Infocus
< http://www.securityfocus.com/infocus/1796 >

Detecting Worms and Abnormal Activities with NetFlow, Part 1

by Yiming Gong
last updated August 16, 2004

Détection de Vers et d'Activités Anormales avec NetFlow, Partie 1

Traduction française personnelle ; Jérôme ATHIAS
Dernière mise à jour : 17/08/2004

Les réseaux d'entreprise sont de plus en plus confrontés aux problèmes de sécurité comme les vers, les scans de ports, dénis de services (DoS) et abus du réseau, et les solutions de contrôle pour détecter rapidement ces activités sont grandement nécessaires. Les pare feu et les systèmes IDS (intrusion detection systems) sont les solutions les plus courantes pour détecter ces activités, mais d'autres technologies comme le NetFlow peuvent s'avérer être d'une aide précieuse.

1. Vue d'ensemble de NetFlow

NetFlow est une technologie de monitoring du trafic réseau développée par Darren Kerr et Barry Bruins de Cisco Systems, depuis 1996. En tant que standard industriel, NetFlow décrit la méthode utilisée par un routeur pour générer des statistiques sur les paires de sockets routées, et c'est maintenant une fonctionnalité intégrée à la plupart des routeurs Cisco comme le Jupiter, l'Extreme et les routeurs et switchs de quelques autres constructeurs.

Lorsqu'un administrateur active l'export NetFlow sur l'interface d'un routeur, les statistiques sur les paquets reçus sur cette interface sont comptées comme « flow » (mouvement) et enregistrées dans un cache dynamique de mouvements (dynamic flow cache).

1.1 Quesqu'un "flow"?

Un Flow est défini comme une séquence unidirectionnelle de paquets (ce qui signifie qu'il y aura deux flows pour chaque session de connexion ; un du serveur vers le client, un du client vers le serveur) entre deux emplacements. Un flow peut être identifié par sept champs clés : l'adresse IP source, l'adresse IP destination, le numéro du port source, le numéro du port destination, le type de protocole, le type de services, et l'interface d'entrée du routeur. Après chaque réception de paquet, un routeur va examiner ces sept champs puis prendre une décision : si le paquet fait partie d'un flow existant, les statistiques de trafic du flow correspondant seront incrémentées, sinon, une nouvelle entrée de flow sera créée.

Cisco explique que, comme un nouveau flow est constamment créé, les enregistrements de flow expirés vont être exportés sous la forme d'un paquet UDP vers une station de monitoring spécifiée par l'utilisateur si l'un des cas suivants se présentent. Ces conditions

sont les suivantes :

- Le protocole de transport indique que la connexion est terminée (TCP FIN), et il existe un petit délai pour permettre l'achèvement de la transmission du signal FIN.
- L'inactivité du trafic dépasse 15 secondes.
- Pour les flows qui restent actifs en continu, les entrées du flow cache expirent toutes les 30 minutes pour assurer le report périodique des flows actifs.

Bon nombre de constructeurs de matériel réseau ont implémentés leur propre version du NetFlow, mais la Version5 est désormais la plus courante. Pour un datagramme en V5, chaque datagramme UDP contient un entête de flow (flow header) et 30 enregistrements de flow (flow records). Chaque enregeristrement de flow est composé de plusieurs champs, qui incluent : les adresses IP source et destination, la prochaine adresse de saut (hop adress), les numéros d'interfaces d'entrée et de sortie, le nombre de paquets dans le flow, le nombre d'octets total dans le flow, les ports source et destination, le protocole, ToS, le numéro de source et destination AS, et les flags TCP (flags OR ou TCP cumulés).

Sur la station qui collecte, un analyseur de flow est nécessaire pour interpréter les données d'exportation de flow en temps réel. Cela peut être un logiciel ou matériel commercial ou une station utilisant des outils open source.

1.2 NetFlow contre les systèmes IDS

En examinant un enregistrement de flow, vous pourrez constater qu'il n'y a pas d'informations sur la charge utile (payload) du paquet dans le champ flow. C'est là l'une des différences principales entre le NetFlow et un système IDS traditionnel. Un enregistrement de flow ne contient aucune information sur les couches supérieures, il contient seulement des caractéristiques du trafic. Résultant de ceci, cela fait perdre au NetFlow la possibilité fouiller profondément dans les paquets et de faire un travail d'analyse de paquet, néanmoins, il reste suffisamment d'informations pour promouvoir des conclusions valables d'après les données. L'avantage de cette méthode est sa grande vitesse. Ignorer les payloads des paquets réduit grandement le processus et rend NetFlow très adapté aux environnements réseau rapides encombrés. En plus de cela, cette caractéristique rend NetFlow très utile pour la détection d'attaque de type zero-day ou mutante pour des cas où les systèmes de détection d'intrusions basés sur des signatures échouent.

Du fait que les données de flow proviennent directement du routeur, un élément central de tout réseau important, NetFlow est capable de fournir une vue unique du trafic global d'un réseau au niveau de l'infrastructure. Il permet également la détection dynamique d'événements liés à la sécurité sur l'infrastructure réseau.

Si ils sont correctement analysés, les enregistrements NetFlow seront particulièrement adéquats pour la détection préventive de vers ou d'activité réseau anormale dans de grands réseaux d'entreprise et des services de FAI. Dans ce document, je vais décrire quelques méthodes d'analyse de base de flows pour la sécurité réseau.

2. Méthodes d'analyse basées sur le Flow

2.1 Top N et Baseline

Une baseline (ligne de fond) est un modèle décrivant ce qu'est une activité réseau

« normale » en concordance avec un schéma de trafic historisé ; tout le trafic qui sort de ce schéma établi sera considéré comme anormal.

Les rapports d'analyse de tendance et de baseline, couramment reportés comme Top N et Analyse de Baseline, constituent la méthode la plus courante et la plus basique pour analyser les flows. A travers cette approche, l'attention est portée sur les enregistrements de flows caractérisés par des « tailles importantes », en particulier la valeur de ces champs de flows qui diffère de manière significative par rapport à une baseline précédente déterminée.

Normalement, il existe deux façons d'utiliser les méthodes de Top N et de Baseline : les sessions Top N et les données Top N.

2.1.1 Session Top N

Une session Top N signifie qu'un hôte unique produit un volume anormalement important de requête de connexion vers une destination unique ou un ensemble de destinations, et que ce volume s'écarte de la baseline établie. La raison la plus probable pour ce genre d'activités est la présence de nouveaux vers, attaques DoS/DDoS, scans réseau ou certaines formes d'abus du réseau.

Les clients normaux connectés à Internet conservent une fréquence de connexion à l'extérieur relativement normale. Mais si un hôte est infecté par un ver, cela agit d'une toute autre manière. Cela lancera toujours un grand nombre de requêtes de connexion vers l'extérieur pour tenter d'autres victimes, et par conséquent, le nombre de requêtes vers l'extérieur deviendra significativement élevé.

De la même manière, lorsqu'un petit malin peu doué ("script kiddie") est en train de scanner un grand nombre d'adresses à la recherche de certains services vulnérables, nous pourrons constater un grand nombre de sessions vers l'extérieur provenant d'une adresse IP unique.

Nous pouvons également utiliser les méthodes de session Top N pour détecter bon nombre d'abus du réseau, comme contrôler les enregistrements des flows pour les requêtes de connexion sur le port 25 envoyées par chaque hôte en temps réel. Pour une période donnée, pour chaque hôte, si les statistiques des requêtes sur le port 25 sont supérieures à une valeur « normale », l'on pourra le considérer comme spammeur ou comme quelqu'un infecté par un type de ver qui se propage par mail. L'Internet ne s'en porterait que mieux si les fournisseurs d'accès se mettaient à utiliser cette technologie et bloquent les spammeurs grâce à cette détection.

2.1.2 Données Top N

Une seconde méthode pour utiliser les méthodes Top N et Baseline est d'utiliser les données Top N. Cela peut être interprété comme une quantité de données très importante échangée pendant une certaine période entre deux ordinateurs du réseau ou entre un ordinateur spécifique et un ensemble d'adresses.

Les hôtes Top N qui transmettent ou reçoivent des données de l'extérieur dans une entreprise peuvent être catégorisées en plusieurs groupes distincts. Si ce schéma change, et qu'un nouvel hôte apparaisse subitement dans la liste des hôtes Top N, une alerte sera

déclenchée.

Voici un exemple illustrant des méthodes d'analyse des données Top N qui furent utilisées pour identifier un problème de sécurité réseau. Un jour, un de nos clients signala un problème d'utilisation de la bande passante et de congestion. Nous avons rapidement activé NetFlow sur l'interface de son routeur pour enregistrer le trafic sur son réseau, et redirigé les données sur notre station de monitoring. Quelques minutes plus tard, un rapport de flow fut créé. Nous avons utilisés des outils d'analyse de flow pour générer un rapport sur les 20 ordinateurs les plus actifs, triés par volume de données. Lorsque le résultat s'afficha sur la console, nous avons constaté qu'une adresse se trouvant en première position possédait un nombre anormalement élevé d'octets transférés. Une analyse plus approfondie des enregistrements de flow révéla que l'ordinateur en question envoyait un grand nombre de requêtes en destination du port 1434, nous avions alors la réponse. La machine était infecté par le ver SQL slammer, et consumait pratiquement toute leur bande passante. Après que notre client ait patché l'ordinateur vulnérable, la situation de leur réseau revint à la normale.

2.2 Comparaison de valeur

La comparaison de valeur est une autre méthode que nous pouvons employer pour identifier des activités réseau anormales en faisant de l'analyse basée sur les flows. Avec cette méthode, les enregistrements de flows seront inspectés et ceux qui révèlent des champs « suspects » seront marqués.

Tous les champs dans un enregistrement de flow peuvent être utilisés pour effectuer un test de valeur, mais les adresses IP source et destination, ainsi que les ports source et destination sont les plus couramment utilisés.

2.2.1 Test de port

D'une manière générale, afin de lancer une attaque, pratiquement chaque attaque visera un port spécifique ouvert. Par exemple, le ver SQL slammer fonctionne sur le port 1434, le troyen Netbus agit sur le port 12345. Les administrateurs ont la possibilité de filtrer tous les enregistrements de flows où les ports de destination sont égaux à des ports spécifiques, afin d'identifier les attaques correspondantes. Cette méthode est très facile à mettre en pratique et peut être utilisée dans la plupart des cas, toutefois elle peut également générer des fausses alertes.

2.2.2 Test d'adresse IP

La correspondance d'adresse IP est encore une autre méthode qui peut être utilisée pour la sécurité avec l'analyse NetFlow. Il existe différentes manières de réaliser un test de correspondance d'adresse IP comme ce qui suit :

(A) Test d'adresses IANA réservées

L'IANA a réservé de larges plages d'adresses Internet qui ne doivent pas être utilisées pour un routage global. Si nous rencontrons un enregistrement de flow contenant une adresse IANA réservée, une alerte devra être donnée.

Un point important que doit réaliser l'administrateur, est qu'en réalisant des tests de correspondance avec des adresses IANA réservées, il ne pourra pas tracer l'hôte

potentiel à travers l'enregistrement du flow s'il utilise des adresses IP falsifiées (spoofed IP addresses). A ce niveau, un autre champ flow peut être utilisé : Ifindex. Nous pouvons vérifier le numéro Ifindex du routeur correspondant dans les enregistrements de flows pour trouver l'interface actuelle du routeur d'où viennent les flows.

J'ai pu rencontré un cas intéressant où des enregistrements NetFlow de l'un de nos clients semblaient étranges ; les enregistrements de flows faisaient apparaître un grand nombre de connexions pour lesquelles le port d'origine était le 80, les adresses sources étaient 127.0.0.1, et les flags TCP de ces enregistrements étaient tous des RST/ASK.

Ce qui suit est un exemple de rapport extrait d'un outil de flow :

```
Sif  SrcIPaddress DIf  DstIPaddress     Pr SrcP DstP Pkts Octets
StartTime         EndTime           Active  B/Pk  Ts  Fl
0059 127.0.0.1    005b 219.140.194.174  06 50   4f3   1   40
0721.21:58:00.593 0721.21:58:00.593 0.000   40    00  14
0059 127.0.0.1    005b 219.148.205.228  06 50   6ef   1   40
0721.21:57:56.533 0721.21:57:56.533 0.000   40    00  14
```

Nous pouvons voir que le port source (ScrP) est 50 en HEXadécimal, ce qui équivaut à 80 en décimal. Et le flag TCP (Fl) est 14 en HEXadécimal, et en binaire 010100, ce qui le RST/ACK TCP. Alors que l'adresse IP source (SrcIPaddress) est falsifiée en 127.0.0.1, d'où vient l'attaquant ?

En utilisant l'Ifindex (Sif) du routeur dans ces enregistrements, l'interface du routeur d'où proviennent ces paquets est rapidement identifiée. J'informai l'administrateur qui était en charge du réseau su cette interface, et après un moment il me répondit : un PC sur son domaine était piraté et un programme de DoS y avait été installé. Le programme était fait pour lancer des attaques DoS sur le port TCP 80 avec une adresse IP spoofée contre un site de sécurité à Guangdong, en Chine, mais l'enregistrement DNS A du site avait été changé en 127.0.0.1. Ainsi, les paquets de l'attaque étaient reçus par le PC lui-même, puis redirigés sur l'adresse IP falsifiée…

(B). Test d'une IP spécifique ou d'une liste d'IPs

Il existe toujours des règles par défaut pour toute entreprise ou fournisseur d'accès lors de la détection de trafic anormal par l'analyse de flow. Certaines de ces règles sont basées sur :

- Trafic sortant

  Pour une entreprise ou un fournisseur d'accès, n'importe quel enregistrement de flow provenant d'une adresse IP ne faisant pas partie de son domaine réseau se doit d'être considéré comme anormal.

- Trafic entrant

  Pour une entreprise ou un fournisseur d'accès, n'importe quel enregistrement de flow où l'adresse IP source fait partie de son domaine en

trafic entrant, se doit d'être considéré comme anormal.

- Adresses connues

  Certains types d'activités anormales peuvent avoir une ou plusieurs adresses IP connues avec lesquelles un contact est établi. Par exemple, lorsque le ver W32/Netsky.c se propage, il envoie une requête DNS aux serveurs DNS suivants ;

  145.253.2.171, 151.189.13.35, 193.141.40.42, 193.189.244.205, 193.193.144.12, 193.193.158.10, 194.25.2.129, 194.25.2.129, 194.25.2.130, 194.25.2.131, 194.25.2.132, 194.25.2.133, 194.25.2.134, 195.185.185.195, 195.20.224.234, 212.185.252.136, 212.185.252.73, 212.185.253.70, 212.44.160.8, 212.7.128.162, 212.7.128.165, 213.191.74.19, 217.5.97.137, 62.155.255.16

  Ainsi, un quelconque enregistrement de flow dans lequel l'adresse de destination est trouvée dans cette liste et pour lequel le port de destination est également le port UDP 53 devra déclencher une alerte, et une analyse postérieure sera requise.

3.0 Conclusion de la première partie

Ceci conclue la première partie de la série de deux articles. Dans la seconde partie, nous verrons comment filtrer nos résultats de flows via les flags TCP, nous parlerons de quelques problèmes ICMP, puis parlerons de quelques outils différents qui existent pour nous aider à mettre en place et analyser notre solution NetFlow.

About the author

Yiming Gong has worked for China Telecom for more than 5 years as a senior system administrator, and now he works as a Technical Manager in China Telecom System Integration Co.Ltd. He also has a personal homepage focusing on network/system security.

Comments on this article can be sent to the editor.
