SecurityFocus™
Infocus
< **http://www.securityfocus.com/infocus/1802** >

Detecting Worms and Abnormal Activities with NetFlow, Part 2
by Yiming Gong
last updated September 23, 2004

Détection de Vers et d'Activités Anormales avec NetFlow, Partie 2

Traduction française personnelle : Jérôme ATHIAS

rnière mise à jour : 15/10/2004

---

1. Bilan de NetFlow

Dans la première partie de cette série d'articles, nous avons examiné ce qu'est NetFlow et comment il peut être utilisé dans la détection préventive de vers, spammeurs, et autre activité réseau anormale pour les grands réseaux d'entreprises et de fournisseurs d'accès Internet. L'article traitait de quelques méthodes les plus courantes d'analyse basée sur le flow : Top N, Baseline et techniques de Comparaison de Valeurs.

Dans cette seconde et dernière partie de l'article, nous allons nous pencher sur trois méthodes supplémentaires d'analyse de flow, incluant la manière de filtrer nos résultats de flow via les flags TCP, afin obtenir une vue plus granuleuse des anomalies réseau.

2. Les flags TCP pour NetFlow

Une tâche difficile lors d'une analyse base sur le flow est que l'administrateur doit examiner une très grande quantité d'enregistrements de flows. S'il compte juste sur le Top N, le baseline et les méthodes de correspondances, l'administrateur va simplement obtenir une vue grossière des anomalies réseaux. Nous avons constaté à de nombreuses occasions qu'il existe des vers et d'autres activités anormales modérément intensives qui apparaissent indéfinissables parmi l'immense quantité de trafic légitime que l'on trouve particulièrement dans un grand réseau d'entreprise. Ces hôtes malicieux ne sortiront pas du lot dans les listes Top N, à moins que nous connaissions à l'avance quels champs clés et valeurs 'greper' – dès lors ce sont toujours des hôtes malicieux qui doivent être traités.

Afin d'identifier les anomalies de manière plus fiable et efficace, une meilleure façon pour rétrécir les enregistrements de flows est nécessaire. Heureusement, pour la plupart des types de vers basés sur TCP et les autres anormalités, il existe un autre champ utile dans les enregistrements de flow : l'analyse basée sur les flags TCP.

Les vers, de part leur nature à la réplication, sont programmés pour rechercher le plus grand nombre de victimes possible. Typiquement, ils envoient des centaines ou même des milliers de sondes à de grands blocs d'adresses IP dans une très petite période de temps. Si un ver a été conçu pour se répandre via TCP (comme le sont la plupart), durant sa propagation il y aura un grand nombre de paquets TCP SYN correspondants envoyés vers l'extérieur lorsqu'il cherche les services vulnérables sur d'autres hôtes.

2.1 Processus typique de scan SYN d'un ver

Etant donné la manière dont un ver se diffuse en dehors du réseau d'entreprise, il existe trois résultats possibles à son scan SYN:

1. La première possibilité est que l'hôte destination est actif, et que le service vulnérable correspondant qui est ciblé tourne.

Comme nous le savons tous, le dialogue en trois phases qui ouvre une connexion TCP normale fait intervenir :

- En premier lieu un client va envoyer un paquet SYN à l'hôte destination
- L'hôte destination répond par un paquet SYN/ACK
- Le client accuse réception de la réponse de l'hôte destination
- La connexion est établie

La Figure 1 suivante, illustre cet accusé de réception.

Figure 1: l'hôte destination est actif et le port TCP est ouvert

Lorsqu'un paquet SYN arrive sur le port destination d'un hôte, si le port est ouvert, la requête SYN envoyé par le ver obtiendra une réponse – indépendamment du fait que le service tournant sur ce port est vulnérable ou pas. De ce fait, le dialogue en trois phases TCP standard sera complet et les paquets ultérieurs transmis à travers les flags TCP comme PUSH et ACK seront suivis. Dans la première partie de cet article nous expliquions qu'un enregistrement NetFlow V5 contient les OR cumulés des flags TCP dans la connexion 'entière'. De ce fait, en utilisant notre approche de NetFlow, nous pourrions nous attendre à voir une combinaison de flags TCP comme ACK/PUSH/SYN/FIN ou ACK/SYN/FIN dans un enregistrement de flow, allant dans les deux sens.

2. Le second résultat possible pour le scan SYN du ver est que l'hôte destination auquel il tente de se connecter ne soit pas actif, comme montré en Figure2.

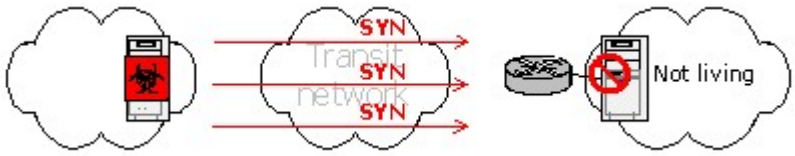

Figure 2: connexion à un hôte destination 'mort'

Du fait que l'hôte destination ne soit pas actif, les requêtes SYN envoyées par le ver ne recevront aucune réponse. Vis à vis de NetFlow, nous obtiendront un enregistrement de flow dans lequel seulement le bit SYN est inscrit de l'hôte vérolé et envoyé à l'hôte destination.

3. Le troisième résultat possible est que le destinataire soit actif mais que les tentatives de connexion du ver soient non fonctionnelles, comme montré en Figure 3.

Figure 3: hôte destination avec port TCP fermé

Cela signifie simplement que l'hôte destination est en effet actif mais que le port auquel le ver tente de se connecter est fermé. Comme nous le savons, pour établir une connexion TCP, un serveur doit écouter sur ce port particulier. Si un client se connecte sur un port non en écoute d'un serveur, le serveur renverra un paquet RST/ACK. Conformément aux guidelines d'implémentation TCP normale, l'hôte cessera immédiatement toute tentative de connexion TCP dès qu'il reçoit un RST. Du point de vue du NetFlow, la combinaison de flag TCP dans les enregistrements de flow ne montrera que les requêtes SYN de l'hôte vérolé vers l'hôte destination.

Jusqu'à présent j'ai exposé trois possibilités lorsqu'un hôte vérolé scan le réseau. Il y a une chose importante à propos des vers à garder en tête : lorsqu'il tente de se propagé, les adresses destination sont généralement générées au hasard, et normalement, il y aura un grand nombre d'hôtes destinations qui seront non actifs ou non fonctionnels. De ce fait, dans le trafic sortant, nous pouvons nous attendre à voir un grand nombre d'octets SYN renseignés dans les enregistrements de flow associés avec l'hôte infecté par le ver. Cette caractéristique est très utile et peut être utilisée comme un point clé pour notre détection de vers basés sur le flow et autres anormalités basées sur TCP. Dans la section 2.2, ci-dessous, nous allons décrire ce processus de détection.

2.2 Trois étapes pour traiter un fichier flow pour les tags TCP

En effectuant une analyse basée sur le flow, un fichier de flow capture peut être traité par les étapes suivantes :

1. La première étape est de chercher dans le fichier de flow, filtrer les enregistrements de flow qui n'ont que le bit SYN défini, extraire les adresses IP de tous les enregistrements de flow, compter les occurrences de chaque IP unique, et finalement trier les enregistrements suivant le nombre d'occurrence de chacun. En suivant cette démarche, nous aboutirons à une liste convenable d'hôtes potentiels. L'administrateur peut mettre en place un seuil en fonction de la taille du réseau et du volume du trafic, ainsi les hôtes dont les compteurs sont au-dessus de ce seuil pourront être considérés comme potentiellement malicieux, et ceux sous le seuil pourront être considérés comme bénins.
2. La seconde étape est de chercher à nouveau dans le fichier de flow pour extraire tous les enregistrements de flow où les adresses IP sources sont celles trouvées dans la liste des « potentiellement malicieux » générée à l'étape 1 précédente. En traitant ainsi une seconde fois le fichier de flow, nous obtiendrons une table détaillée de connexion pour chaque hôte potentiel. Les résultats de cette recherche seront utilisés pour notre troisième et dernière étape dans ce processus, et nous aideront pour mieux identifier le comportement de nos hôtes suspects.
3. La troisième étape va nous fournir des données très significatives sur les hôtes infectés par un ver sur notre réseau. En premier lieu examinez le résultat donné par l'étape 2, puis pour chaque hôte comptez le nombre d'apparitions de chaque port destination unique. Effectuons un tri sur le nombre d'occurrences, nous obtenons une table avec l'adresse IP et ses ports actifs correspondants. Ce qui suit est un exemple de résultat généré par un petit script shell qui effectue cette tâche, écrit par l'auteur.

```
Hôte potentiel1: 61.236.123.225
------
  84 tentatives sur le dstport 1025
  76 tentatives sur le dstport 80
  72 tentatives sur le dstport 2745
  64 tentatives sur le dstport 3127
  48 tentatives sur le dstport 6129
```

Pour un hôte malicieux qui tente toujours de se connecter à un ou plusieurs ports spécifiques, nous pouvons le découvrir dans nos rapports en vérifiant les services enregistrés correspondants pour le port destination de l'hôte le plus actif dans la matrice. En d'autres termes, en utilisant l'exemple ci-dessus, l'hôte 61.236.123.225 a été infecté par le ver W32.gaobot.sa du fait qu'il scanne pour trouver la backdoor (porte dérobée) MyDoom sur le port 3127, la backdoor Bagle sur le port 2745, et le port de Dameware en 6129. Ce sont clairement les caractéristiques du ver W32.gaobot.sa.

Si notre but est d'évaluer la propagation d'un ver de l'extérieur vers l'intérieur de notre réseau, les enregistrements de flow provenant de l'extérieur qui ont les flags TCP RST/ACK définis doivent être examinés. Nous savons qu'un port fermé renverra un RST/ACK à une requête TCP, comme ce fut montré en Figure 3. Si un ver est en train de scanner un grand nombre d'hôtes actifs pour une certaine vulnérabilité, ces

hôtes avec des ports fermés renverront un RST/ACK. Pour le trafic entrant, si un hôte destination (pas source!) dans les enregistrements de flow reçoit trop de réponses RST/ACK, l'administrateur doit vérifier sérieusement cette IP destination, car elle semble bien être infectée par un ver.

3. Problèmes ICMP

Un des objectifs d'ICMP est de fournir un retour sur les problèmes dans l'environnement de communication du réseau. Quelquefois un type/code ICMP dans les enregistrements de flow peut aussi être utilisé pour nous aider à localiser les hôtes malicieux potentiels.

La première chose à noter est qu'il n'y a aucun champ de flow qui est directement nommé comme un type ICMP ou code ICMP, comme se fut déduit dans la partie 1 de cette série d'articles. Certaines personnes ont suggérées, ensuite, que le type et code des requêtes ICMP ne peuvent pas être identifiées dans les données de flow capturées car cette information n'est pas spécifiquement enregistrée. Cela signifie-t-il que nous ne pouvons pas utiliser le type/code ICMP pour notre analyse basée sur le flow ? Non. Dans les faits, nous pouvons : le type et code ICMP sont en fait enregistrés dans les données NetFlow, elles sont simplement stockées dans le champ port destination dans l'enregistrement de flow.

Pour les outils de flow, lorsque l'un d'eux doit obtenir le type et numéro de code ICMP, nous avons juste à vérifier le champ port destination (dstPort). Si le nombre apparaît en hexa, nous devrons le convertir en décimal, et vice versa.

Ce qui suit est un exemple de résultat des outils de flow dans lequel le dstPort est en décimal.

```
srcIP            dstIP            prot  srcPort  dstPort  octets      packets
135.169.9.116    137.54.111.144   1     0        2048     28          1
135.169.9.116    137.62.249.241   1     0        2048     28          1
135.230.255.66   136.129.9.27     1     0        769      112         2
135.32.252.50    136.129.9.27     1     0        769      56          1
```

Nous pouvons voir que le champ protocole (prot) est 1, ce qui signifie ICMP. Le port destination est 2048, ce qui donne 800 en hexa. Ici le 8 signifie type ICMP 8, et 00 est le champ code pour le type ICMP 8, qui signifie pas de code. Ainsi, nous pouvons conclure que 800 est une requête ICMP echo. De la même manière, 769 équivaut à 301 en hexa, qui est le type ICMP 3 et code 01, qui signifie hôte ICMP non disponible.

Il y a deux types de paquets ICMP intéressants qui peuvent être utilisés pour la détection d'activité anormale basée sur le flow lors de l'analyse du trafic entrant d'un réseau. Il est également possible d'utiliser les méthodes de recherche de correspondances pour faire une analyse de flow ICMP, comme nous allons le voir.

3.1 Destination ICMP non disponible

Conformément aux directives de l'implémentation d'ICMP, si le réseau destination ou l'hôte destination est non disponible, la passerelle DOIT envoyer des messages de destination non disponible à l'hôte source, comme illustré en Figure 4.

Figure 4: destination non disponible

3.2 Port ICMP non disponible

Pour les requêtes UDP, les hôtes avec des ports fermés doivent envoyés des messages port indisponible à l'hôte source. Si un ver se répand par UDP, il devrait déclencher beaucoup d'enregistrements de flow port non disponible dans les paquets retournés. Ceci est illustré en Figure 5.

Figure 5: hôte destination avec un port UDP fermé

Si un hôte a un volume anormal de port/hôte/réseau ICMP non disponible dans les enregistrements de flow, cela peut indiquer que l'hôte agit anormalement.

3.3 Méthodes de comparaisons à des modèles

Une autre méthode d'analyse de flow basée sur ICMP est la recherche de correspondances. Certains vers et attaques réseaux sont transmis vers l'extérieur en utilisant ICMP, comme nous l'avons vu avec le ver W32.Nachi.worm. Quand un hôte est infecté par un ver, il va envoyé des requêtes echo ICMP vers l'extérieur avec une longueur fixe de 92 octets. Ainsi nous n'avons simplement qu'à filtrer les enregistrements de flow avec un type ICMP 8 qui ont un paquet long de 92 octets, et les hôtes infectés par ce ver seront débusqués.

4. Zones spéciales dans l'Entreprise

Dans les réseaux d'entreprises, il y a toujours quelques serveurs qui sont protégés avec des pare feux. En principe ces serveurs doivent être protégés et n'ont que des ports déterminés ouverts à l'extérieur. A l'exception de ces ports déterminés, toute connexion établie entre le serveur et l'extérieur doit être interdite.

Nous pourrions utiliser cette caractéristique pour surveiller la sécurité des serveurs en utilisant NetFlow.

4.1 Trafic entrant

Si nous trouvons n'importe quel enregistrement de flow où l'IP destination contient une IP de serveur, mais que le port destination n'est pas dans la liste des ports fonctionnels du serveur et qu'en plus les flags TCP dans l'enregistrement de flow contient un ACK (mais pas un RST/ACK), une alerte doit être déclenchée.

Les suggestions précédentes indique peut être deux points. Premièrement, cela nous dit que le pare-feu devant l'hôte a un problème, du fait qu'il a laissé s'établir une connexion (qui devrait être interdite). Une exception à cela serait que la connexion lancée de l'extérieure contienne incorrectement seulement un paquet ACK ; par conséquent, ce type de connexion n'aurait pas du apparaître. Deuxièmement, l'apparence de cet enregistrement de flow indique également que le serveur semble avoir un port anormal ouvert à l'extérieur !

4.2 Trafic sortant

Lorsque nous voyons n'importe quel enregistrement de flow dans lequel l'IP source contient une IP d'un serveur mais que le port source n'est pas dans la liste de ports fonctionnels du dit serveur, et que de plus les flags TCP dans l'enregistrement de flow ne sont pas des RST/ACK, une alerte doit être déclenchée.

Ainsi, si nous repérons n'importe quelle donnée transférée en même temps que ce qui précède, une alerte rouge doit être déclenchée immédiatement! Il est fort probable que le serveur ait été pénétré. Une backdoor a peut être été activée, et un nouveau service a pu être activé.

5. Directives d'implémentation

Jusqu'ici nous avons parlé de différentes méthodes qui peuvent être utilisées pour la détection basée sur le flow de vers et d'autres activités réseau anormales, cependant aucune instruction d'implémentation n'a été fournie. En réalité, si vous suivez les principaux points de ces méthodes comme décrit dans la série d'articles, l'implémentation serait alors honnêtement très simple.

Pour les instructions sur comment activer NetFlow sur un routeur particulier, les lecteurs peuvent consulter le site Internet du constructeur correspondant. Quelques exemples de configurations NetFlow pour les routeurs populaires Cisco et Jupiter peuvent être trouvés ici: http://www.splintered.net/sw/flow-tools/docs/flow-tools-examples.html

Alors qu'il existe à la fois des solutions commerciales ou open source pour l'analyse de fichiers de flow, l'auteur lui-même préfère la solution open source. Les produits commerciaux ont normalement des fonctions intégrées non évolutives qui sont difficiles à étendre. Le plus important, les outils d'analyse de flow commerciaux n'ont pas la flexibilité des options open source existantes.

Pour les solutions open source nous avons beaucoup de choix, comme cflowd, SiLK, et flow-tools. Tous ceux ci fonctionnent très bien et sur beaucoup de plateformes UNIX différentes.

cflowd
cflownd est l'outil classique d'analyse de trafic flow. Il peut être trouvé sur http://www.caida.org/tools/measurement/cflowd, malgré tout notez qu'il n'est plus supporté par CAIDA du tout, donc considérez un des autres outils suivants.

SiLK
SiLK est une collection d'outils NetFlow développé par le CERT/AC pour faciliter l'analyse de la sécurité dans de grands réseaux. Il consiste en un système packagé et une suite d'analyse. SiLK permet aux administrateurs de traiter les données de flow avec une grande flexibilité, mais de mon point de vue, SiLK nécessite encore quelques révisions et améliorations pour le rendre plus confortable. SiLK peut être trouvé sur http://silktools.sourceforge.net/.

Flow-tools
Flow-tools est un programme puissant et utile pour le travail sur NetFlow. Il existe quelques extensions (add-ons) disponibles, et par dessus tout il fournit une plus grande flexibilité et plus de contrôles que beaucoup des autres outils. Flow-tools peut être trouvé sur http://www.splintered.net/sw/flow-tools/.

En complément de cela, il existe quelques autres programmes comme FlowScan et CUFlow qui peuvent être employés pour un travail d'analyse basée sur le flow. Tous ceux ci peuvent être considérés comme des outils précieux.

6. Résumé

Cette série d'articles a présentée la détection de vers et d'activités anormales basée sur le l'analyse du flow. La première partie parlait du concept basique de NetFlow, puis les deux premières des cinq méthodes d'analyse basées sur le flow furent misent en avant. La seconde partie de l'article a présenté les trois dernières méthodes d'analyse. En résumé, ces trois méthodes d'analyse sont le Top N et le Baseline, la Recherche Comparative, les flags TCP, les problèmes ICMP et les zones spéciales des grandes entreprises.

Il n'y a pas de balle en argent pour la détection de sécurité sur les grandes infrastructures réseau, mais avec NetFlow nous pouvons obtenir un meilleur aperçu du trafic circulant sur l'ensemble de notre réseau – et le faire mieux fonctionner.

A propos de l’auteur

Yiming Gong a travaillé pour China Telecom pendant plus de 5 ans comme administrateur système senior, et maintenant il travaille comme Manageur Technique à China Telecom System Integration Co.Ltd. Il a également une page personnelle axée sur la sécurité système/réseau.

Des commentaires sur cet article peuvent être envoyés à l’éditeur.
