TP – Analyse de flux et outils Netflow : Nfdump et Nfsen

Table des matières

Introduction à la surveillance et analyse de trafic:

A l'heure actuelle, les réseaux d'entreprise sont composés de plusieurs types de réseaux interconnectés. Les entreprise utilisent une grande variété de systèmes et d'applications sur ces réseaux. L'équipe d'administrateurs réseau doit être capable de fournir un environnement opérationnel, sécurisé, fiable (peu de pannes, peu d'impact si panne, et rapidement corrigée) et efficace pour supporter les activités quotidiennes de l'entreprise.

De plus, surveiller le trafic à un ou plusieurs endroits dans le réseau d'un opérateur (Orange, etc) permet à l'opérateur de comptabiliser le trafic, d'identifier des applications qui posent problème (comme le P2P), d'identifier la source des problèmes et les corriger, et de faire de l'ingénierie de trafic (comment répartir le trafic sur le réseau).

1 Contexte

1.1 Netflow

Netflow est un outil Cisco qui permet à un routeur d'exporter vers un collecteur des informations sur les flux IP le traversant.

Définition : un flux IP est un ensemble de paquets IP ayant en commun le quadruplet (Src IP, Src Port, Dst IP, Dst Port). Si le temps entre 2 paquets successifs ayant ces propriétés est supérieur à un certain seuil (par exemple 5 min), on considère que les nouveaux paquets font partie d'un deuxième flux. Le seuil est un paramètre à fixer.

Ces informations sur les flux le traversant sont envoyées, sous la forme de trames Netflow (version 5 ou 9) vers un (ou plusieurs) collecteurs selon le schéma suivant :

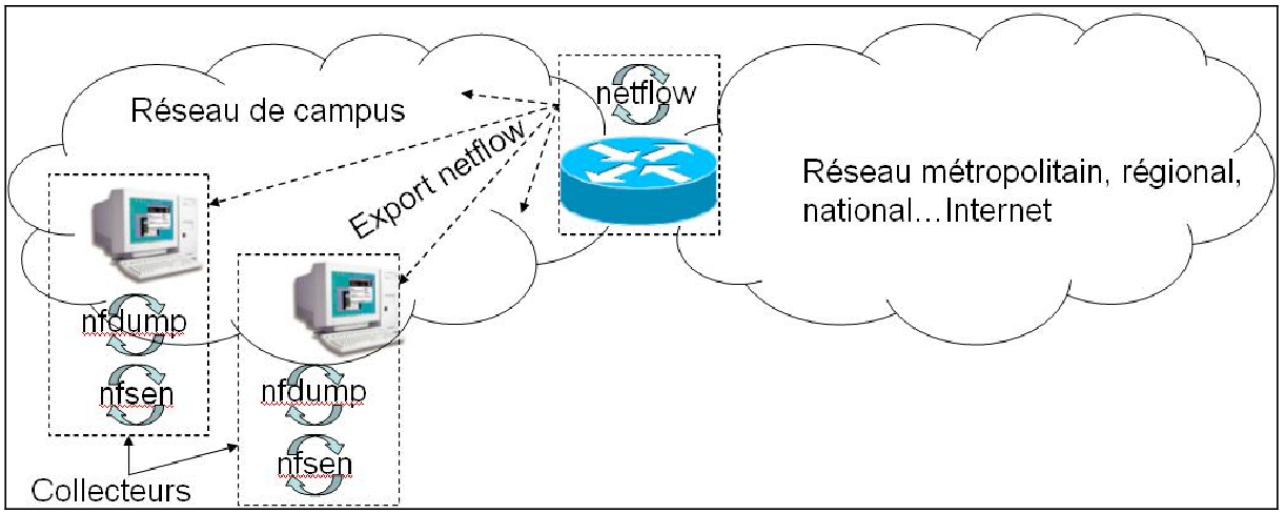

Illustration 1: Collecte de flux avec le module Netflow

L'analyse des flux est réalisée avec le logiciel Nfdump qui fonctionne en ligne de commande. Il y a également une interface graphique à cet outil, qui s'appelle Nfsen (outil graphique se basant sur nfdump). Nfsen sera abordé en second.
(Plus d'infos à : http://nfdump.sourceforge.net/ et http://nfsen.sourceforge.net/ et man pour les options de chacun des outils)

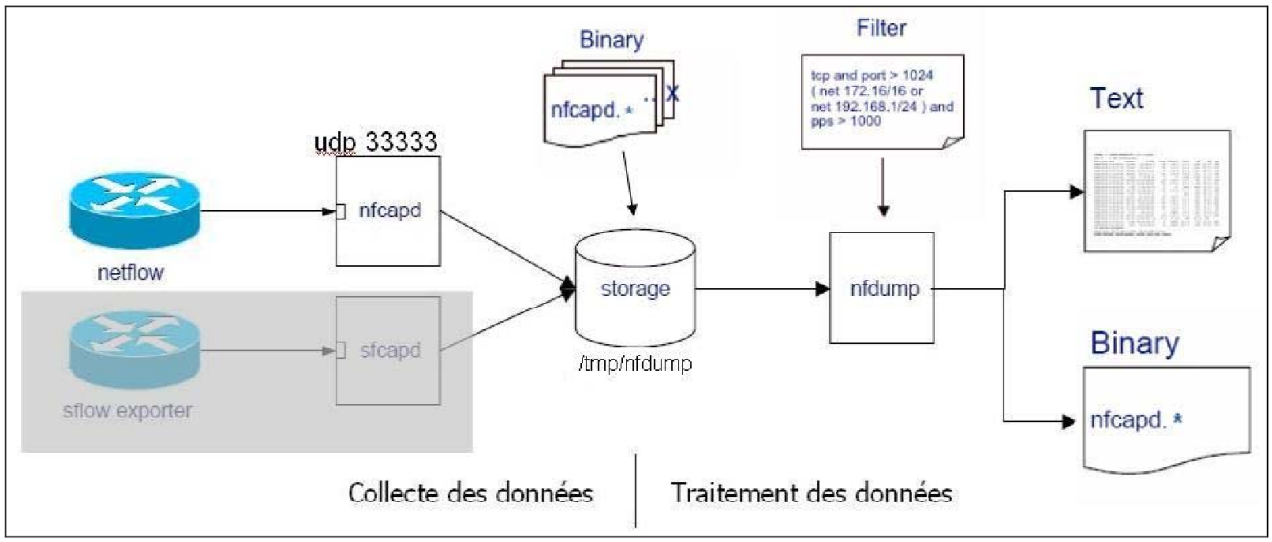

Illustration 2: Module Netflow sur routeur et données Netlow récupérées avec nfdump sur la machine d'administration

1.2 L'outil Nfdump

Nfdump est un ensemble d'outils en ligne de commande permettant la collecte, le stockage et le traitement des enregistrements de flux, compatibles avec netflow v5, v7, v9 et sflow. Les commandes importante de cette suite sont :

- **nfcapd** - Capture des netflow venant du routeur et enregistrement cyclique sous la forme de fichiers. Une rotation automatique sur les fichiers s'effectue (par défaut toutes les 5mn.).

- **nfdump** – Traitement des fichiers générés par nfcapd. Récupère les enregistrements de flux stockés par nfcapd pour effectuer des mesures/statistiques (top N par IP, ports…). (La syntaxe d'utilisation est similaire à celle de tcpdump ou wireshark.)

Les enregistrements de flux collectés sont stockés, par défaut toutes les 5 min, dans un nouveau fichier sous la forme : *nfcapd.YYYYMMddhhmm*. Par exemple, le fichier *nfcapd.200709181140* contient les données collectées le 18 sept. 2007 de 11h40 à 11h45.
Pour distinguer les enregistrements de flux provenant de routeurs différents, on les répartit dans des répertoires différents. Pour le TP, un routeur exportera des trames netflow, ensuite une de vos machine virtuelle exportera des données Netflow. Ce qui fera donc d'abord 1 puis 2 sources.

1. La **collecte** est effectuée par nfcpad en écoute sur le port (UDP 3333 pour le TP) à destination duquel le routeur (ou la machine exportant en netflow) exporte ses enregistrements de flux : `nfcapd –w –D –l /home/nfdump –p 3333`
   D : mode Daemon
   -*w* : permet de faire une rotation des fichiers de manière arrondie. Pour une valeur par défaut de l'intervalle de rotation qui est de 5mn, la rotation des fichiers s'alignera sur 0, 5,10...
   -*l* /tmp/nfdump : répertoire de stockage des données reçues sous forme de fichier nfcapd.*.
   -*p* en écoute sur le port 3333

2. L'**analyse des données** netflow avec Nfdump peut être faite sur un simple fichier (option –*r*) ou sur un ensemble de fichier (option –R) :
   `nfdump –r /tmp/netflow/nfcapd.YYYYMMddhhmm`
   `nfdump –R /tmp/netflow/nfcapd.YYYYMMddhh00:nfcapd.YYYYMMddhh55`

- **Aggrégation de flux** : pour simplifier le format d'affichage, on peut agréger les flux grâce à l'option *–a* qui réunie sur une même ligne les flux ayant les mêmes caractéristiques suivantes : protocole, adresse IP source et destination, port source et destination. Il est possible de n'agréger qu'en fonction de certaines caractéristiques avec l'option –A <scheme>, par exemple n'agréger les flux qu'en fonction de l'adresse IP src et du port destination s'exécute avec l'option `–a –A srcip,dstport`.

- **Filtrage** : En fonction des besoins (Analyse d'incident, détection de scans, pistage d'une machine, métrologie par port/srcip.../tos), Nfdump peut filtrer (sélectionner) les flux affichés. Par exemple, si on ne souhaite afficher que le trafic http à destination du serveur 10.0.2.3 sur une période d'1/2h :
  `nfdump –R /tmp/netflow/nfcapd.Y…hh00:nfcapd.Y…hh30 'dst ip 10.0.2.3 and dst port 80'.`

- **Statistiques Top N** : L'option `–s type[/orderby]` permet de faire des top N sur les enregistrements de flux (où N est configurable avec l'option `–n num`, `–n 0` affiche tous les enregistrements) en fonction d'une caractéristique (*type* : ip, proto, dstip, srcip, srcport…) et de manière ordonnée (*orderby* : décroissant par nombre de flux, nombre d'octets, de paquets…).

2 Architecture réseau considérée

Nous considérons un réseau constitué de 2 VM Debian : VM1 va être la machine d'administration recevant les données Netflow envoyées par un routeur imité par VM2. Précisément, VM2 va être une machine du réseau renvoyant les paquets envoyés par un routeur à une machine d'administration.

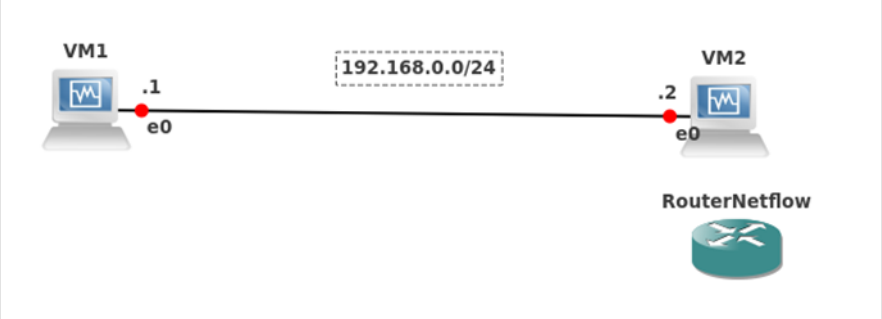

Illustration 3: Réseau pour le TP

Création des machines et configuration réseau :

- Créer VM1 en lui attachant le disque virtuel *Debian_simple.vdi*, téléchargeable depuis http://www.i3s.unice.fr/~sassatelli/Debian_simple.vdi
  Re-générez son adresse MAC.
- Créer VM2 comme clone attaché de VM1 (pour économiser de l'espace disque). Re-générez son adresse MAC.
- Dans la configuration du matériel virtuel, attribuez une carte Ethernet à chaque VM, que vous configurez en NAT ou en accès par pont sur votre carte physique active.

3 Installation de Nfdump sur VM1

Installez pour commencer Synaptic et gcc.

Dépendances à installer au préalable :

Utiliser *Synaptic Package Manager* pour installer les packages (avec les dépendances proposées): rrdtool, apache2 (il va falloir qu'un serveur http tourne sur votre machine), libapache2-mod-php5, flex, librrd-dev, byacc-j, librrdtool-oo-perl, libio-socket-inet6-perl, autoconf.

Installation de Nfdump :

Télécharger Nfdump depuis : http://sourceforge.net/projects/nfdump/
le décompresser et l'installer :

```
tar -zxvf nfdump-1.6.12.tar.gz
cd <repertoire_créé>
```

Passer en mode root

```
./configure --enable-nfprofile --enable-nftrack
make
make install
```

Après l'installation, assurez-vous que les fichiers suivants soient installés. Dans */usr/local/bin*, vous devez au moins trouver : *nfdump, nfcapd, nfexpire, nfprofile, nftrack*.

4 Configuration du routeur Netflow simulé par VM2

La surveillance du trafic est particulièrement importante sur les routeurs de bord, permettant l'accès à l'Internet. Nous allons analyser une telle trace.
Il va falloir télécharger (pas tout de suite) le premier fichier de la trace n°7 de

http://www.simpleweb.org/wiki/Traces
Ce fichier n'est pas une capture réalisée par nfcapd, c'est-à-dire un fichier nfcapd, mais un fichier pcap, correspondant à une capture, réalisée avec tcpdump ou wireshark, des trames envoyées par l'agent netflow du routeur.
Pour analyser le trafic de ce routeur de bord d'université, il va donc falloir d'abord re-créer les fichiers lisibles par nfdump, donc issus d'une capture par nfcapd.
Pour ce faire, vous allez modifier cette trace grâce à l'outil tcprewrite, et ensuite la rejouer grâce à tcpreplay, pour que nfcapd génère des fichiers de trace qu'enfin vous allez pouvoir analyser.

- Sur VM2 (pour la suite) : Télécharger le premier fichier de la trace n°7 de http://www.simpleweb.org/wiki/Traces
- Téléchargez et installez selon les instructions la dernière version de tcpreplay depuis : http://tcpreplay.appneta.com/wiki/installation.html
- Changer la configuration réseau des 2 VM en Internal Network, puis configurez leurs interfaces selon les adresses IP indiquées en figure 3 (pour cela vous pouvez éditer */etc/network/interfaces* puis `ifdown` et `ifup`).

Ce sera de cette VM2 que vous allez rejouer le trafic vers VM1. Pour cela :

- Changement de l'adresse MAC de destination (la destination étant VM1) du trafic de la trace, pour que le trafic rejoué arrive bien sur VM1:
- `# tcprewrite --enet-dmac=@MACdest --infile=netflow --outfile=out1`
- Génération d'un fichier intermédiaire pour changement d'adresses IP :
- `# tcpprep --auto=bridge --pcap=out1 --cachefile=input.cache`
- Changement de l'adresse IP de destination (la destination étant VM1) :
- `# tcprewrite --endpoints=@IPdest:@IPsource --portmap=9500:3333 --cachefile=input.cache --infile=out1 --outfile=out2`
- Rejouer cette trace Netflow avec la commande *tcpreplay* (en mode root).
- `# tcpreplay -i eth0 out2`
- **Afin de vérifier que VM1 reçoit bien les paquets Netflow (et ainsi crée bien ses fichiers nfcapd), lancer un Wireshark (ou tcpdump) sur VM1.**

5 Capture de données Netflow du routeur et filtrage par Nfdump

5.1 Faire fonctionner nfcapd et nfdump

Wireshark
Avant de lancer ces commandes nfdump, assurez-vous que le routeur est bien en train d'envoyer des packets vers votre machine, sur le bon port UDP. Utilisez Wireshark pour cela. Les flux collectés depuis le routeur sont des paquets UDP.

Nfcapd
Nfcapd peut être lancé par tout utilisateur dans le répertoire de stockage voulu : prenez */home/rt/nfcaptures* (à créer).

- `nfcapd -w -t 300 -p 3333 -l /home/rt/nfcaptures &`

-w => crée un nouveau fichier à chaque intervalle de -t *5* minutes (300s), en écoutant les paquets à destination du port *-p 3333* , dans le répertoire *-l output directory.*

Plus d'options avec

- `nfcapd -help`

Vérifiez qu'un processus nfcapd a bien été lancé en écoute sur le port UDP 3333 :

- `netstat -lu -p`

5.2 Manipulation de NfDump pour l'extraction d'information

Pour voir les données capturées par nfcapd :

- `nfdump -r /home/rt/nfcaptures/nfcapd.(timeslot)`

où `nfcapd.(timeslot)` est un fichier particulier, ou

- `nfdump -R /home/rt/nfcaptures/nfcapd.(timeslotdébut):/home/rt/nfcaptures/nfcapd.(timeslotfin)`

Vous pouvez arrêter le replay de VM2 et stopper nfcapd sur VM1 (en mode root, `nstat -lu -program` et `kill -9 <PID>`.)

En vous servant des 2 premières pages de l'énoncé du TP (qui devraient suffire) ou de http://nfdump.sourceforge.net/, et de la commande `nfdump --help` , répondez aux questions suivantes en indiquant dans le rapport les commandes que vous avez utilisées.

- Pour un fichier d'enregistrement de flux de 5mn :
  - Afficher les flux

Q1: `nfdump -r nfcapd._ _ _ _ _ _ _ _ _ _`

  - Afficher les statistiques par protocole (regroupement des flows par protocole de couche 4).

Q2: `nfdump -r nfcapd._ - _ _ _ _ _ _ _`

  - Afficher les flux de manière agrégée et repérer le couple srcip, dstip comportant le plus de flux

Q3: `nfdump -r nfcapd._ - _ _ _ _ _ _ _ _`

Q4: A quoi ces flux correspondent-ils ?

- Sur une période de 10mn : filtres et recherche
  - Afficher seulement les flux à destination du port TCP 80

Q5: `nfdump -R nfcapd._ _:nfcapd._ _ _ - _ '_ _ _ _'`

  - Quelle est l'adresse IP du serveur utilisant le port TCP 23 ?

Q6: `nfdump -R nfcapd._ _:nfcapd._ _ _ _ _ _`

  - A quoi correspondent les flux avec le port UDP 67 ?

Q7: `nfdump -R nfcapd._ _:nfcapd._ _ _ _ _ _ _ _ _ _ _ _ _ _`

6 Installation de Nfsen

Pour éviter de devoir extraire en ligne de commande les informations qui nous intéressent des Netflow, l'outil Nfsen permet d'avoir une interface Web pour interpréter graphiquement les fichiers générés par le processus nfcapd. Il s'agit donc d'installer la suite logicielle Nfsen, ainsi qu'un serveur Web local (Apache) pour au final visualiser les informations de trafic dans une page Web.

Télécharger Nfsen depuis http://sourceforge.net/projects/nfsen/ et le décompresser.

6.1 Créer un utilisateur nfsen et donner à l'utilisateur apache les permissions nfsen

Un nouvel utilisateur et un nouveau groupe sont créés pour permettre à des commandes externes de fonctionner depuis l'interface web.Vous devez être root pour effectuer les commandes suivantes.

- `/usr/sbin/useradd -m nfsen`
- `passwd nfsen` (prenez le login pour mdp)
- `/usr/sbin/groupadd nfsen`
- `/usr/sbin/usermod -G nfsen nfsen`

Create a new group and adding it to apache user group.

- `/usr/sbin/groupadd nfsenadmin`
- `/usr/sbin/usermod -a -G nfsenadmin nfsen` (adding user nfsen to nfsenadmin group)
- `/usr/sbin/usermod -a -G nfsenadmin www-data` (adding user apache to nfsenadmin group)

6.2 Créer un répertoire HTML DIR

Avant d'installer Nfsen, il faut encore créer un répertoire HTML où seront stockées les pages Web représentant les données Netflow.

- `mkdir -p /usr/local/nfsen/www/htdocs/nfsen`

Changer les droits du répertoire à *nfsen:nfsenadmin* pour ne pas avoir de problèmes de droit plus tard.

- `chown -R nfsen:nfsenadmin /usr/local/nfsen`

6.3 Configuration de Nfsen : le fichier *nfsen.conf*

Quand vous installez nfsen, ou changez sa configuration, tout se passe dans le fichier nfsen.conf situé dans le répertoire etc/. S'y positionner :

- `cd /home/rt/Avousdevoir/nfsen-1.3.6p1/etc`

Editez *nfsen-dist.conf* et sauvegardez-le en *nfsen.conf* après avoir fait les changements suivants (ne pas toucher aux autres options) :

- $BASEDIR=”/usr/local/nfsen”
- $HTMLDIR=”${BASEDIR}/www/htdocs/nfsen”
- $USER=”nfsen”
- $WWWUSER=”nfsen”
- $WWWGROUP=”nfsenadmin”
- %sources = (

‘FromVM2’ => { ‘port’ => ‘3333’, ‘col’ => ‘#0f0000’, ‘type’ =>’netflow’ },) ;

Ne pas oublier d'enlever les autres sources (peer1 et peer2).

Faîtes un reconfig de nfsen puis un reload.

Puis :

- `cd ..`
- `./install.pl etc/nfsen.conf`

Et Nfsen est installé sur votre système.

6.4 Configurer l'interface web

Vous devez vous souvenir qu'Apache est un serveur HTTP open-source, qui va être utilisé par Nfsen. Créer un fichier *nfsen.conf* dans le répertoire */etc/apache2/conf.d*. Mettez-y les lignes suivantes :

alias /nfsen "/usr/local/nfsen/www/htdocs/nfsen"
<directory "/usr/local/nfsen/www/htdocs/nfsen">
</directory>
Après avoir sauvé et fermé, redémarrez le serveur web : `/etc/init.d/apache2 restart`

6.5 Faire fonctionner Nfsen

Après l'installation, démarrer Nfsen : `/usr/local/nfsen/bin/nfsen start`

Pour le stopper : `/usr/local/nfsen/bin/nfsen stop`

Relancez, si vous l'aviez arrêté, le replay de la trace depuis VM2.
Et connectez-vous sur *http://localhost/nfsen/nfsen.php*
Cliquez sur le premier graphique.
Les erreurs se voient dans le fichier syslog : tail /var/log/syslog

7 Analyse dans Nfsen

Vérifiez que vous pouvez visualiser votre trafic, et refaire un des filtres de la partie précédente. Explorez les menus.

8 Analyse de traces complexes

8.1 Analyser les données Netflow envoyées par le routeur (VM2)

Utilisez une page Web répertoriant les ports TCP et UDP pour repérer les services dans les question suivantes. De même, aidez-vous de la page de Nfsen, notamment pour la syntaxe des filtres si vous avez des problèmes.

Q8- Avant même que la toute la trace soit rejouée, et au maximum pour les 20 premières minutes de trafic (qui arrive très vite sur votre VM_Netflow1), analysez les fichiers générés par nfcapd, en utilisant Nfsen (dans l'explorateur web que vous aviez lancé sur VM_Netflow1). Positionnez donc les curseurs de la fenêtre du trafic en conséquence.

Q9- En utilisant un filtre, faire afficher (« lister ») les flux à destination du port TCP 80 à l'aide d'un filtre.

Q10- Mesurer et écrire la quantité de bytes de ces flux en utilisant en plus une agrégation sur le protocole.

Q11- Trouver et écrire l'adresse IP des serveurs DNS, Web et NTP générant la majorité des flows grâce aux options de Stat TopN.

Q12- Afficher le Top 20 des connexions (c'est-à-dire les quadruplets (srcip,dstip,srcport,dstport)) par nombre d'octets des flux udp (voir éventuellement page web de Nfdump pour aide sur Top N stats). Ecrire les 2 premières.

Q13- Afficher le Top 20 des services par nombre de flux. Ecrire les 3 premiers.

Q14- A quoi correspondent les 3 premiers ?

Q15- Faire la même chose par nombre d'octets (i.e., quantité de trafic). Que constate-t'on quant au nombre d'octet pour ces 4 services, en comparaison du nombre de flux, et expliquez ?

Q16- Afficher le Top 10 des adresses IP les plus consommatrices en débit. Ecrire la première.

Q17- Donner la nature du trafic émis par la station la plus consommatrice (revenez dans List Flows pour cela).

Q18- Afficher et écrivez le Top 3 des réseaux /24 échangeant le plus de trafic .

Créer un alerte :
En fonction des données netflow reçues du routeur, Nfdump est capable de déclencher des alarmes, configurées au préalable par l'administrateur réseau, et de prévenir celui-ci en cas de déclenchement. Les alarmes peuvent être configurées directement au travers de Nfsen. Documentez-vous sur le page de nfsen pour
Q19 : Créer une alarme se déclenchant si le nombre de flows par seconde dépasse la moyenne vue sur 10min, ou une valeur absolue que vous fixerez (vérifier qu'elle se déclenche bien dans ces conditions).

8.2 Utilisation du plugin php Port Tracker

Pour l'installer, suivez les instructions de */home/rt/AVousdevoir/nfsen-1.3.6p1/contrib/PortTracker/INSTALL*.
Visualisez l'utilisation des ports de façon à répondre aux questions suivantes. Allez sur la page web

de nfsen pour vous aider à identifier les répertoires mentionnés.
Attention : Pour pouvoir afficher les ports souhaités, vous devez désactiver (« *skip* ») les ports 65527 à 65535.

Q20- Quelle est la fraction de http et https par rapport au trafic tcp total ?

Q21- Sur quels 2 ports le trafic udp peut être majoritaire ? Vérifiez votre réponse.

Q22- Parmi smtp, pop, pops, imap, imaps, quel est l'ordre d'utilisation de ces protocoles mail ?