Groupe de travail Réseau
Request for Comments : 2873
Catégorie : En cours de normalisation

Traduction Claude Brière de L'Isle

X. Xiao, Global Crossing
A. Hannan, iVMG
V. Paxson, ACIRI/ICSI
E. Crabbe, Exodus Communications

juin 2000

Traitement TCP du champ Préséance IPv4

Résumé
Le présent mémoire décrit un conflit entre TCP [RFC0793] et DiffServ [RFC2475] sur l'utilisation des trois bits de gauche de l'octet Type de service dans un en-tête IPv4 [RFC0791]. Dans un réseau qui contient des nœuds à capacité DiffServ, un tel conflit peut causer l'échec des connexions TCP ou peut causer le rétablissement indésirable de connexions TCP. Le présent mémoire propose une modification à TCP pour résoudre le conflit.

Parce que l'octet de classe de trafic IPv6 [RFC2460] n'a pas de signification définie sauf ce qui est précisé dans la [RFC2474], et en particulier qu'il ne définit pas de bits de préséance ou de paramètre de sécurité, il n'y a pas de conflit entre TCP et DiffServ sur l'utilisation des bits dans l'octet de classe de trafic IPv6.

1. Introduction

Dans TCP, chaque connexion a un ensemble d'états associés. De tels états sont reflétés par un ensemble de variables mémorisées dans le bloc de contrôle TCP, (TCB, *TCP Control Block*) des deux extrémités. De tells variables peuvent inclure le numéro de prise locale et distante, la préséance de la connexion, le niveau de sécurité et le compartiment, etc. Les deux extrémités doivent s'accorder sur le réglage de la préséance et les paramètres de sécurité afin d'établir une connexion et de la garder ouverte.

Il n'y a pas de champ dans l'en-tête TCP qui indique la préséance d'un segment. À la place, le champ Préséance dans l'en-tête du paquet IP est utilisé comme indication. Le niveau de sécurité et le compartiment sont portés de même dans l'en-tête IP, mais comme options IP plutôt que comme un champ d'en-tête fixé. À cause de cette différence, le problème de la préséance discuté dans le présent mémoire ne s'applique pas à eux.

TCP exige que la préséance (et les paramètres de sécurité) d'une connexion reste inchangée durant la vie de la connexion. Donc, pour une connexion TCP établie avec une préséance, la réception d'un segment avec une préséance différente indique une erreur. La connexion doit être réinitialisée (pages 36, 37, 40, 66, 67, 71 de la [RFC0793].

Avec l'arrivée de DiffServ, des nœuds intermédiaires peuvent modifier le codet de service différencié (DSCP, *Differentiated Services Codepoint*) [RFC2474] de l'en-tête IP pour indiquer le comportement par bond (PHB, *Per-hop Behavior*) [RFC2475], [RFC2597], [RFC2598] désiré. Le DSCP inclut les trois bits anciennement connus comme champ de préséance. Comme toute modification de ces trois bits sera considérée comme illégale par les points d'extrémité qui ont connaissance des préséances, elles peuvent causer l'échec de l'établissement des connexions, ou peuvent causer la réinitialisation de connexions établies.

2. Terminologie

Segment : l'unité de données que TCP envoie à IP.

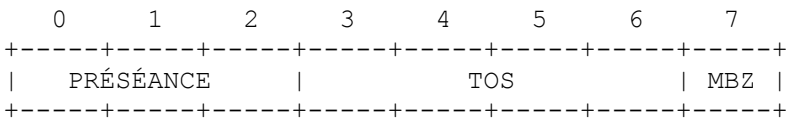
Champ Préséance : les trois bits de gauche de l'octet TOS *(type de service)* d'un en-tête IPv4. Noter que dans DiffServ, ces trois bits peuvent être ou non utilisés pour noter la préséance du paquet IP. Il n'y a pas de champ Préséance dans l'octet

Classe de trafic dans IPv6.

Champ TOS : bits 3-6 dans l'octet TOS de l'en-tête IPv4 [RFC1349].

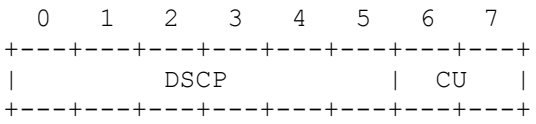
Champ MBZ : Must Be Zero *(doit être zéro)*

La structure de l'octet TOS est décrite ci-dessous :



Champ DS : l'octet TOS d'un en-tête IPv4 est renommé champ Services différenciés (DS) par DiffServ.

La structure du champ DS est décrite ci-dessous :



DSCP : codet de service différencié, les 6 bits de gauche du champ DS.

CU : actuellement non utilisé.

Comportement par bond (PHB, *Per-hop Behavior*) : description chez un nœud conforme aux services différenciés du traitement de transmission observable en externe appliqué à un agrégat de comportement.

3. Description du problème

La manipulation du DSCP pour réaliser le PHB désiré par les nœuds à capacité DiffServ peut entrer en conflit avec l'utilisation par TCP du champs de préséance. Ce conflit peut causer des problèmes aux mises en œuvre de TCP qui se conforment à la [RFC0793]. D'abord, au paragraphe 3.4 de la [RFC0793], le point 2 du sous-titre "Génération d'un signal de réinitialisation" déclare :
"Si la connexion est dans un état non synchronisé (LISTEN, SYN-SENT, SYN-RECEIVED), et si le segment entrant acquitte quelque chose qui n'a pas été encore envoyé (le segment porte un ACK non recevable), ou si un segment entrant est sur un niveau de sécurité ou un compartiment de sécurité ne correspondant pas exactement à celui demandé pour la connexion, un segment RST *(réinitialisation)* est émis. Si notre segment SYN n'a pas été acquitté et si le niveau de préséance du segment entrant est supérieur à celui attendu, on pourra relever le niveau de préséance de la connexion (si l'application et le système d'exploitation l'autorisent) ou émettre un "reset" ; ou si le niveau de préséance du segment entrant est inférieur à celui requis, on continuera à le traiter comme si le niveau de préséance correspondait exactement (si le TCP distant ne peut augmenter le niveau de préséance pour répondre à nos exigences locales, cela sera détecté dans les prochains segments reçus, et la connexion sera alors fermée). Si notre segment SYN a été acquitté (peut être dans ce segment entrant) le niveau de préséance doit correspondre exactement au niveau de préséance local. Dans le cas contraire un "reset" doit être émis."

Cela conduit au problème n° 1 : Pour un module TCP à capacité de préséance, si durant le processus de synchronisation de TCP les champs de préséance des paquets SYN et/ou ACK sont modifiés par les nœuds intermédiaires, résultant en ce que le paquet ACK reçu a une préséance différente de celle prise par ce module TCP, la connexion TCP ne peut pas être établie, même si les deux modules sont en fait d'accord sur une préséance identique pour la connexion.

Ensuite, au point 3 du même sous-titre, la [RFC0793] déclare :
"Si la connexion est dans un état synchronisé (ESTABLISHED, FIN-WAIT-1, FIN-WAIT-2, CLOSE-WAIT, CLOSING, LAST-ACK, TIME-WAIT), tout segment inacceptable (en dehors d'un numéro de séquence de fenêtre ou d'un numéro d'accusé de réception inacceptable) doit déclancher seulement un segment d'accusé de réception vide contenant le numéro de séquence envoyé actuel et un accusé de réception indiquant le prochain numéro de séquence qu'on s'attend à recevoir, et la connexion reste dans le même état".

Cela conduit au problème n° 2 : Pour un module TCP à capacité de préséance, si le champ Préséance d'un segment reçu d'une connexion TCP établie a été changé en route par les nœuds intermédiaires de sorte qu'il est différent de la préséance

spécifiée durant l'établissement de la connexion, la connexion TCP sera réinitialisée.

Chacun des problèmes 1 et 2 a un problème en miroir. Ils sont cause que les connexions TCP qui doivent être réinitialisées selon la [RFC0793] ne sont pas réinitialisées.

Problème n° 3 : Une connexion TCP peut être établie entre deux modules TCP qui prennent des préséances différentes, parce que les champs Préséance des paquets SYN et ACK sont modifiés par les nœuds intermédiaires, avec pour résultat que les deux modules pensent qu'ils sont en accord pour la préséance de la connexion.

Problème n° 4 : Une connexion TCP a été établie normalement par deux modules TCP qui prennent la même préséance. Mais au milieu de la transmission des données, un des modules TCP change la préséance de ses segments. Selon la [RFC0793], la connexion TCP doit être réinitialisée. Dans un environnement à capacité DiffServ, si la préséance des segments est altérée par des nœuds intermédiaires de telle sorte qu'elle conserve la valeur attendue lors de l'arrivée à l'autre module TCP, la connexion ne sera pas réinitialisée.

4. Proposition de modification à TCP

La modification proposée à TCP est que TCP doit ignorer la préséance de tous les segments reçus. Plus précisément :

(1) Dans le processus de synchronisation de TCP, les modules TCP des deux extrémités doivent ignorer les champs Préséance des paquets SYN et SYN ACK. La connexion TCP sera établie si toutes les conditions spécifiées par la [RFC0793] sont satisfaites excepté la préséance de la connexion.

(2) Après l'établissement d'une connexion, chaque extrémité envoie des segments avec sa préséance désirée. La préséance prise par une extrémité de la connexion TCP peut être la même ou peut être différente de celle prise par l'autre extrémité (parce que la préséance est ignorée pendant l'établissement de la connexion). Les champs Préséance peuvent aussi être changés par les nœuds intermédiaires. Dans l'un ou l'autre cas, la préséance des paquets reçus sera ignorée par l'autre extrémité. La connexion TCP ne sera réinitialisé ni dans un cas ni dans l'autre.

Les problèmes 1 et 2 sont résolus par la modification proposée. Les problèmes 3 et 4 deviennent des non problèmes parce que TCP doit ignore la préséance. Dans un environnement à capacité DiffServ, les deux cas décrits dans les problèmes 3 et 4 devraient être permis.

5. Considérations sur la sécurité

Une mise en œuvre de TCP qui termine une connexion à réception d'un segment qui a un champ de préséance incorrect, sans considération de la correction des numéros de séquence dans l'en-tête du segment, présente une menace sérieuse de déni de service, car tout ce qu'un attaquant doit faire pour mettre fin à une connexion est de deviner les numéros d'accès et ensuite d'envoyer deux segments avec des valeurs de préséance différentes ; une d'elles est certaine de mettre fin à la connexion. En conséquence, le changement proposé dans le présent mémoire au traitement TCP procurerait un gain significatif en termes de résilience de cette mise en œuvre de TCP.

D'un autre côté, les règles de traitement plus strictes de la [RFC0793] rendent en principe les attaques en usurpation d'identité plus difficiles dans TCP, car l'attaquant ne doit pas seulement deviner le numéro de séquence TCP initial de la victime, mais aussi son réglage de préséance.

Finalement, les questions de sécurité de chaque groupe de PHB sont traitées dans la spécification particulière de ce groupe de PHB [RFC2597], [RFC2598].

6. Remerciements

Nos remerciements à Al Smith pour sa relecture attentive et ses commentaires.

7. Références

[RFC0791] J. Postel, éd., "Protocole Internet - Spécification du protocole du programme Internet", STD 5, septembre 1981.

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

[RFC1349] P. Almquist, "Type de service dans la suite de protocole Internet", juillet 1992. *(Remplacée par* RFC2474*)*

[RFC2460] S. Deering et R. Hinden, "Spécification du protocole Internet, version 6 (IPv6) ", décembre 1998. *(MàJ par* 5095*,* 6564 *; D.S)*

[RFC2474] K. Nichols, S. Blake, F. Baker et D. Black, "Définition du champ Services différenciés (DS) dans les en-têtes IPv4 et IPv6", décembre 1998. (*MàJ par* RFC3168, RFC3260) *(P.S.)*

[RFC2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang et W. Weiss, "Architecture pour services différenciés", décembre 1998. *(MàJ par* RFC3260)

[RFC2597] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Groupe PHB Transmission assurée", juin 1999. (*MàJ par* RFC3260)*PS*

[RFC2598] V. Jacobson, K. Nichols, K. Poduri, "PHB Transmission expédiée", juin 1999. (*Obsolète, voir* RFC3246) *(P.S.)*

8. Adresse des auteurs

Xipeng Xiao
Global Crossing
141 Caspian Court
Sunnyvale, CA 94089
USA
tél : +1 408-543-4801
mél : xipeng@gblx.net

Alan Hannan
iVMG, Inc.
112 Falkirk Court
Sunnyvale, CA 94087
USA
tél : +1 408-749-7084
mél : alan@ivmg.net

Edward Crabbe
Exodus Communications
2650 San Tomas Expressway
Santa Clara, CA 95051
USA
tél : +1 408-346-1544
mél : edc@explosive.net

Vern Paxson, ACIRI/ICSI
1947 Center Street
Suite 600
Berkeley, CA 94704-1198
USA
tél: +1 510-666-2882
mél : vern@aciri.org

9. Déclaration complète de droits de reproduction

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