1 Log visualization at CNES (Part II)

1.1 Background

For almost 2 years now, CNES has set up a team dedicated to "log analysis". Its role is multiple:

- This team is responsible for analyzing the logs after an incident in order to understand what happened during the incident and the depth of the attack.
- The team is also responsible for providing expertise to projects that request their logs to be analyzed.

Soon, the team have faced several critical issues during investigations of incidents:

- What are the contents of logs provided for the analysis of this incident? Basically, what we find inside (firewall logs, applications, systems, sensor intrusion detection).
- How to quickly filter logs which are useful for this incident analysis, among the few mega bytes and millions of lines provided as input information.
- How to quickly visualize these logs to ask the right questions and to provide answers.

After some research, no existing tools seemed satisfactory. Then the CNES decided to develop its own tools.

1.2 The solutions

The solutions chosen by the CNES to its problems are:

- A database. All the data are injected into a database in order to make SQL queries.
- Regular expressions. All the data are extracted using regular expressions. These regular expressions can efficiently filter useful log events from a significant volume of logs.
- Graphical tools. The views provided by these graphical tools enable to consider certain phenomena, to find answers and especially to not hide the logs as a whole. If one locks himself in a predefined scenario, it is then impossible to see what is happening aside.

The following paragraphs describe the 3 graphical tools used internally by the CNES "log analysis team" as part of its log analysis missions.

1.3 The CubeCnes program

This program is a tool for viewing logs. It is based on an initial development by the CEA (French Atomic Energy Commission) and a concept introduced by Stephen Lau in "The Spinning Cube of Potential Doom" (http://www.nersc.gov/nusers/security/TheSpinningCube.php) published in December 2003.

Some features have been added during the development by the CNES (generalization of the concept of "points", adding of regular expressions, adding of new dimensions such as "color", "size" and "lifetime" of points, intuitive user interface, etc.).

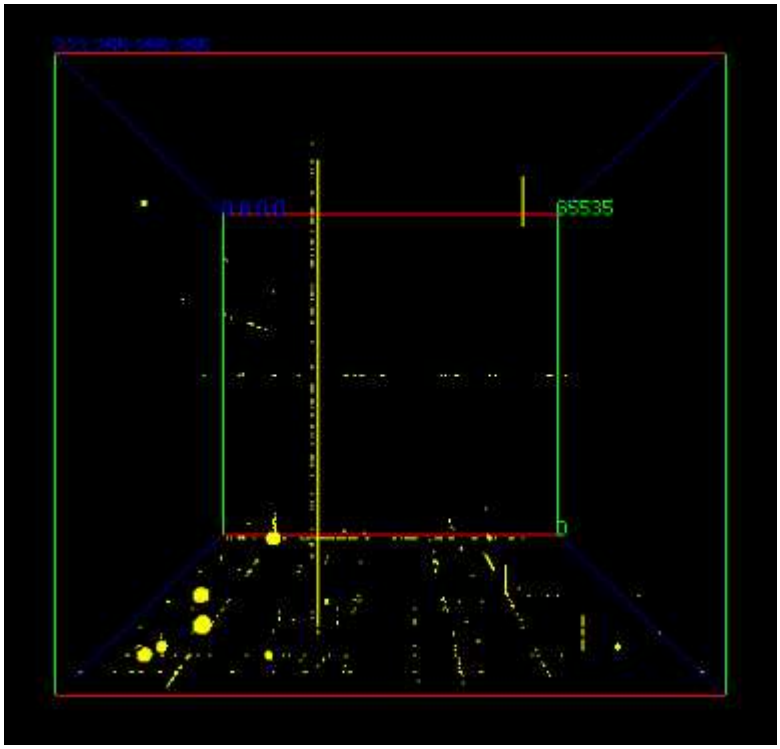
The goal of this program is:

- to extract 6 fields or numerical values from a line of logs with a regular expression,
- to represent the 6 fields by a cloud of points in a 3-dimensional space. The 6 values are represented by the 3 "classical" coordinates of the "X, Y, Z" point, the size of the point, its color and lifespan.

The CubeCnes program is particularly suitable for viewing logs from a filtering equipment. The logs generated by these facilities include a lot of digital information easily represented in a graph:

- Date and Time
- IP source and destination
- Protocol

• Port number source and destination



This screenshot shows a real CubeCnes view while visualizing firewall logs. The X axis (horizontal) represents the internal IP addresses of CNES, the Y axis (vertical) represents the source port of the request, the Z axis (depth) is the IP address, the color represents the protocol number (in this case, UDP is yellow), the size of the points is the number of occurrences, the lifetime parameter is not used (infinite life).

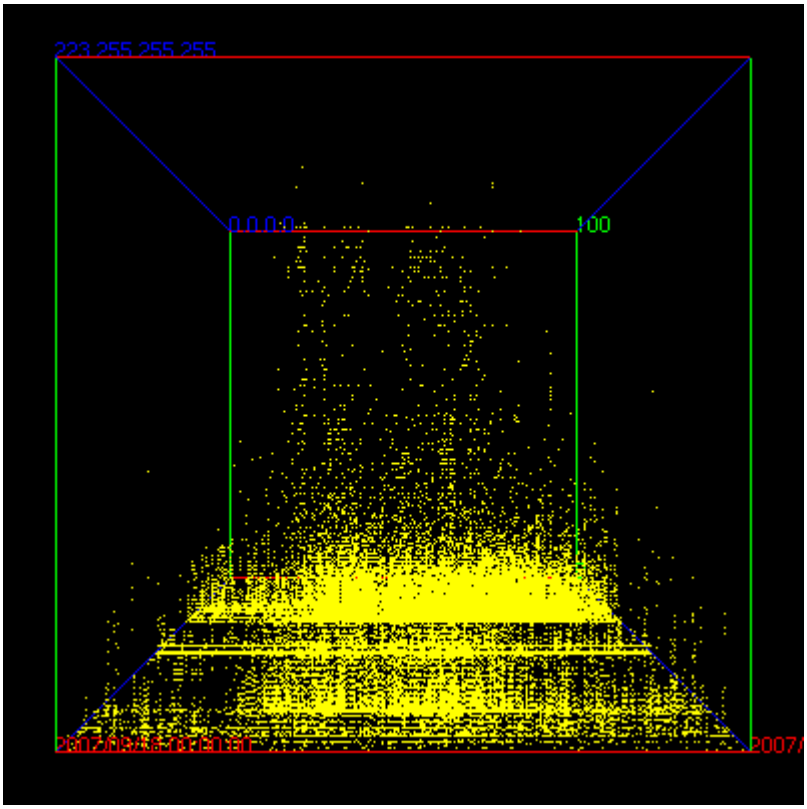
The big vertical yellow line on the diagram should make the operator wonder about its origin. In fact after analysis, this line is only due to a single internal machine which was badly configured (more than 100 000 rejected requests generated a day). It was using (or at least trying to use) an external DNS server instead of the internal ones. All the requests were thus rejected by the firewall.

The CubeCnes can also be used to highlight the network scans. In such a case, if the "deny" logs of a firewall are injected, a vertical (for a port scan) or horizontal (for network scan) line will appear on the cube.

The CNES also uses the CubeCnes to verify that the filtering policy is in accordance with the specification of this policy. To do this, the specification of the filtering policy is injected on one side (in the form of an Excel file converted into text file) and the "allow" logs from the filtering equipment on the other side. The cube is configured not to display the "allow" logs, the ones authorized by the filtering policy (CubeCnes feature). Thus, all the points displayed by the cube are flows authorized by the device but not authorized by the filtering policy.

This feature allows to highlight flows unspecified by the policy filtering.

A third function of the cube is to generate statistics in 3D:

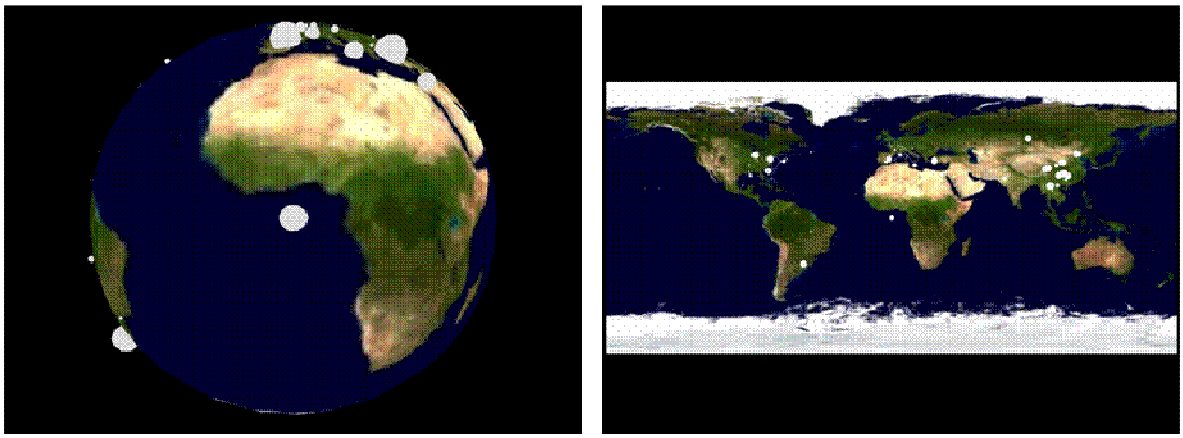


This other view of the CubeCnes program shows the number of accesses to an internal CNES web server per second and per external IP address. The X axis (horizontal) represents the time (from 0:00 to 23:59) during the day of 18 September 2007. The Y axis (vertical) represents the number of hits per IP address, and Z (depth) is the IP address of the requester. This diagram shows the peaks of activity during the day.

After investigation, the jobs of the robots indexing the web (such as Google and Yahoo) can be recognised by the form of continuous horizontal lines. All day long, they make requests on the web server.

1.4 The SphereCnes program

The SphereCnes program has been developed following a specific request, "What is the main source in terms of geographical location of the flows rejected by the CNES external firewall?"



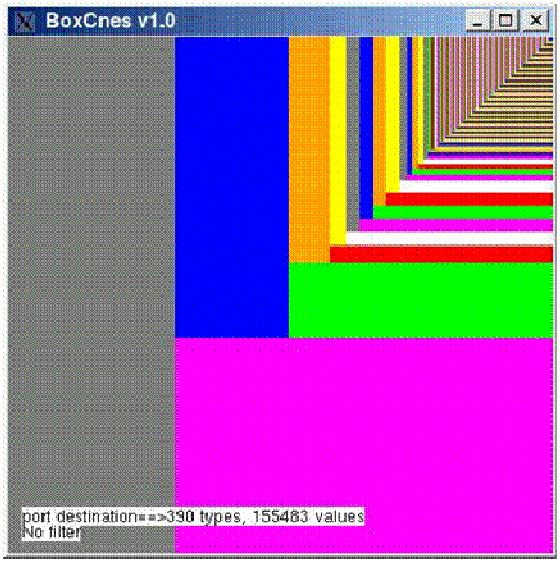
This real view shows a world map (in 2D or 3D) with the location of IP source addresses rejected by the CNES firewall. Bigger is the size of points, greater is the number of rejected flows from a source. Details are extracted from the log files by using a regular expression.

The view presented by the SphereCnes program can be used, for example, during IT security awareness sessions to present two ideas: "The Internet is dangerous" and "Here are our attackers." SphereCnes is based on a free geo-localisation database from "geoip MaxMind". This database provides the geographical position (latitude and longitude) of any IP address (if it is referenced at worldwide WhoIs bases). This database is updated monthly. A more precise database containing more information is available with a paid subscription to MaxMind.

The SphereCnes program is particularly suited for viewing logs of the accesses to a server (FTP, HTTP) or for viewing the "allow or deny" logs of a firewall.

1.5 The BoxCnes program

The BoxCnes program's goal is to quickly view the distribution of data extracted from log files by a regular expression.



The given view allows to quickly know what information is most present to focus on them first.
This real view shows that on 155,483 lines of logs processed, there are 390 destination port numbers and that only 4 port numbers represent 75% of the lines processed.

2 Conclusion

Visualizing logs allows to use a great correlation tool, which is the operator's eye. It also helps not to care of the various scenarios of intrusion detection and therefore to keep an overall view and not a truncated one (what is not expected by the scenarios is not visible).

On the other hand, the log visualization requires:

- A certain quality of logging (many logs, synchronized machines, a high level of logging, etc.). All logs can be helpful.
- The operator must have some experience (and even a confirmed experience) in this exercise.
- The operator must have the knowledge of the contents of logs and the network topology in order to interpret the displayed results. He must also be able to interact with network or system administrators to understand or interpret certain results.

The tools presented in this article (CubeCnes, SphereCnes and BoxCnes) are developed internally and belong to the CNES. You can contact M. Yvon Klein (Yvon.Klein@cnes.fr) for loans, exchanges, partnerships or more ideas about these tools.

3 Several URL of interest

- 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://www.maxmind.com/