Si les ressources ne sont pas partagées à travers plusieurs flux, alors cette attaque n'est pas possible parce que l'homologue distant ne va pas être capable de consommer plus de mémoires tampon qu'il n'en a été alloué au flux. Le pire scénario est celui où l'homologue distant peut consommer plus de mémoires tampon de réception que n'en a alloué l'ULP local, résultant en ce qu'aucune mémoire tampon n'est disponible, ce qui pourrait causer que le flux de l'homologue distant à l'homologue local soit supprimé, et que toutes les ressources allouées soit libérées.

Si les mémoires tampon de réception de données locales sont partagées entre plusieurs flux, alors l'homologue distant peut tenter de consommer plus que sa juste part des mémoires tampon de réception, causant un manque de mémoires tampon de réception chez un flux différent, et donc, de causer éventuellement la suppression de l'autre flux. Par exemple, si l'homologue distant a envoyé suffisamment de messages non étiquetés de un octet, cela pourrait être capable de consommer toutes les ressources de la file d'attente de réception partagée localement, avec peu d'efforts de sa part.

Une méthode que l'homologue local pourrait utiliser pour reconnaître qu'un homologue distant tente d'utiliser plus que sa juste part des ressources est de terminer le flux (causant la libération des ressources allouées). Cependant, si l'homologue local est suffisamment lent, il est encore possible que l'homologue distant monte une attaque de déni de service. Une contre-mesure qui peut protéger contre cette attaque est de mettre en œuvre une notification de bas niveau. La notification de bas niveau alerte l'ULP si le nombre de mémoires tampon dans la file d'attente de réception est inférieur à un certain seuil.

Si toutes les conditions suivantes sont vraies, alors l'homologue local ou l'ULP local peut dimensionner la quantité de mémoires tampon de réception locales allouée à la file d'attente de réception pour s'assurer qu'une attaque de DoS peut être stoppée :

* une notification de bas niveau est activée, et
* l'homologue local est capable de limiter la durée qu'il lui faut pour remplir les mémoires tampon de réception, et
* l'homologue local tient des statistiques pour déterminer quel homologue distant consomme les mémoires tampon.

Les conditions ci-desssus permettent à la notification de bas niveau d'arriver avant que les ressources soient épuisées, et donc, l'homologue ou ULP local peut prendre une action corrective (par exemple, terminer le flux de l'homologue distant attaquant).

Une attaque différente, mais qui présente des similitudes, est si l'homologue distant envoie un nombre significatif de paquets déclassés et si le RNIC a la capacité d'utiliser la mémoire tampon d'ULP (c'est-à-dire, la mémoire tampon non étiquetée pour DDP ou la mémoire tampon consommée par un message de type Send pour RDMAP) comme une mémoire tampon de réassemblage. Dans ce cas, l'homologue distant peut consommer un nombre significatif de mémoires tampon d'ULP, mais jamais envoyer asssez de données pour permettre à la mémoire tampon d'ULP d'être saturée à l'ULP.

Une contre-mesure efficace est de créer une notification de bas niveau qui alerte l'ULP si il y a plus qu'un nombre spécifié de mémoires tampon de réception "en cours" (partiellement consommées, mais pas saturées). La notification est générée quand plus que le nombre spécifié de mémoires tampon sont en cours simultanément sur un flux spécifique (c'est-à-dire, des paquets ont commencé d'arriver pour la mémoire tampon, mais la mémoire tampon n'a pas encore été livrée à l'ULP).

Une contre-mesure différente est que le moteur RNIC fournisse un moyen de limiter la capacité de l'homologue distant de consommer des mémoires tampon de réception flux par flux. Malheureusement, cela exige qu'une grande quantité d'état soit suivie dans chaque RNIC flux par flux.

Donc, si un moteur RNIC fournit la capacité de patager les mémoires tampon de réception sur plusieurs flux, la combinaison du moteur RNIC et du gestionnaire de ressource privilégié DOIT être capable de détecter si l'homologue distant tente de consommer plus que sa juste part des ressources afin que l'homologue local ou l'ULP local puisse appliquer des contre-mesures pour détecter et empêcher l'attaque.

6.4.3.2 Homologue distant ou local qui attaque une CQ partagée

Pour une vue d'ensemble du modèle d'attaque de file d'attente d'achèvement partagée, voir au paragraphe 7.1.

L'homologue distant peut attaquer une CQ partagée en consommant plus que sa juste part d'entrées de CQ en utilisant une

des méthodes suivantes :
* L'ULP permet à l'homologue distant de faire que l'ULP local réserve un nombre spécifié d'entrées de CQ, éventuellement en laissant des entrées insuffisantes pour les autres flux qui partagent la CQ.
* Si l'homologue distant, l'homologue local, ou l'ULP local (ou toute combinaion de ceux-ci) peut attaquer la CQ en submergeant la CQ jusqu'à saturation, alors le processus d'achèvement sur les autres flux qui partagent cette file d'attente d'achèvement peut être affecté (par exemple, la file d'attente d'achèvement déborde et cesse de fonctionner).

La première méthode d'attaque peut être évitée si l'ULP ne permet pas à un homologue distant de réserver des entrées de CQ, ou si il y a un intermédiaire de confiance, comme un gestionnaire de ressource privilégié. Malheureusement, il est souvent irréaliste de ne pas permettre à un homologue distant de réserver des entrées de CQ, en particulier si le nombre d'entrées d'achèvement dépend d'autres paramètres d'ULP négociés, comme la quantité de mémoire tampon requise par l'ULP. Donc, une mise en œuvre DOIT utiliser un gestionnaire de ressource privilégié pour contrôler l'allocation d'entrées de CQ. Voir au paragraphe 2.1, "Composants", la définition d'un gestionnaire de ressource privilégié.

Une façon dont un homologue local ou distant peut tenter de submerger une CQ avec des achèvements est d'envoyer des messages RDMAP/DDP de longueur minimum pour causer autant d'achèvements par seconde que possible (achèvements de réception pour l'homologue distant, achèvements d'envoi pour l'homologue local). Si c'est l'homologue distant qui attaque, et on suppose que la ou les files d'attente de réception de l'homologue local ne sont pas à court de mémoires tampon de réception (si elles le sont, c'est alors une attaque différente, documentée au paragraphe 6.4.3.1 "Plusieurs flux partageant des mémoires tampon de réception") alors il serait possible à l'homologue distant de consommer plus que sa juste part des entrées de file d'attente d'achèvement. Selon la mise en œuvre de CQ, cela pourrait causer le débordement de CQ (si elle n'est pas assez grande pour traiter tous les achèvements générés) ou qu'un autre flux ne soit pas capable de générer des entrées de CQ (si le RNIC a le contrôle des flux sur la génération des entrées de CQ dans la CQ). Dans l'un et l'autre cas, la CQ va cesser de fonctionner correctement, et tous les flux qui attendent des achèvements sur la CQ vont cesser de fonctionner.

Cette attaque peut se produire sans considération de si tous les flux associés à la CQ sont dans le même domaine de protection ou des domaines différents - le problème clé est que le nombre d'entrées de file d'attente d'achèvement est inférieur au nombre de toutes les opérations en instance qui peuvent causer un achèvement.

L'homologue local peut se protéger de ce type d'attaque en utilisant une des méthodes suivantes :
* dimensionner la CQ au niveau approprié, comme spécifié ci-dessous (noter que si la CQ existe actuellement et a besoin d'être redimensionnée, redimensionner la CQ n'est pas obligé pour réussir dans tous les cas, de sorte que le redimensionnement de CQ devrait être fait avant de dimensionner la file d'attente d'envoi et la file d'attente de réception sur le flux) OU
* accorder moins de ressources que ce que l'homologue distant a demandé (ne pas fournir le nombre de mémoires tampon de réception de données demandé).

Le dimensionnement approprié de la CQ dépend de si le ou les ULP locaux vont poster autant de ressources pour les diverses files d'attente que la taille de la file d'attente le permet. Si on peut faire confiance au ou aux ULP locaux pour poster un nombre de ressources plus petit que la taille de la file d'attente de la ressource spécifique, alors une CQ de dimension correcte signifie que la CQ est assez grande pour contenir l'état d'achèvement pour toutes les mémoires tampon de données en instance (mémoires tampon d'envoi et de réception) ou :

CQ_MIN_SIZE = SUM(MaxPostedOnEachRQ) + SUM(MaxPostedOnEachSRQ) + SUM(MaxPostedOnEachSQ)

Où :
MaxPostedOnEachRQ = nombre maximum de demandes qui peut causer un achèvement qui va être posté sur une file d'attente de réception spécifique.
MaxPostedOnEachSRQ = nombre maximum de demandes qui peut causer qu'un achèvement va être posté sur une file d'attente de réception partagée spécifique.
MaxPostedOnEachSQ = nombre maximum de demandes qui peut causer qu'un achèvement va être posté sur une file d'attente d'envoi spécifique.

Si l'ULP local doit être capable de remplir complètement les files d'attente, ou ne peut pas être de confiance pour observer une limite plus petite que les files d'attente, alors la CQ doit être dimensionnée pour s'accommoder du nombre maximum d'opérations qu'il est possible de poster à tout moment. Donc, l'équation devient :

CQ_MIN_SIZE = SUM(SizeOfEachRQ) + SUM(SizeOfEachSRQ) + SUM(SizeOfEachSQ)

Où :

SizeOfEachRQ = nombre maximum de demandes pouvant causer un achèvement qui peut toujours être posté sur une file d'attente de réception spécifique.
SizeOfEachSRQ = nombre maximum de demandes pouvant causer un achèvement qui peut toujours être posté sur une file d'attente de réception spécifique partagée.
SizeOfEachSQ = nombre maximum de demandes pouvant causer un achèvement qui peut toujours être posté sur une file d'attente d'envoi spécifique.

MaxPosted*OnEach*Q et SizeOfEach*Q varient flux par flux ou par file d'attente de réception partagée.

Si l'ULP partage une CQ à travers plusieurs flux qui ne partagent pas de confiance mutuelle partielle, il DOIT alors mettre en œuvre un mécanisme pour assurer que la file d'attente d'achèvement ne déborde pas. Noter qu'il est possible de partager des CQ même si les homologues distants qui accèdent aux CQ ne sont pas de confiance si l'une des deux formules ci-dessus est mise en œuvre. Si on peut faire confiance à l'ULP pour ne pas poster plus que MaxPostedOnEachRQ, MaxPostedOnEachSRQ, et MaxPostedOnEachSQ, alors la première formule s'applique. Si on ne peut pas faire confiance à l'ULP pour respecter la limite, alors la seconde formule s'applique.

6.4.3.3 Attaque de la file d'attente de demandes RDMA Read

La file d'attente de demandes RDMA Read peut être attaquée si l'homologue distant envoie plus de demandes RDMA Read que la profondeur de la file d'attente de demandes RDMA Read chez l'homologue local. Si la file d'attente de demandes RDMA Read est une ressource partagée, cela pourrait corrompre la file d'attente. Si la file d'attente n'est pas partagée, le pire cas est alors que le flux en cours ne soit plus fonctionnel (par exemple, supprimé). Une approche pour résoudre la file d'attente de demandes RDMA Read partagée serait de créer des seuils, similaires à ceux décrits au paragraphe 6.4.3.1, "Plusieurs flux partagent des mémoires tampon de réception". Une approche plus simple est de ne pas partager les ressources de file d'attente de demandes RDMA Read entre les flux ou d'appliquer des limites strictes de consommation par flux. Donc, la consommation de ressources de file d'attente de demandes RDMA Read DOIT être contrôlée par le gestionnaire de ressource privilégié afin que les flux RDMAP/DDP qui ne partagent pas de confiance mutuelle partielle ne partagent pas de ressources de file d'attente de demandes RDMA Read.

Si le problème est une faute dans la mise en œuvre de l'homologue distant, mais pas une attaque malveillante, il peut être résolu en demandant au RNIC de l'homologue distant de réduire les demandes RDMA Read. En configurant de façon appropriée le flux chez l'homologue distant à travers un agent de confiance, on peut faire que le RNIC ne transmette pas de demandes RDMA Read au delà de la profondeur de la file d'attente de demandes RDMA Read chez l'homologue local. Si le flux est correctement configuré, et si l'homologue distant soumet plus de demandes que ce que peut contenir la file d'attente de demandes RDMA Read de l'homologue local, la demande va être mise en file d'attente au RNIC de l'homologue distant jusqu'à ce que les demandes précédentes s'achèvent. Si le flux de l'homologue distant n'est pas configuré correctement, le flux RDMAP est terminé quand il arrive chez l'homologue local plus de demandes RDMA Read qu'il ne peut en traiter (en supposant que la recommandation du paragraphe précédent est mise en œuvre). Donc, une mise en œuvre de RNIC DEVRAIT fournir un mécanisme pour contrôler le nombre de demandes RDMA Read en instance. La configuration de cette limite sort du domaine d'application du présent document.

6.4.4 Utilisation de chemins de code non optimaux

Une autre forme d'attaque de DoS est de tenter de contraindre à des chemins de données qui peuvent consommer une quantité de ressources disproportionnée. Un exemple pourrait être si les cas d'erreur sont traités sur un "chemin lent" (consommant les ressources de calcul de l'hôte ou du RNIC) et qu'un attaquant génère un nombre excessif d'erreurs pour tenter de consommer ces ressources. Noter que pour la plupart des erreurs RDMAP ou DDP, le flux attaquant va simplement être supprimé. Donc, pour que cette forme d'attaque soit efficace, l'homologue distant doit contraindre à des chemins de données qui ne causent pas la suppression du flux.

Si une mise en œuvre de RNIC contient des "chemins lents" qui ne résultent pas en la suppression du flux, il est recommandé qu'une mise en œuvre fournisse la capacité de détecter cette condition et permette à un administrateur d'agir, incluant éventuellement de supprimer administrativement le flux RDMAP associé au flux qui contraint les chemins de données, qui consomment une quantité de ressources disproportionnée.

6.4.5 Invalidation à distance d'une STag partagée sur plusieurs flux

Si un homologue local a activé une STag pour l'accès à distance, l'homologue distant pourrait tenter d'invalider à distance la STag en utilisant le message RDMAP Send avec Invalidate ou Send avec SE et Invalidate. Si la STag est seulement

valide sur le flux en cours, le seul effet est que l'homologue distant ne peut plus utiliser la STag ; donc, il n'y a pas de problème de sécurité.

Si la STag est valide sur plusieurs flux, alors l'homologue distant peut empêcher les autres flux d'utiliser cette STag en utilisant la fonction d'invalidation à distance.

Donc, si les flux RDDP ne partagent pas de confiance mutuelle partielle (c'est-à-dire, si l'homologue distant peut tenter d'invalider prématurément à distance la STag) l'ULP NE DOIT PAS permettre une STag qui serait valide sur plusieurs flux.

6.4.6 L'homologue distant attaque une CQ non partagée

L'homologue distant peut attaquer une CQ non partagée si l'homologue local ne dimensionne pas correctement la CQ. Par exemple, si l'homologue local permet que la CQ traite les achèvements des mémoires tampon de réception, et si la file d'attente de la mémoire tampon de réception est plus longue que la file d'attente d'achèvement, un débordement peut éventuellement se produire. L'effet sur le flux de l'attaquant est catastrophique. Cependant, si un RNIC n'a pas en place les protections appropriées, une attaque pour submerger la CQ peut alors aussi causer la corruption et/ou la terminaison d'un flux sans rapport. Donc, un RNIC DOIT s'assurer que si une CQ déborde, aucun des flux qui n'utilisent pas la CQ NE DOIT être affecté.

6.5 Élévation de privilège

L'architecture de sécurité RDMAP/DDP différencie explicitement trois niveaux de privilège : non privilégié, privilégié, et gestionnaire de ressource privilégié. Si un ULP non privilégié est capable d'élever son niveau de privilège à celui d'ULP privilégié, la transposition d'une liste d'adresses physiques en une STag peut fournir l'accès local et à distance à toute localisation d'adresse physique sur le nœud. Si un ULP en mode privilégié est capable de se promouvoir à être gestionnaire de ressources, il lui est alors possible d'effectuer des attaques de type déni de service où des quantités substantielles de ressources locales pourraient être consommées.

En général, l'élévation de privilège est un problème spécifique de mise en œuvre locale et donc qui sort du domaine d'application du présent document.

7. Attaques provenant d'homologues locaux

Cette section décrit les attaques locales qui sont possibles contre le système RDMA défini à la Figure 1, "Modèle de sécurité RDMA" et les ressources de moteur RNIC définies au paragraphe 2.2.

7.1 ULP local attaquant une CQ partagée

Les attaques de DoS contre une file d'attente d'achèvement partagée (CQ - voir au paragraphe 2.2.6, "Files d'attente d'achèvement") peuvent être causées par l'ULP local ou l'homologue distant si l'un ou l'autre tente de causer plus d'achèvements que sa juste part du nombre d'entrées ; donc, potentiellement d'affamer un autre ULP sans relation de façon qu'aucune entrée de file d'attente d'achèvement ne soit disponible.

Une entrée de file d'attente d'achèvement peut éventuellement être malicieusement consommée par l'achèvement de la file d'attente d'envoi ou celui de la file d'attente de réception. Dans le premier cas, l'attaquant est l'ULP local, dans le second, l'attaquant est l'homologue distant.

Une forme d'attaque peut survenir lorsque les ULP locaux peuvent consommer des ressources sur la CQ. Un ULP local qui est lent à libérer les ressources sur la CQ en ne récoltant pas assez rapidement l'état d'achèvement pourrait bloquer tous les autres ULP locaux qui tentent d'utiliser cette CQ.

Pour ces raisons, un RNIC NE DOIT PAS permettre le partage d'une CQ à travers des ULP qui ne partagent pas de confiance mutuelle partielle.

7.2 Homologue local attaquant la file d'attente de demandes RDMA Read

Si les ressources de la file d'attente de demandes RDMA Read sont groupées sur plusieurs flux, une attaque est si l'ULP

local tente d'allouer de façon inéquitable les ressources de la file d'attente de demandes RDMA Read pour ses flux. Par exemple, un ULP local ttente d'allouer toutes les ressources disponibles sur une file d'attente de demandes RDMA Read spécifique pour ses flux, déniant ainsi la ressource aux ULP avec qui il partage la file d'attente de demandes RDMA Read. Le même type d'argument s'applique même si la demande RDMA Read n'est pas partagée, mais qu'un ULP local tente d'allouer toutes les ressources du RNIC quand la file d'attente est créée.

Donc, l'accès aux interfaces qui allouent des entrées de file d'attente de demandes RDMA Read DOIT être restreint à un homologue local de confiance, tel qu'un gestionnaire de ressource privilégié. Le gestionnaire de ressource privilégié DEVRAIT empêcher un ULP local d'allouer plus que sa juste part de ressources.

7.3 ULP local attaquant la transposition de PTT et de STag

Si un ULP non privilégié est capable de manipuler directement les tableaux de traduction de page de RNIC (qui traduisent une STag en adresse d'hôte) il serait possible que l'ULP non privilégié puisse pointer le tableau de traduction de page sur les mémoires tampon d'un flux ou ULP sans relation et, par là, être capable d'obtenir l'accès à des informations du flux/ULP sans relation.

Comme exposé à la Section 2, "Modèle architectural", l'introduction d'un gestionnaire de ressource privilégié pour arbitrer les demandes de transposition est une contre-mesure efficace. Cela permet au gestionnaire de ressource privilégié de s'assurer qu'un ULP local peut seulement initialiser le tableau de traduction de page (PTT) pour pointer sur ses propres mémoires tampon.

Donc, si les ULP non privilégiés sont pris en charge, le gestionnaire de ressource privilégié DOIT vérifier que l'ULP non privilégié a le droit d'accéder à une mémoire tampon de données spécifique avant de permettre une STag pour laquelle l'ULP a des droits d'accès à être associé à une mémoire tampon de données spécifique. Cela peut être fait quand le tableau de traduction de page est initialisé à accéder à la mémoire tampon de données ou quand la STag est initialisée à pointer sur un groupe d'entrées de tableau de traduction de page, ou les deux.

8. Considérations sur la sécurité

Prière de voir dans la Section 5, "Attaques qui peuvent être atténuées avec la sécurité de bout en bout", la Section 6, "Attaques provenant des homologues distants", et la Section 7, "Attaques provenant des homologues locaux" une analyse détaillée des attaques et des contre-mesures normatives pour atténuer les attaques.

De plus, les appendices fournissent un résumé des exigences de sécurité pour des audiences spécifiques. L'Appendice A, "Problèmes d'ULP pour les protocoles client/serveur RDDP", donne un résumé des questions de mise en œuvre et des exigences pour les applications qui mettent en œuvre un style traditionnel d'interaction client/serveur. Il donne des directives d'applicabilité supplémentaires au texte normatif des Sections 5, 6, et 7. L'Appendice B, "Résumé des exigences de mise en œuvre de RNIC et d'ULP" donne un résumé pratique des exigences normatives pour les mises en œuvre.

9. Considérations relatives à l'IANA

Les considérations relatives à l'IANA ne sont pas traitées par ce document. Toutes les considérations relatives à l'IANA résultant de l'utilisation de DDP ou RDMA doivent être traitées dans les normes pertinentes.

10. Références

10.1 Références normatives

[RFC0793] J. Postel (éd.), "Protocole de commande de transmission – Spécification du protocole du programme Internet DARPA", STD 7, septembre 1981. (*Remplacée par* RFC9293)

[RFC2401] S. Kent et R. Atkinson, "Architecture de sécurité pour le protocole Internet", novembre 1998. (*Obsolète, voir* RFC4301)

[RFC2402] S. Kent et R. Atkinson, "En-tête d'authentification IP", novembre 1998. (*Obsolète, voir* RFC4302, 4305)

[RFC2406] S. Kent et R. Atkinson, "Encapsulation de charge utile de sécurité IP (ESP)", novembre 1998. *(Ob., voir* RFC4303)

[RFC2409] D. Harkins et D. Carrel, "L'échange de clés Internet (IKE)", novembre 1998. *(Obsolète, voir la* RFC4306)

[RFC3723] B. Aboba et autres, "Protocoles de sécurisation de mémorisation de blocs sur IP", avril 2004. *(P.S.)*

[RFC4960] R. Stewart, éd., "Protocole de transmission de commandes de flux (SCTP)", septembre 2007. (*Remplace* RFC2960, RFC3309 ; *P.S. ; Remplacée par* RFC9260)

[RFC5040] R. Recio et autres, "Spécification d'un protocole d'accès direct à une mémoire distante", octobre 2007. *(P.S. ; MàJ par* RFC7146 *)*

[RFC5041] H. Shah et autres, "Placement direct des données sur transports fiables", octobre 2007. *(P.S. ; MàJ par RFC 7146)*

10.1 Références pour information

[INFINIBAND] "InfiniBand Architecture Specification Volume 1", release 1.2, InfiniBand Trade Association standard, <http://www.infinibandta.org/specs>. Les verbes sont documentés au chapitre 11.

[RFC3530] S. Shepler et autres, "Protocole de système de fichiers réseau (NFS) v. 4", avril 2003. (*P.S. ; remplacée par* RFC7530)

[RFC3552] E. Rescorla, B. Korver, "Lignes directrices pour la rédaction d'une section de considérations sur la sécurité dans les RFC", juillet 2003. (BCP0072)

[RFC3720] J. Satran et autres, "Interface Internet des systèmes de petits ordinateurs (iSCSI)", avril 2004. *(Remplacée par* RFC7143)

[RFC4301] S. Kent et K. Seo, "Architecture de sécurité pour le protocole Internet", décembre 2005. *(P.S.) (Remplace la* RFC2401)

[RFC4346] T. Dierks et E. Rescorla, "Protocole de sécurité de la couche Transport (TLS) version 1.1", avril 2006. *(Remplace* RFC2246 *; Remplacée par* RFC5246 *; MàJ par* RFC4366, 4680, 4681, 5746, 6176, 7465, 7507, 7919)

[RFC4347] E. Rescorla, N. Modadugu, "Sécurité de la couche de transport de datagrammes", avril 2006. *(P.S.)*

[RFC4949] R. Shirey, "Version 2 du glossaire de la sécurité sur Internet", août 2007. (*Remplace* RFC2828) (FYI0036) *(Info.)*

[RFC5045] C. Bestler et autres, "Applicabilité du protocole d'accès direct à une mémoire distante (RDMA) et du placement direct des données (DDP)", octobre 2007. *(Information)*

[RFC5046] M. Ko et autres, "Extensions pour l'accès direct à une mémoire distante (RDMA) à l'interface système de petit ordinateur à l'Internet (iSCSI)", octobre 2007. *(P.S. ; Remplacée par* RFC7145)

[RFC5056] N. Williams, "Sur l'utilisation des liaisons de canaux pour sécuriser les canaux", novembre 2007. *(P.S.)*

[RFC5661] S. Shepler, M. Eisler, D. Noveck, "Système de fichiers réseau (NFS) version 4.1 : Protocole", janvier 2010. *(P.S. ; MàJ par* [RFC8178, RFC8434 *; remplacée par* RFC8881)

[VERBS-RDMAC] "RDMA Protocol Verbs Specification", RDMA Consortium standard, avril 2003, <http://www.rdmaconsortium.org/home/draft-hilland-iwarp-verbs-v1.0-RDMAC.pdf>.

[VERBS-RDMAC-Overview] "RDMA enabled NIC (RNIC) Verbs Overview", présentation de transparents par Renato Recio, avril 2003, <http://www.rdmaconsortium.org/home/RNIC_Verbs_Overview2.pdf>.

Appendice A. Problèmes d'ULP pour les protocoles RDDP client/serveur

Cette Section est un appendice normatif du document centré sur les exigences de mise en œuvre d'ULP client/serveur pour assurer une mise en œuvre de serveur sécurisée.

Les sections précédentes mentionnaient des attaques spécifiques et leurs contre-mesures. Cette Section résume les attaques et contre-mesures qui ont été définies dans les sections précédentes, qui sont applicables à la création d'un serveur d'ULP sécurisé (par exemple, d'application). Un serveur d'ULP est défini comme un ULP qui doit être capable de communiquer avec de nombreux clients n'ayant pas nécessairement une relation de confiance les uns avec les autres, et pour assurer que chaque client ne peut pas attaquer un autre client par des interactions de serveurs. De plus, le serveur peut souhaiter utiliser plusieurs flux pour communiquer avec un client spécifique, et ces flux peuvent partager une confiance mutuelle. Noter que cette section suppose une mise en œuvre conforme de RNIC et de gestionnaire de ressource privilégié - donc, elle se concentre spécifiquement sur les questions de mise en œuvre de serveur d'ULP (par exemple, application).

Tous les détails des sections précédentes sur les attaques et contre-mesures s'appliquent au serveur; donc, les exigences qui sont répétées dans cette section utilisent les "doit", "devrait", et "peut" non normatifs. Dans certains cas, les déclarations normatives DEVRAIT pour l'ULP dans le corps principal de ce document sont transformées en déclarations DOIT pour le serveur d'ULP parce que les conditions de fonctionnement peuvent être rafinées pour rendre inapplicables le DEVRAIT. Si un DEVRAIT antérieur est changé en un DOIT dans cette section, il est explicitement noté et il utilise des déclarations normatives en majuscules.

La liste qui suit résume les attaques pertinentes que des clients peuvent monter sur le serveur partagé en requalifiant les déclarataions normative précédentes pour être spécifiques du client/serveur. Noter que chaque client/serveur d'ULP peut employer des opérations explicites RDMA (RDMA Read, RDMA Write) de différentes façons. Donc, lorsque approprié, "ULP local", "homologue local", et "homologue distant" sont utilisés à la place de "serveur" ou "client", afin de conserver la pleine généralité de chaque exigence.

* **Usurpation d'identité**

  * Les paragraphes 5.1.1 à 5.1.3. Pour la protection contre de nombreuses formes d'attaques par usurpation d'identité, activer IPsec.

  * Au paragraphe 6.1.1, en utilisant une STag sur un flux différent. Pour s'assurer qu'un client ne peut pas accéder aux données d'un autre client via l'utilisation de la STag de l'autre client, le serveur d'ULP doit soit dimensionner une STag à un seul flux, soit utiliser un domaine de protection unique par client. Si un seul client a plusieurs flux qui partagent une confiance mutuelle partielle, alors la STag peut être partagée entre les flux associés en utilisant un seul domaine de protection parmi les flux associés (voir au paragraphe 5.4.4, "ULP qui fournissent la sécurité", pour des problèmes supplémentaires). Pour empêcher un partage involontaire de STag au sein des flux associés, un serveur d'ULP devrait utiliser les STag d'une façon telle qu'il soit difficile de prédire le numéro de STag de la prochaine STag allouée .

* **Altération**

  * Au paragraphe 6.2.2 "Modification d'une mémoire tampon après indication". Avant que l'ULP local opère sur une mémoire tampon qui a été écrite par l'homologue distant en utilisant un RDMA Write ou RDMA Read, l'ULP local DOIT s'assurer que la mémoire tampon ne peut plus être modifiée en invalidant la STag pour l'accès à distance (noter que ceci est plus fort que le DEVRAIT du paragraphe 6.2.2). Ceci peut être fait soit en révoquant explicitement les droits d'accès distants pour la STag quand l'homologue distant indique que l'opération est achevée, soit en vérifiant que l'homologue distant a invalidé la STag par la capacité RDMAP Invalidate. Si l'homologue distant n'a pas invalidé la STag, l'ULP local révoque alors explicitement les droits d'accès distants de la STag.

* **Divulgation d'informations**

  * Au paragraphe 6.3.2, en utilisant RDMA Read pour l'accès à des données périmées. Dans un environnement de serveur d'utilisation générale, il n'y a pas de raisons fortes d'exiger qu'une mémoire tampon soit initialisée avant que la lecture à distance soit activée (et un énorme inconvénient de partage de données involontaire). Donc, un ULP local DOIT (ceci est plus fort que le DEVRAIT du paragraphe 6.3.2) s'assurer qu'aucune données périmées ne sont contenues dans une mémoire tampon avant que des droits d'accès en lecture distants soient accordés à un

homologue distant (cela peut être fait en mettant à zéro le contenu de la mémoire, par exemple).

* Au paragraphe 6.3.3, "Accès à une mémoire tampon après le transfert". Cette atténuation est déjà couverte par le paragraphe 6.2.2 (ci-dessus).

* Au paragraphe 6.3.4, "Accès à des données non prévues avec une STag valide". L'ULP doit régler la base et les limites de la mémoire tampon quand la STag est initialisée pour exposer seulement les données à restituer.

* Au paragraphe 6.3.5, "RDMA Read dans une mémoire tampon RDMA Write". Si un homologue a seulement l'intention qu'une mémoire tampon soit exposée pour l'accès en écriture distant, il doit régler les droits d'accès à la mémoire tampon comme activant seulement l'accès en écriture distant.

* Au paragraphe 6.3.6, "Utilisation de plusieurs STag qui se transposent en la même mémoire tampon". L'exigence du paragraphe 6.1.1 (ci-dessus) atténue cette attaque. Une mémoire tampon de serveur est exposée à seulement un client à la fois pour s'assurer qu'aucune divulgation ou altération d'informations ne se produit entre les homologues.

* Au paragraphe 5.3, "Espionnage fondé sur le réseau". Les services de confidentialité devraient être activés par l'ULP si cette menace pose problème.

* **Déni de service**

  * Au paragraphe 6.4.3.1, "Plusieurs flux partagent des mémoires tampon de réception". La taille de l'empreinte de mémoire d'ULP peut être importante pour certains serveurs d'ULP. Si un serveur d'ULP attend un trafic réseau significatif de plusieurs clients, utiliser une file d'attente de mémoires tampon de réception par flux lorsque il y a grand nombre de flux peut consommer des quantités substantielles de mémoire. Donc, une file d'attente de réception qui peut être partagée par plusieurs flux est intéressante. Cependant, à cause des attaques mentionnées dans ce paragraphe, partager une seule file d'attente de réception entre plusieurs clients doit seulement être fait si un mécanisme est en place pour assurer qu'un client ne peut pas consommer les mémoires tampon de réception au delà de ses limites, définies par chaque ULP. Pour plusieurs flux au sein d'un seul client d'ULP (qui partagent probablement une confiance mutuelle partielle) ces frais généraux supplémentaires peuvent être évités.

  * Au paragraphe 7.1 "ULP local attaquant une CQ partagée". Les atténuations normatives de RNIC exigent que le RNIC n'active pas le partage d'une CQ si les ULP locaux ne partagent pas une confiance mutuelle partielle. Donc, lorsque l'ULP n'a pas la permission d'activer cette caractéristique dans un mode non sûr, si les deux ULP locaux partagent la confiance mutuelle partielle, ils doivent se comporter de la manière suivante :
    1) Le dimensionnement de la file d'attente d'achèvement se fonde sur la taille de la file d'attente de réception et des files d'attente d'envoi, comme mentionné en 6.4.3.2, "Homologue distant ou local qui attaque une CQ partagée".
    2) L'ULP local s'assure que les entrées de CQ sont collectées assez fréquemment pour respecter les règles du paragraphe 6.4.3.2.

  * Au paragraphe 6.4.3.2, "Homologue distant ou local qui attaque une CQ partagée". Deux atténuations sont spécifiées dans ce paragraphe - une exige une taille de plus mauvais cas pour la CQ, et peut être mise en œuvre entièrement au sein du gestionnaire de ressource privilégié. La seconde approche requiert la coopération avec le serveur d'ULP local (pas pour poster trop de mémoires tampon) et active une plus petite CQ à utiliser. Dans certains environnements de serveur, la confiance partielle du serveur d'ULP (mais pas des clients) est acceptable ; donc, la plus petite CQ atténue complètement l'attaque distante. Dans d'autres environenments, le serveur d'ULP local pourrait aussi contenir des éléments non sûrs qui peuvent attaquer la machine locale (ou avoir des fautes). Dans ces environnements, le pire cas de taille de CQ doit être utilisé.

  * Au paragraphe 6.4.3.3, "Attaque de la file d'attente de demandes RDMA Read". Ce paragraphe exige qu'un gestionnaire de ressource privilégié de serveur ne permette pas le partage des files d'attente de demandes RDMA Read à travers plusieurs flux qui ne partagent pas de confiance mutuelle partielle pour un ULP qui effectue des opérations RDMA Read sur les mémoires tampon du serveur. Cependant, parce que le serveur d'ULP sait quels flux partagent le mieux la confiance mutuelle partielle, cette exigence peut être reflétée à l'ULP. L''exigence de l'ULP (c'est-à-dire, du serveur) dans ce cas, est qu'il NE DOIT PAS permettre que des files d'attente de demandes RDMA Read soient partagées entre les ULP qui n'ont pas une confiance mutuelle partielle.

  * Au paragraphe 6.4.5, "Invalidation à distance d'une STag partagée sur plusieurs flux". Cette atténuation est déjà couverte par le paragraphe 6.2.2 (ci-dessus).

Appendice B. Résumé des exigences de mise en œuvre de RNIC et d'ULP

Cet Appendice est pour information.

Voici un résumé des exigences de mise en œuvre pour le RNIC :
* 3 Partage de confiance et de ressources
* 5.4.5 Exigences pour l'encapsulation IPsec de DDP
* 6.1.1 Utilisation d'une STag sur un flux différent
* 6.2.1 Débordement de mémoire tampon - réponse RDMA Write ou Read
* 6.2.2 Modification d'une mémoire tampon après indication
* 6.4.1 Consommation des ressources du RNIC
* 6.4.3.1 Plusieurs flux partageant des mémoires tampon de réception
* 6.4.3.2 Homologue distant ou local qui attaque une CQ partagée
* 6.4.3.3 Attaque de la file d'attente de demandes RDMA Read
* 6.4.6 Homologue distant qui attaque une CQ non partagée
* 6.5 Élévation de privilège 39
* 7.1 ULP local attaquant une CQ partagée
* 7.3 ULP local attaquant la transposition de PTT et de STag

Voici un résumé des exigences de mise en œuvre pour l'ULP au-dessus du RNIC :
* 5.3 Divulgation d'informations - espionnage fondé sur le réseau
* 6.1.1 Utilisation d'une STag sur un flux différent
* 6.2.2 Modification d'une mémoire tampon après indication
* 6.3.2 Utilisation de RDMA Read pour accéder à des données périmées
* 6.3.3 Accès à une mémoire tampon après le transfert
* 6.3.4 Accès à des données non prévues avec une STag valide
* 6.3.5 RDMA Read dans une mémoire tampon RDMA Write
* 6.3.6 Utilisation de plusieurs STag qui se transposent en la même mémoire tampon
* 6.4.5 Invalidation à distance d'une STag partagée sur plusieurs flux

Appendice C. Taxonomie de la confiance partielle

Cet Appendice est pour information.

La confiance partielle est définie comme quand une partie veut supposer qu'une autre partie va s'abstenir d'une attaque ou ensemble d'attaques spécifique, les parties sont dites être dans un état de confiance partielle. Noter que l'homologue de confiance partielle peut tenter un ensemble différent d'attaques. Cela peut être approprié pour de nombreux ULP où tout effet adverse de la trahison est facilement confiné et ne fait pas peser d'autres risques sur les autres clients ou ULP.

Les modèles de confiance décrits dans cette section ont trois caractéristiques distinctives principales. Le modèle de confiance se réfère à un ULP local et un homologue distant, qui sont entendus être les instances d'ULP local et distant qui communiquent via RDMA/DDP.

* Partage de ressources local (oui/non) - Quand les ressources locales sont partagées, elles le sont à travers un groupement de flux RDMAP/DDP. Si les ressources locales ne sont pas partagées, les ressources sont dédiées flux par flux. Les ressources sont définies au paragraphe 2.2, "Ressources". L'avantage de ne pas partager les ressources entre les flux est que cela réduit les types d'attaques possibles. L'inconvénient est que les ULP pourraient être à bout de ressources.

* Confiance partielle locale (oui/non) - La confiance partielle locale est déterminée sur la base de si le groupement local des flux RDMAP/DDP (qui est normalement égal à un ULP ou groupe d'ULP) se fait mutuellement confiance pour ne pas effectuer un ensemble spécifique d'attaques.

* Confiance partielle distante (oui/non) - Le niveau de confiance partielle distante est déterminé sur le fait que l'ULP local d'un flux RDMAP/DDP spécifique accorde une confiance partielle à l'homologue distant du flux (voir la définition de la confiance partielle à la Section 1, "Introduction").

Toutes les combinaisons des caractéristiques de confiance sont supposées être utilisées par les ULP. Le présent document

analyse spécifiquement cinq modèles de confiance d'ULP qui sont supposés être d'usage courant. Les modèles de confiance sont :

* NS-NT - Ressources locales non partagées, pas de confiance locale, pas de confiance distante ; normalement, un serveur d'ULP qui veut fonctionner dans le mode le plus sûr possible. Toutes les atténuations d'attaque sont en place pour assurer un fonctionnement robuste.

* NS-RT - Ressources locales non partagées, pas de confiance locale, confiance partielle distante ; normalement, un ULP d'homologue à d'homologue qui a, par une méthode qui sort du domaine d'application du présent document, authentifié l'homologue distant. Noter que sauf si une forme d'authentification à clé est utilisée flux par flux RDMA/DDP, il peut n'être pas possible que des attaques par interposition se produisent.

* S-NT - Ressources locales partagées, pas de confiance locale, pas de confiance distante ; normalement, un serveur d'ULP qui fonctionne dans un environnement sans confiance où la quantité de ressources requise est soit trop grande, soit trop dynamique pour être dédiée à chaque flux RDMAP/DDP.

* S-LT - Ressources locales partagées, confiance locale partielle, pas de confiance distante ; normalement, un ULP qui fournit une couche de session et utilise plusieurs flux, pour fournir du débit supplémentaire ou des capacités de reprise sur défaillance. Tous les flux au sein de l'ULP local se font une confiance partielle les uns aux autres, mais ne font pas confiance à l'homologue distant. Ce modèle de confiance peut être approprié pour des environnements d'incorporation.

* S-T - Ressources locales partagées, confiance locale partielle, confiance partielle distante ; normalement, une application répartie, comme une application de base de données répartie ou une application d'ordinateur à hautes performances (HPC, *High Performance Computer*) qui est destiné à fonctionner dans une grappe. Du fait des exigences extrêmes de ressources et de performances, l'application s'authentifie normalement avec tous ses homologues et ensuite fonctionne dans un environnement de confiance. Les homologues d'application sont tous dans un seul domaine de faute d'application et dépendent les uns les autres de leur bon comportement quand ils accèdent aux structures de données. Si un homologue distant de confiance a un défaut de mise en œuvre qui résulte en un mauvais comportement, l'application entière pourrait être corrompue.

Les modèles NS-NT et S-NT, ci-dessus, sont typiques du réseautage Internet - ni l'ULP local ni l'homologue distant ne sont de confiance. Parfois, des optimisations peuvent être faites qui permettent le partage de tableaux de traduction de page sur plusieurs ULP locaux ; donc, le modèle S-LT peut être avantageux. Le modèle S-T est normalement utilisé quand l'adaptation de ressources à travers un grand ULP parallèle rend impossible d'utiliser un autre modèle. Les problèmes d'adaptation de ressources peuvent être dus aux performances autour de l'adaptation ou parce que simplement il n'y a pas assez de ressources. Le modèle NS-RT est probablement le modèle qui a le moins de chances d'être utilisé, mais il est présenté pour être complet.

Remerciements

Sara Bitan, Microsoft Corporation ; mél : sarab@microsoft.com
Allyn Romanow, Cisco Systems ; mél : allyn@cisco.com
Catherine Meadows, Naval Research Laboratory ; mél : meadows@itd.nrl.navy.mil
Patricia Thaler, Agilent Technologies, Inc. ; mél : pat_thaler@agilent.com
James Livingston, NEC Solutions (America), Inc. ; mél : james.livingston@necsam.com
John Carrier, Cray Inc. ; mél : carrier@cray.com
Caitlin Bestler, Broadcom ; mél : cait@asomi.com
Bernard Aboba, Microsoft Corporation ; mél : bernarda@windows.microsoft.com

Adresse des auteurs

James Pinkerton
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
téléphone : +1 (425) 705-5442
mél : jpink@windows.microsoft.com

Ellen Deleganes
P.O. Box 9245
Brooks, OR 97305
USA
téléphone : (503) 642-3950
mél : deleganes@yahoo.com

Déclaration complète de droits de reproduction