

# 1 La visualisation des logs au CNES

## 1.1 Historique

Depuis près de 2 ans maintenant, le CNES a mis en place une « cellule d'analyse de logs ». Son rôle est multiple :

- Cette cellule est chargée d'analyser les logs suite à un incident afin de comprendre ce qui s'est passé lors de l'incident ainsi que la profondeur de l'attaque.
- Elle est aussi chargée de fournir de l'expertise aux différents projets qui en font la demande en ce qui concerne la journalisation.

Très vite, la cellule a été confrontée à plusieurs problèmes cruciaux lors des investigations sur incident :

- Quel est le contenu des logs fournis pour l'analyse de cet incident. En gros, qu'est ce que l'on trouve dedans (logs firewall, applicatifs, systèmes, sonde de détection d'intrusion).
- Comment filtrer très rapidement les logs utiles à l'analyse de cet incident parmi les quelques mégas octets et millions de lignes fournis en entrée.
- Comment visualiser rapidement ces logs afin de poser les bonnes questions et de fournir les réponses.

Après un temps de recherche, aucun outil donnant une représentation qui paraissait satisfaisante n'a été trouvé. Le CNES a donc choisi de développer ces outils en interne.

## 1.2 Les solutions

Les solutions choisies par le CNES à ces problèmes sont :

- Une base de données. Toutes les données utiles sont injectées dans une base de données afin de pouvoir faire des requêtes SQL.
- Des expressions régulières. Toutes les données sont extraites en utilisant des expressions régulières. Ces expressions régulières permettent de filtrer de manière efficace les logs utiles parmi un volume parfois important de journaux.
- Des outils graphiques. Les vues fournies par ces outils graphiques permettent en effet de s'interroger sur certains phénomènes, trouver des réponses et surtout ne masquent pas la globalité des logs. En effet, si l'on s'enferme dans un scénario prédéfini, on risque de ne plus voir ce qui se passe à côté.

Les paragraphes suivants décrivent les 3 outils graphiques utilisés en interne au CNES par la « cellule d'analyse des traces » dans le cadre de ses missions d'analyse des logs.

## 1.3 Le logiciel CubeCnes

Le programme CubeCnes est un outil de visualisation de logs. Il reprend un développement initial fait au CEA et un concept présenté par Stephen LAU dans son article « The Spinning Cube of Potential Doom » (<http://www.nersc.gov/nusers/security/TheSpinningCube.php>) paru en décembre 2003.

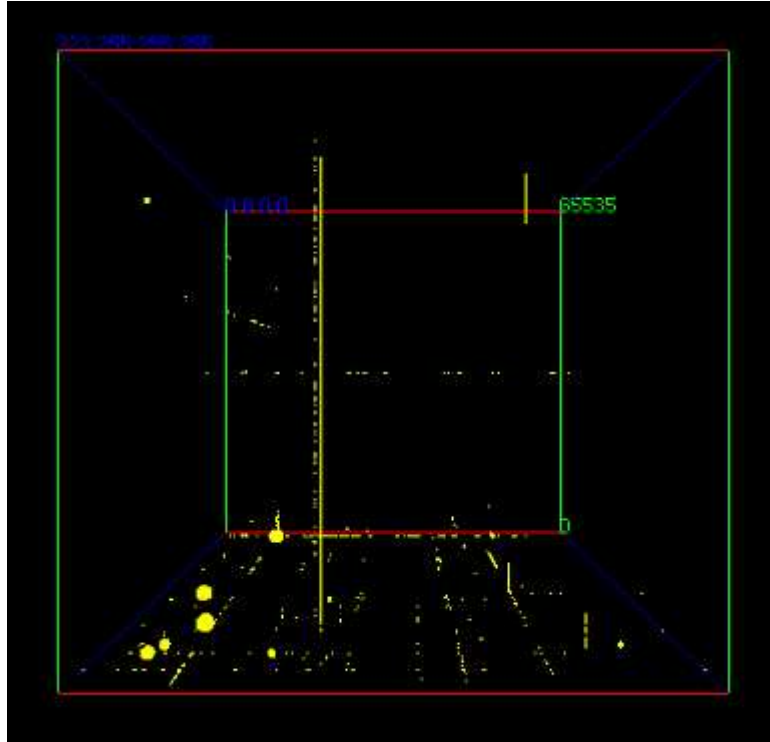
Certaines fonctionnalités intéressantes ont été ajoutées lors du développement au CNES (généralisation du concept de « point », ajout des expressions régulières, ajout des dimensions « couleur », « taille » et « durée de vie » du point, ergonomie de l'interface utilisateur plus intuitive, ...).

Le but de ce programme est :

- d'extraire 6 champs ou valeurs numériques d'une ligne de logs à l'aide d'une expression régulière
- de représenter ces 6 champs sous la forme d'un nuage de points dans un espace à 3 dimensions. Les 6 valeurs sont représentées par les 3 coordonnées « classiques » du point X, Y, Z, la taille du point, la couleur du point et la durée de vie du point.

Le programme CubeCnes est particulièrement adapté à la visualisation de logs issus d'un équipement de filtrage. En effet, les logs générés par ces équipements comportent de nombreuses informations sous forme numérique facilement représentables dans un graphique :

- Date et heure
- Adresse IP source et destination
- Protocole
- Numéro de port source et destination



Cette copie d'écran montre une vue réelle du CubeCnes en train de visualiser les logs d'un firewall. L'axe des X (horizontal) représente les adresses IP internes du CNES, l'axe des Y (vertical) représente le port source de la requête, l'axe Z (profondeur) représente l'adresse IP destination, la couleur représente le numéro de protocole (dans ce cas, UDP est représenté en jaune), la taille du point représente le nombre d'occurrences, le paramètre durée de vie n'est pas utilisé (durée de vie infinie).

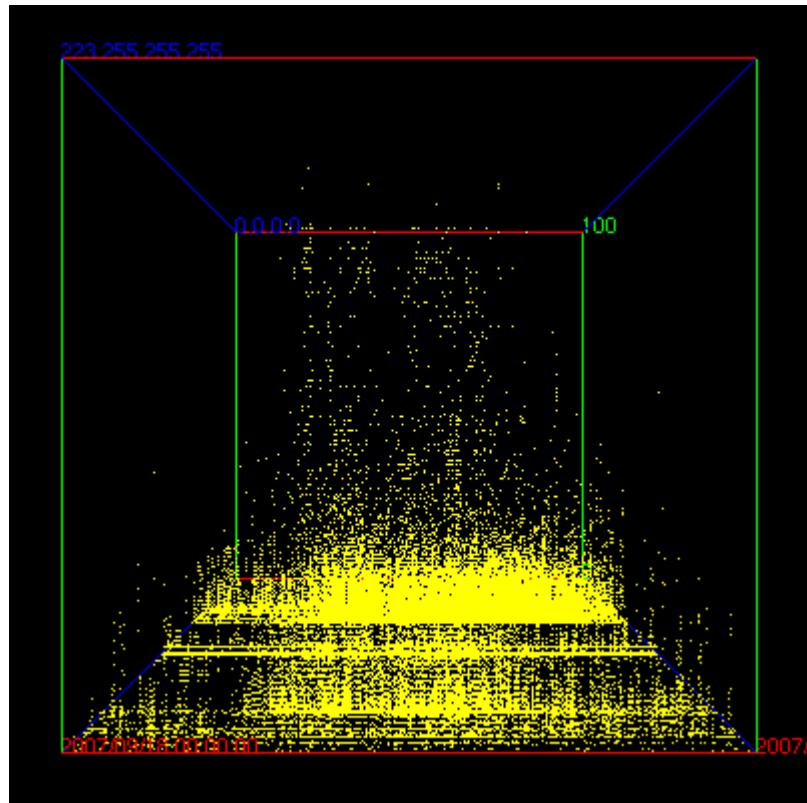
Le grand trait vertical jaune qui apparaît sur le diagramme doit amener l'opérateur à se questionner sur son origine. En fait, après analyse, ce trait est uniquement dû à une seule machine en interne qui était mal configurée (plus de 100 000 rejets sur une journée tout de même). Elle utilisait (ou tout au moins tentait d'utiliser) un serveur DNS externe au lieu d'utiliser le DNS interne. Donc toutes les requêtes sont rejetées par le firewall.

Le CubeCnes peut aussi être utilisé pour mettre en évidence les balayages réseau. En effet, dans un tel cas de figure, si on injecte les logs « deny » d'un firewall, un trait vertical (pour un balayage de ports) ou un trait horizontal (pour un balayage de machines) sera visible sur le cube.

Au CNES, le CubeCnes est aussi utilisé afin de vérifier que la politique de filtrage implantée sur un équipement de filtrage est bien conforme à la spécification de cette politique. Pour cela, on injecte d'une part la spécification de la politique de filtrage (sous la forme d'un fichier Excel transformé en fichier texte) et d'autre part les logs « allow » de l'équipement de filtrage. On configure le cube de manière à ne pas afficher les logs « allow » autorisés par la politique de filtrage (fonctionnalité du CubeCnes). Ainsi, tous les points affichés par le cube sont des flux autorisés par l'équipement mais non autorisés par la politique de filtrage.

Cette fonctionnalité permet de mettre en évidence les flux non spécifiés par la politique de filtrage (mise en place de flux pour des besoins de debug ou encore des erreurs lors de l'implémentation et de l'implantation de la politique de filtrage dans l'équipement par exemple).

Une troisième fonction du cube est la génération en 3D de statistiques :

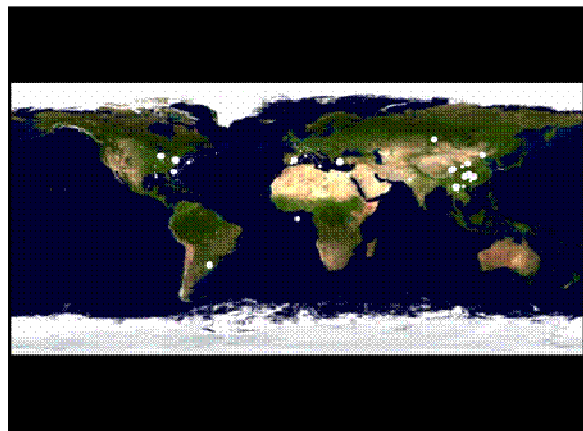
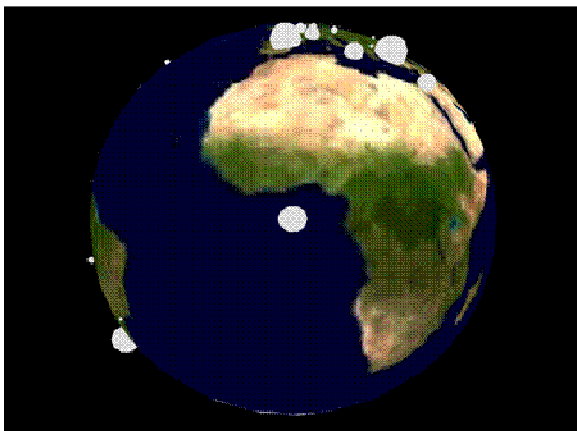


Cette autre vue du logiciel CubeCnes montre le nombre d'accès à un serveur WWW interne du CNES par seconde et par adresse IP externe. L'axe des X (horizontal) représente l'heure (de 0H00 à 23H59) dans la journée du 18 septembre 2007, l'axe Y (vertical) représente le nombre de hits par adresse IP, et l'axe Z (profondeur) représente l'adresse IP du demandeur. Ce graphique montre bien les pics d'activité durant la journée.

On reconnaît aussi (après investigation) le travail des robots d'indexation du WEB (google et autres yahoo) sous la forme des traits horizontaux continus. Toute la journée, ils exécutent des requêtes sur le serveur WWW.

#### 1.4 Le logiciel SphereCnes

Le logiciel SphereCnes a été mis au point suite à une demande spécifique, « Quelle est la source principale en termes de localisation géographique des flux rejetés par le firewall externe du CNES ? ».



Cette vue réelle du logiciel montre une carte terrestre (en 3D ou en 2D) avec la localisation géographique sur la carte du monde des adresses IP sources rejetées par le firewall du CNES. Plus la taille du point est importante, plus le nombre de flux rejetés depuis cette source est important. Les données affichées sont extraites des fichiers de logs à l'aide d'une expression régulière.

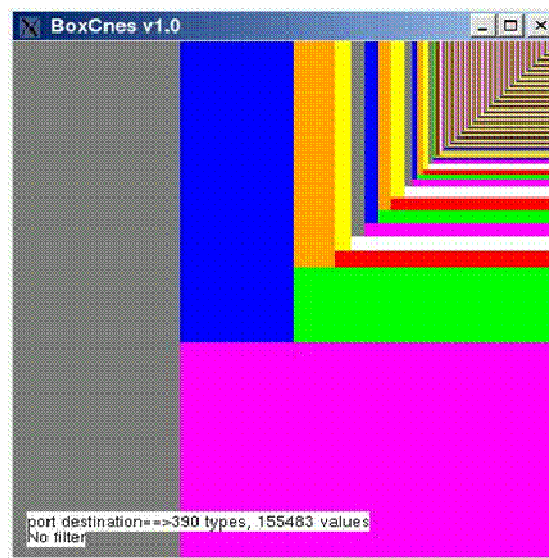
La vue présentée par le logiciel SphereCnes peut être utilisée, par exemple, lors de sessions de sensibilisation à la sécurité pour présenter deux idées : « Internet, c'est dangereux » et « Voici nos agresseurs ».

Le logiciel SphereCnes repose sur la base de données de géo-localisation gratuite Geolp de [MaxMind](#). Cette base de données donne la position géographique sous forme de latitude et de longitude d'une adresse IP quelconque (pour peu qu'elle soit référencée au niveau des bases Whols mondiales). Cette base de données est mise à jour de manière mensuelle. Une base de données plus précises et possédant plus d'informations est disponible mais elle nécessite un abonnement payant auprès de MaxMind.

Le logiciel SphereCnes est particulièrement adapté à la visualisation des logs d'accès à un serveur (FTP, HTTP) ou bien à la visualisation des logs (allow ou deny) d'un firewall.

## 1.5 Le logiciel BoxCnes

Le but du programme BoxCnes est d'afficher rapidement la répartition des données extraites des fichiers de logs par une expression régulière.



La vue offerte par le logiciel permet de connaître rapidement quelles sont les informations les plus présentes afin de se focaliser sur ces dernières dans un premier temps.

Cette vue réelle du logiciel montre que sur les 155483 lignes de logs lues, il y a 390 numéros de ports destination différents et que 4 numéros de ports seulement représentent 75 % du nombre de lignes lues.

## 2 Conclusions

La visualisation de logs permet d'utiliser un formidable outil de corrélation, c'est-à-dire l'œil de l'opérateur. Elle permet aussi de s'affranchir des différents scénarios de détection d'intrusion et donc de garder une vision globale et non tronquée (ce qui n'est pas prévu par les scénarios n'est pas visible).

Par contre, la visualisation de logs demande :

- Une certaine qualité de journalisation (de nombreux logs, des machines synchronisées dans le temps, un niveau de journalisation élevé, ...). Tous les logs peuvent être utiles.
- L'opérateur doit avoir une certaine expérience (voire même une expérience certaine) dans cet exercice.
- L'opérateur doit avoir une connaissance du contenu des logs et de la topologie réseau afin de pouvoir interpréter les résultats affichés. Il doit aussi pouvoir dialoguer avec les exploitants des équipes réseau ou systèmes afin de comprendre ou d'interpréter certains résultats.

Les outils présentés dans cet article (CubeCnes, SphereCnes et BoxCnes) sont développés en interne, ils appartiennent donc au CNES. Vous pouvez prendre contact avec Monsieur Yvon Klein ([Yvon.Klein@cnes.fr](mailto:Yvon.Klein@cnes.fr)) pour des prêts, échanges, partenariats ou d'autres idées encore au sujet de ces outils.

### 3 Quelques URL

- <http://code.google.com/p/davix/>
- <http://www.nersc.gov/nusers/security/TheSpinningCube.php>
- <http://vraf.free.fr/>
- <http://afterglow.sourceforge.net/>
- [http://phoenix.servhome.org/cube6d\\_fr.php](http://phoenix.servhome.org/cube6d_fr.php)
- <http://www.maxmind.com/>