3D Graph Visualisation of Web Normal and Malicious Traffic

I. Xydas[1], G. Miaoulis[1], P.-F. Bonnefoi[2], D. Plemenos[2], D. Ghazanfarpour[2]
[1] Technological Educational Institute of Athens, Ag.Spiridona St., 12210 Athens, Greece
{yxydas@teiath.gr, gmiaoul@teiath.gr}
[2] University of Limoges, XLIM Laboratory, CNRS, UMR 6172
83, rue d'Isle, 87000 Limoges, France
{bonnefoi@unilim.fr, plemenos@unilim.fr, djamchid.ghazanfarpour@unilim.fr}

Abstract

Once a web site has been made operational by a company, organisation or individual there is a wish to know the details regarding the connections to the site. In addition, there is a great interest to monitor the activity profile of the web site in terms of how many hits are received, where they come from, the relationship between this activity and increased revenues of the business and so on. Due to the complexity and volume of data involved in these tasks the only way to manage all of the information is to present it using a visual paradigm. Furthermore, web sites are likely to be regularly scanned and attacked by both automated and manual means. Companies, organisations and individuals are making every effort to build and maintain secure web sites. In this paper we will present an ongoing surveillance prototype system which offers a visual aid to the web analyst by monitoring and exploring 3D graphs. The system offers a visual surveillance of the web traffic for both normal and malicious activity. Web requests are presented as 3D directed graphs. Colours are used on the 3D graphics to indicate malicious attempts or anomalous traffic and the analyst has the ability to perform visual data analysis by navigating online into the web request payload, of either normal or malicious traffic.

Keywords: Web visualisation, web security, intrusion detection, web attacks, expert systems, neural networks, anomaly detection, network security, surveillance aid.

1. Introduction

There are an incredible number of web analysis packages available, commercial or research systems, that present information about the content, structure and usage of web sites. Web analysers come in all shapes and sizes. Some are better at representing structure, whereas others are more optimised for looking at the content. Commercial offerings help manage large web sites by providing graphic navigation, analysis techniques and conceptual navigation through data. With web security and intrusion detection however there is a lack of visualisation tools to offload the monitoring tasks, so that anomalies can be easily flagged for analysis and immediate response by the security analyst. Information presented in a visual format is learned and remembered better than information presented textually or verbally. The human brain is structured so that visual processing occurs rapidly and in parallel. Given a complicated visual scene, humans can immediately pick out important features in a matter of milliseconds. They are limited in terms of attention and memory but they excel at the processing of visual information.

The strength of Visualisation to amplify cognition and the lack of visualisation tools for ID analysts have lead us to design such a visual tool and create a prototype system which is described in this paper. It is a surveillance aid for the web and security analyst, offering him the possibility to navigate into the payload of the web request for further analysis and adequate response and providing him with a user friendly visual tool to detect anomalies in web requests by exploring 3D graphs to understand quickly the kind of undergoing attack by means of colours. The system looks into web requests to detect "fingerprints" which are special characters or chains of characters. These fingerprints are then passed to an expert system to decide if they constitute a malicious request or attack [12]. The output of the expert system is then transformed to a 3D graph for visual interpretation and in parallel is kept for statistical analysis. Web attacks can be either rejected by the web server or can be successful due to security weaknesses. If penetration occurs action must be taken by the security analyst as the prototype system does not deal with resolving the damage caused by an attack. It is solely a surveillance device.

The rest of this paper is organised as follows: section 2 presents related work, section 3 describes the visualisation prototype system and section 4 presents some snapshots of normal and malicious web traffic. Finally, concluding remarks and ideas for future work appear in section 5.

Proceedings of the Information Visualization (IV'06)
0-7695-2602-0/06 $20.00 © 2006 IEEE

2. Related work

Numerous research systems have been developed for Web Visualisation. 3D Visualisation of the World Wide Web in [1] reviews recent work in the area of 3D visualisation of web structure, browsing history, web searches, evolution of the web and presence and activities of multiple users. Periscope, a system for 3D visualisation of web search results is described in [2]. Memospace, a visualisation tool for web navigation is presented in [3] and Visual Web Mining is presented in [4]. Visualisation has been used in various areas including security such as the VISUAL system in [5], which is a home-centric Visualisation tool of Network Traffic for security administration. Visualisation has been also used in [6] for a passive visual fingerprinting of network attack tools, such as nmap, superscan, nessus, nikto and others. Visual tools have been also used to visualise logs of IDS systems, such as the SnortView [7], a 2D visualisation system of Snort logs and a Web-based system for Intrusion Detection in [8], which captures the network traffic from the Snort IDS [13] and using a data mining system displays the traffic with a web browser, filtered by source/destination host, protocol or alert, using bar graphs or pie-charts. Interesting work has been presented by Axelsson in [9] where a 3D visualisation has been used to detect web malicious traffic. The logs of a web server were processed and a log reduction system based on frequencies used in order to select the traffic for the visualisation of the web requests and the detection of unauthorised traffic. 3D visualisation is done on preselected traffic, including both normal and malicious traffic and the operator had to navigate into the sub graphs and the graph tails in order to detect malicious or suspect traffic. In recent work [10] Axelsson presented an IDS system based on a bayesian classifier combined with a 2D visual tool called Bayesvis. Finally, it is worth referring to the research on the Information Visualisation systems protection such as presented in [11].

Our work focused on creating an ongoing surveillance tool offering the security analyst a novel visual tool for monitoring and diagnostic needs. We would like to offer an online tool which is capable of dealing with real network traffic in addition to processing stored web logs. Visualisation has been designed in such a way that the operator is not overwhelmed with uninteresting normal traffic. The security analyst immediately detects the malicious traffic by spotting the coloured information on the screen and he has the option of displaying only the unauthorised coloured traffic, by removing the normal black traffic. This allows him to navigate quickly into the web request data for a precise diagnosis and quick response. Malicious traffic is detected by the expert system's knowledge base and a self-organising neural network is used for the web attacks classification. Our approach differs to that of Axelsson's in that we are dealing with real time data in addition to web logs processing, we are using colouring in the 3D visualisation for quick interpretation and diagnosis and furthermore, we have implemented two artificial neural networks, both an unsupervised and a supervised one, for the web class and attack classification. In addition, we have expanded the signature method for ID to detect backdoor intrusions and code execution attempts by high level applications such as HTML, Java, SQL, Perl, Php and Access db. Finally, we must emphasize that the whole system is developed in Linux and all system modules are written in standard C language, offering speed and portability to any operating system and platform, even on small portable computers.

3. Visualisation prototype system

Our visualisation prototype system consists of the following modules: The data capture module, the pre-processor module, the knowledge base module, the statistical analysis module and the graph generator module.

The data capture module selects data either online from the Internet traffic or offline from the web server logs. The pre-processor module examines the web requests to detect malicious traffic and its output is then forwarded to the knowledge base module to predict the type of unauthorised traffic. Then, both normal and malicious traffic are processed by the graph generator module for visualisation. Additionally, all traffic is kept for statistical analysis.

Figure 1. shows the architecture of the visualisation prototype system. Each module is described in detail below.

The two most popular web servers are Microsoft Internet Information Services (IIS) and the open source Apache web server. The IIS web server (versions 4, 5 and 6) of the Library of the Technological Educational Institute of Athens has been used in order to study the various types of attacks and to create the knowledge data base of the system.

Modern web servers offer optional features which improve convenience and functionality at the cost of increased security tasks. These optional features are taken into consideration in our design in addition to traditional types of web attacks (Unicode, directory traversal, buffer overflow, Server-Side Includes-SSI, Cross Site Scripting-CSS, mail and CGI attacks). Different kinds of application insertion attempts are detected such as HTML, JavaScript, SQL, Perl, Access and Php. In addition IIS indexing vulnerabilities, IIS highlight, illegal postfixes, IIS file insertion (.stm), IIS proxy attempts and IIS data access vulnerabilities (msadc) are detected as well. All .asa, .asp and Java requests are tested for URI (Uniform Resource Identifier) legal syntax according to standards, meaning that a corresponding query not in the form <?key=value> is illegal. Trojan/backdoor upload requests

are detected as well. These backdoors are left by worms such as Code Red, Sadmin/IIS and Nimda. Backdoor attempts for apache and IIS servers are detected when web requests are asking for the corresponding password files (.sam and .htpasswd). Finally, command execution attempts are detected for both Windows (.exe, .bat, .sys, .com, .ini, .sh, .dll and other) and Unix (cat, tftp, wget, ls and other) environments.

The web attack classes used, with the associated visualisation colour, are the following:

CMD Unix or Windows command execution attempt (crimson)
INS Code insertions of HTML, Perl, SQL JavaScript, SQL, Perl, Access db (dark orchid)
TBA Trojan backdoor attempt (deep pink)
MAI Different mails such as sendmail, formail, email etc. (forest green)
BOV Buffer overflow (cyan)
CGI CGI scripts (gold)
IIS IIS server attacks (blue)
CSS Cross Site Scripting or Server Side Include (coral)
MISC Miscellaneous, Coldfusion, Unicode and malicious web request options such as PROPFIND, CONNECT, OPTIONS, SEARCH, DEBUG, PUT and TRACE (dark orange).

If the pre-processor detects even one fingerprint its output is forwarded to a neural network for classification. Neural networks (NN) represent a class of very powerful, general-purpose tools that have been successfully applied to prediction, classification and clustering problems. The input vector to the expert system was prepared by the preprocessor which detected the various fingerprints in a web request. The presence of a specific fingerprint is indicated in the input vector as 1 (true) and its absence as 0 (false or unknown).

A total of 30 fingerprints (dimension 30) were used in the model to group all the different types of web known attacks. The used NN is a multilayer network with one hidden layer, using the *generalised delta rule (backpropagation algorithm)* for learning and the sigmoid function as activation function. For applying the gradient descent method to the training of the network we used the *continuous updating* approach. Initially, for the prediction of the network output the "winner-takes-all" method was used, that is the output with the biggest value (rated between 0 and 1) determined the class of the web attack. Later, during the evaluation of the network performance a threshold was used instead of the "winner-takes-all" mechanism. Using a threshold of 0.8 we achieved the best results with a neural network performance of about 92%.

The predicted attack by the neural network is then used to create a coloured directed graph in **dot** form of the well known GraphViz [15] package, using the corresponding *DOT* language. This language describes three kinds of objects: graphs, nodes and edges. A graph G is a tuple (V,E), where V is the set of nodes and E is the set of edges (subset of the Cartesian product VxV). The language has a large number of attributes that affect the graph drawing.

The payload of a web request is cut in nodes and the directed edges are the links between these nodes from left to right. Therefore, a web request from an IP source 217.229.196.17 with payload GET /hact/graphics/ blackwell.jpg has as nodes the words "217.229.196.17", "GET", "hact", "graphics", "blackwell.jpg" and as "directed edges" the links between these nodes from left to right:
217.229.196.17 --> GET --> hact --> graphics --> blackwell.jpg.

When each web request with its IP source address and the requested data is visualised in a 3D graph the web analyst can navigate into the graph for a quick interpretation and evaluation in case of a malicious attempt. Timestamps were not added to the graph as graphs are displayed in real time and the objective here is to keep the display as simple as possible.

There are two graphs generated with the GraphViz package. One graph contains real time traffic, e.g. both normal and possible malicious traffic and the other does not contain normal but only the possible malicious traffic. Normal traffic is visualised in black and malicious traffic in 9 different colours, one for each attack class. This visual separation was necessary because normal traffic overloads the display and the security analyst cannot interpret quickly the malicious attempts. When visualising both normal and malicious traffic the security analyst spends more time navigating through the graph trying to eliminate normal traffic by zooming into the coloured part of the display, than he would if he had only a coloured graph to contend with.

These two *dot* coloured graphs are then visualised with Tulip [14], a 3D graph visualisation tool, supporting various graph algorithms and extensive features for interactive viewing and graph manipulation.

Fig. 2a, 3a, 4a and 5a show normal and malicious web traffic and Fig. 2b, 3b, 4b and 5b only the malicious traffic for the same events. In Fig. 2b the cyan graph indicates a buffer overflow attempt from IP 195.249.40.234, the dark orchid graph a Perl insertion attempt from IP 62.195.136.174 and the deep pink graphs backdoor attempts with PhpMyAdmin and iisadmin commands from IP 66.194.6.80 etc. In addition to these attacks, in Fig. 3b the green graph indicates a formail attempt and the blue graph an IIS attempt. In Fig. 4b the crimson graph indicates a command execution attempt, the deep pink graph a Trojan backdoor attempt from IP 203.163.130.94 and the cyan graph buffer overflow attempts from 4 different IP addresses. Finally in Fig. 5b the deep pink graph shows a backdoor attempt with the recent Linux/Lupper worm aka luppi worm.

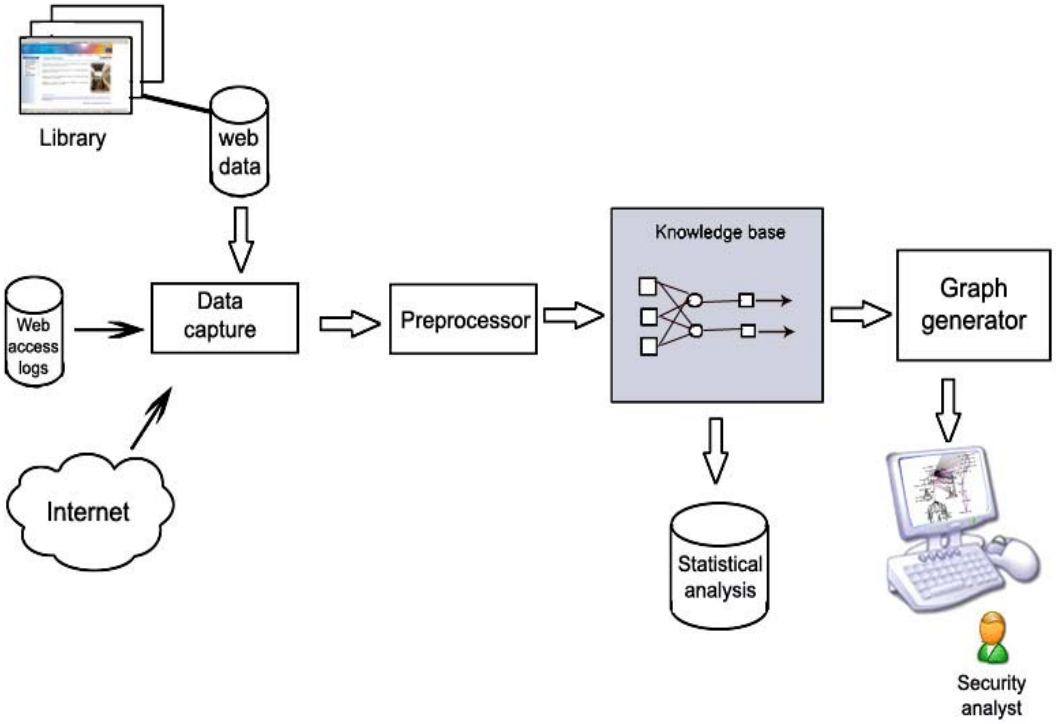

Figure 1. Visualisation system architecture

4. Snapshots

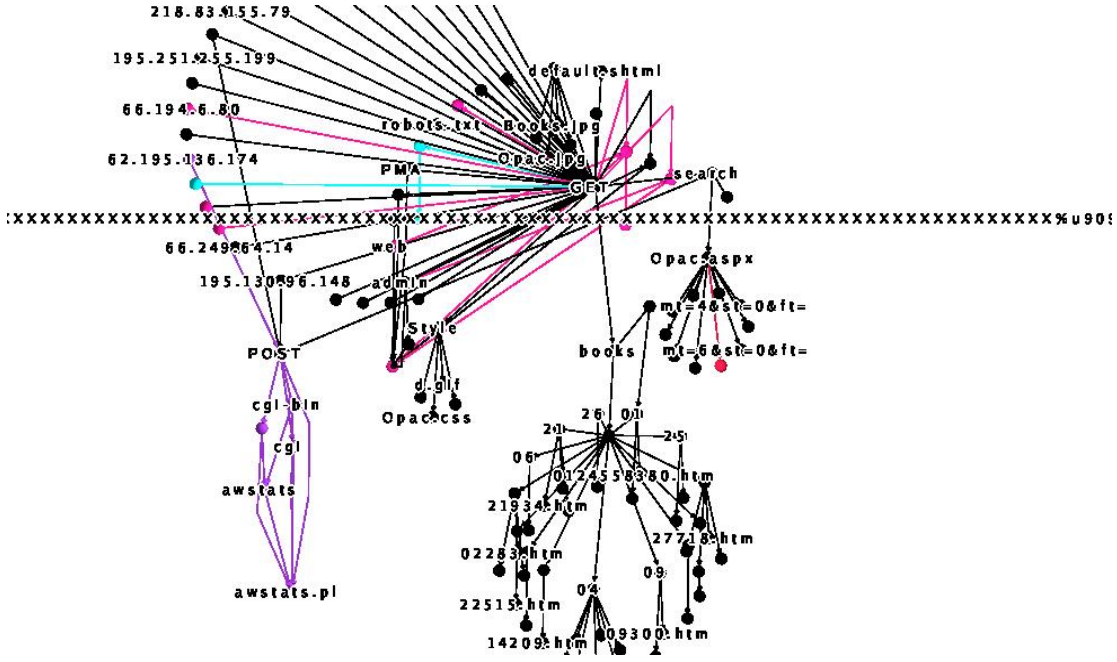

Figure 2a. Normal and malicious traffic (web logs 2005)

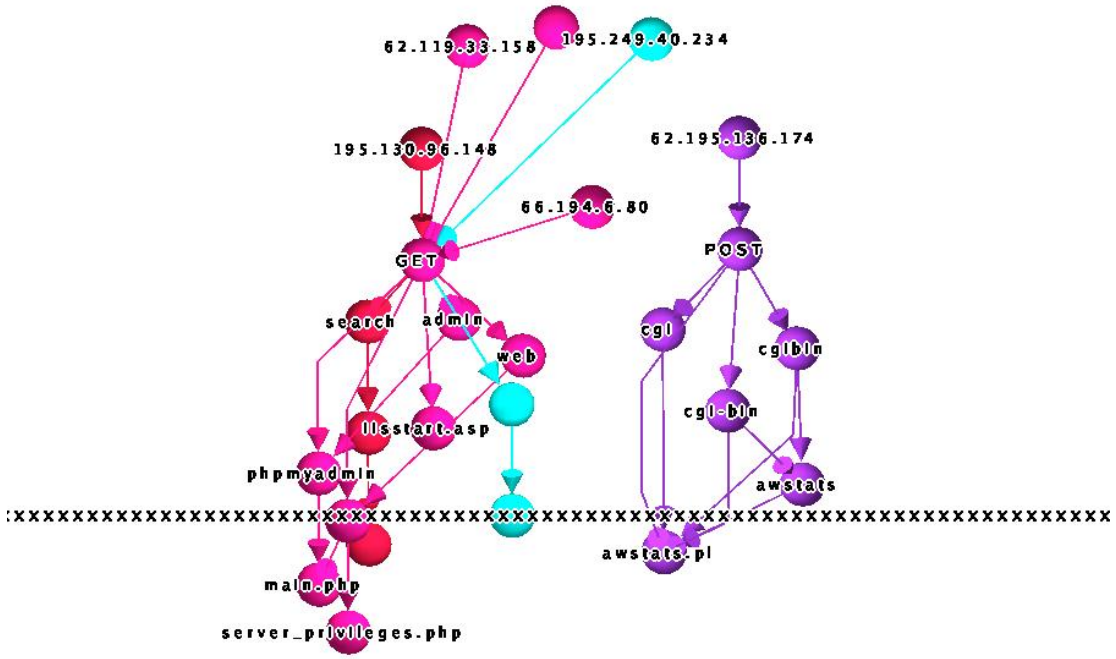

Figure 2b. Malicious only traffic (web logs 2005)

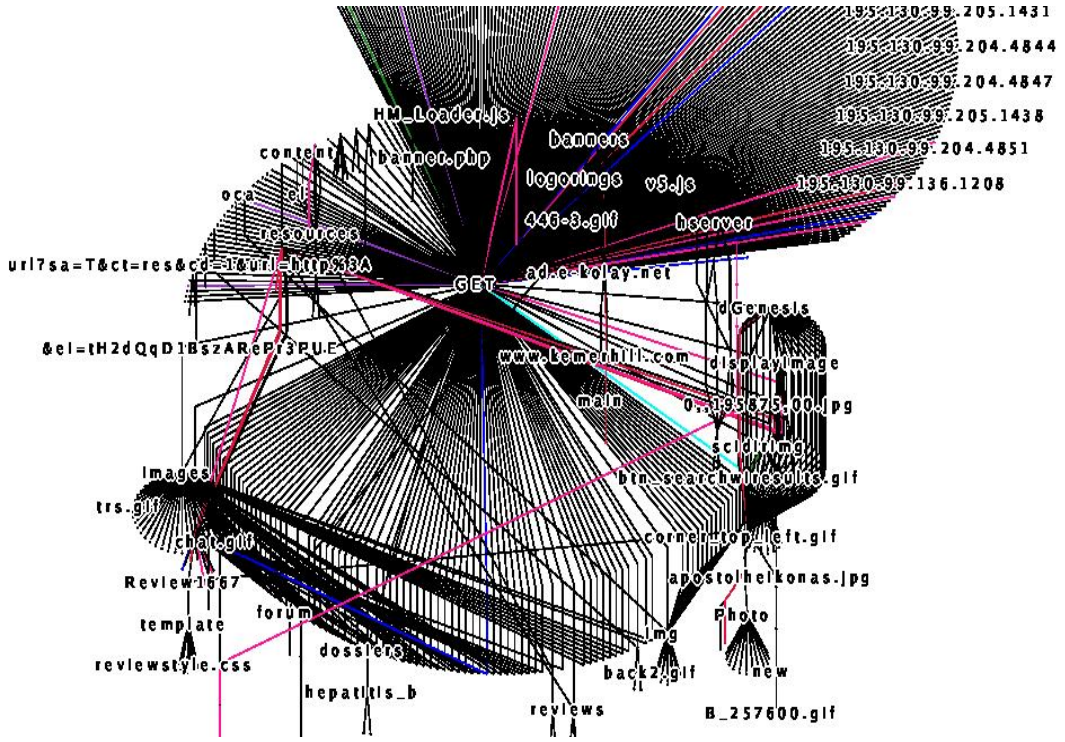

Figure 3a. Normal and malicious traffic (online data 14/6/2005)

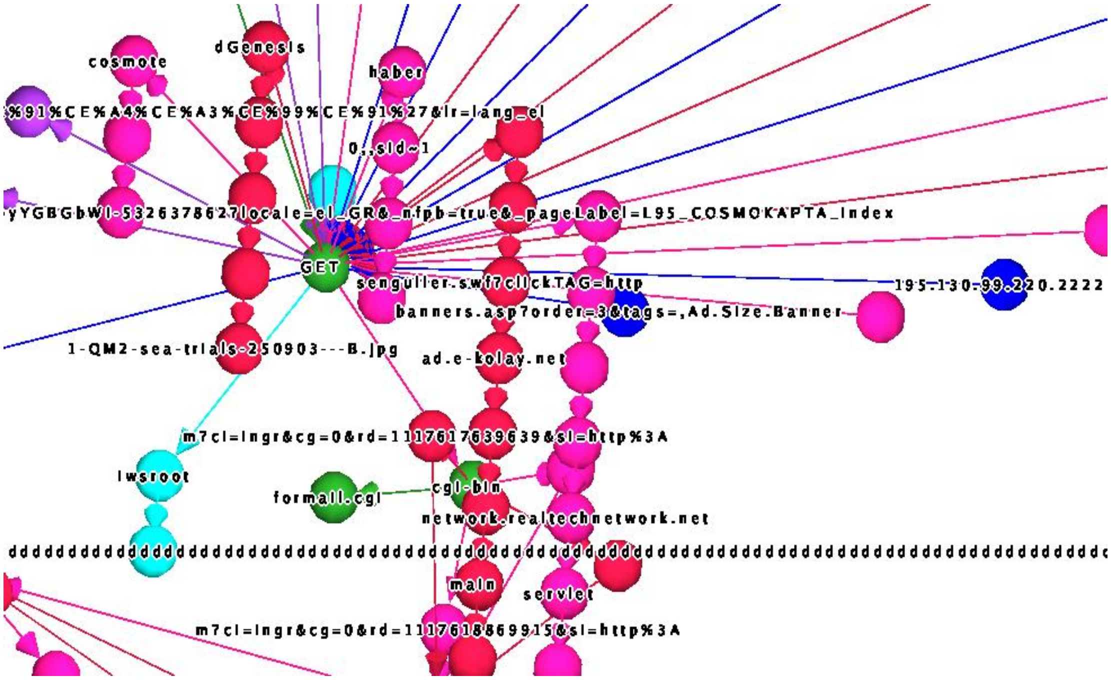

Figure 3b. Malicious only traffic (online data 14/6/2005)

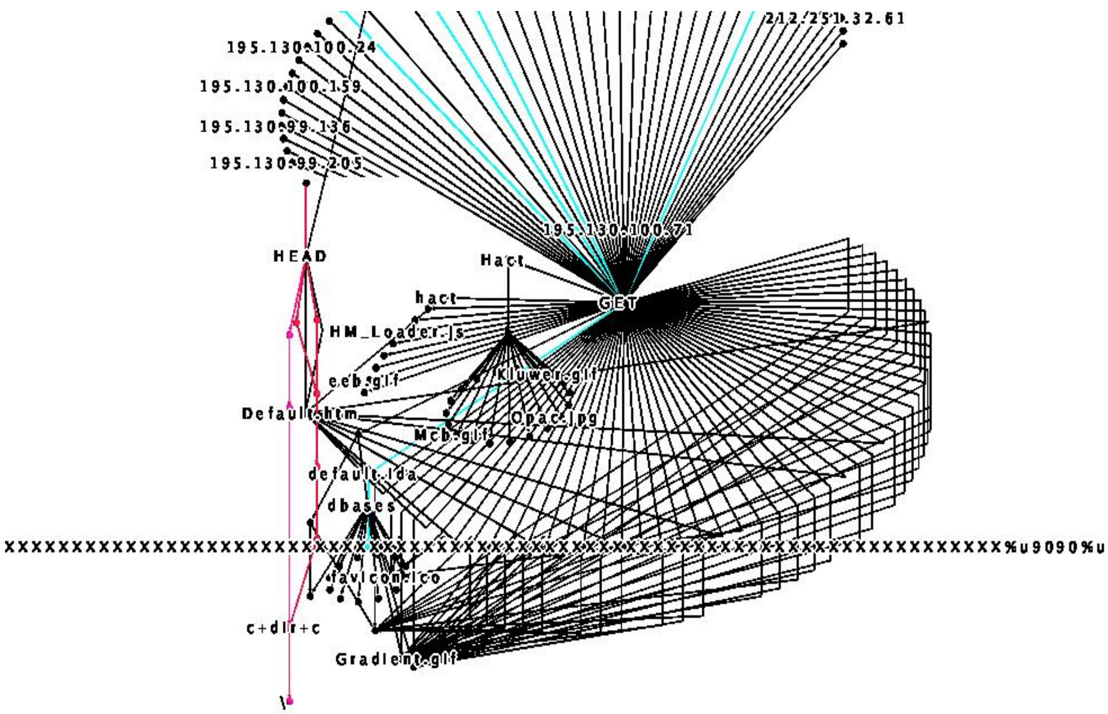

Figure 4a. Normal and malicious traffic (web logs 2003)

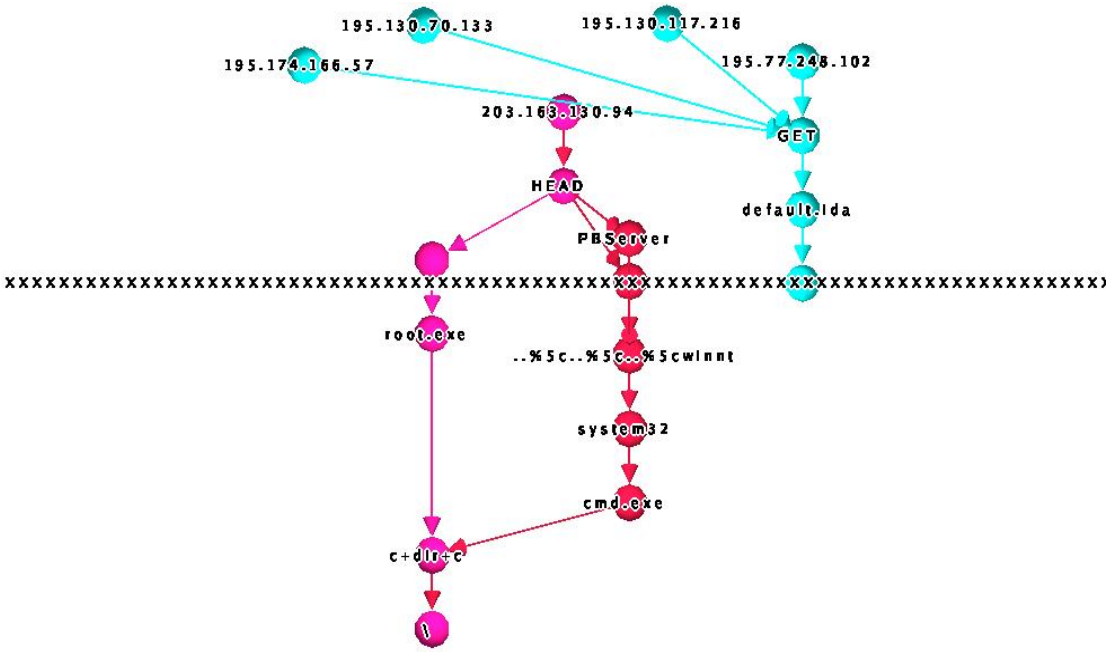

Figure 4b. Malicious only traffic (web logs 2003)

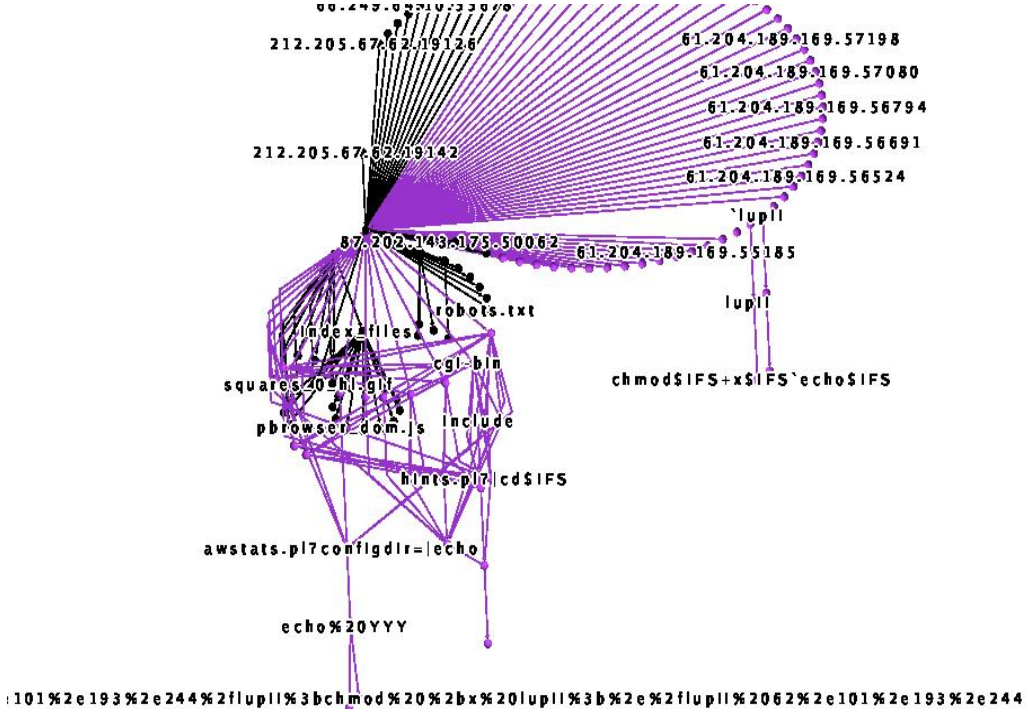

Figure 5a. Normal and malicious traffic (online data 9/11/2005)

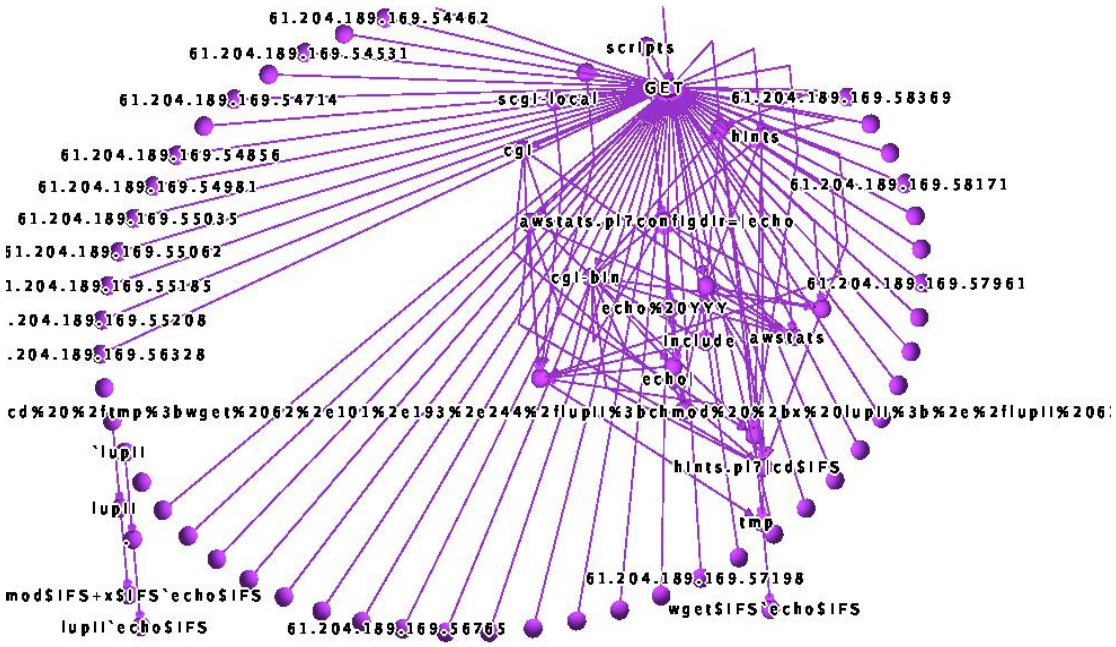

Figure 5b. Malicious only traffic, luppi worm (online data 9/11/2005)

5. Conclusion

Visualisation allows the web analyst to audit the analytical process, since the operator is examining the web traffic directly and online and is making iterative decisions about what is being presented. Visualisation offers a powerful means of analysis that can help the operator uncover hacker trends or strategies that are likely to be missed with other non visual methods. With our work we have contributed to web visualisation the following:

-On going visualisation of the web traffic.

-A new visualisation of web traffic that enables rapid perception and detection of unauthorised traffic.

-Capability to isolate malicious traffic for immediate analysis and response.

-Use of expert system knowledge base for rapid classification of attacks.

-A surveillance aid for the web and security analyst.

-A visualisation prototype system ideal for educational purposes and untrained users to understand web server security state.

Network data analysis is a very important but time consuming task for any administrator. A significant amount of time is devoted to sifting through text-only log files and messages generated by networks tools in order to secure networks. This project has demonstrated that visualisation considerably reduces the time required for data analysis and at the same time provides insights which might otherwise be missed during textual analysis.

The web traffic surveillance could be expanded to other basic but popular internet services, such as mailing or DNS. Combining traditional or novel analytical methods with visual presentation techniques can generate a very robust approach to network security. Visualisation and artificial intelligence can be incorporated in intrusion detection systems to produce more powerful security systems capable of dealing with the new attack challenges.

6. References

[1] S.Benford, I.Taylor, D.Brailsford, B.Koleva, M.Craven, M.Fraser, G.Reynard, C.Greenhalgh. Three Dimensional Visualization of the World Wide Web, ACM Computing Surveys (CSUR), Vol. 31, Number 4es, ACM Press, Dec. 1999.

[2] W.Wiza, K.Walczak, W. Cellary. Periscope - A System for Adaptive 3D Visualization of Search Results, *Proceedings of the 9th international conference on 3D Web technology*, p.29-40, ACM Press, Apr. 2004.

[3] J.Waniek, H.Langner, F.Schmidsberger. MemoSpace: A visualization tool for Web navigation, *Special interest tracks and posters of the 14th international conference on World Wide Web*, p.900-901, ACM Press, May 2005.

[4] A.Youssefi, D.Duke, M.Zaki. VisualWebMining, *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, p.394-395, ACM Press, May 2004.

[5] R.Ball, G.A.Fink, C.North. Home-Centric Visualization of Network Traffic for Security Administration, *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security,* p.55-64, ACM Press, Oct. 2004.

[6] G.Conti, K.Abdullah. Passive Visual Fingerprinting of Network Attack Tools, *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security,* p.45-54, ACM Press, Oct. 2004.

[7] H.Koike, K.Ohno. SnortView: Visualization System of Snort Logs, *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security,* p.143-147, ACM Press, Oct. 2004.

[8] A.Nalluri and D.C.Kar. A web-based system for Intrusion Detection, Journal of Computing Sciences in Colleges, Vol.20 Issue 4, p.274-281, Consortium for Computing Sciences in Colleges (CCSC), USA, Apr. 2005.

[9] S.Axelsson. Visualising Intrusions: Watching the Webserver. *Security and Protection in Information Processing Systems, IFIP 18th World Computer Congress, TC11 19th International Information Security Conference (SEC 2004),* p.259-274, Toulouse, France, Kluwer, Aug. 2004.

[10] S.Axelsson. Combining a Bayesian Classifier with Visualisation: Understanding the IDS, *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security,* p.99-108, ACM Press, Oct. 2004.

[11] G.Conti, M.Ahamad, J.Stasko. Attacking Information Visualization System Usability, Overloading and Deceiving the Human, *Proceedings of the 2005 Symposium On Usable Privacy and Security (SOUPS '05),* p.89-100, Pittsburgh PA, USA, ACM Press, July 2005.

[12] Fingerprinting Port 80 Attacks, A look into web server and web application attack signatures, admin@cgisecurity.com, 2002.

[13] Snort software, http://www.snort.org

[14] Tulip software, http://www.tulip-software.org

[15] Graph Visualization software, http://www.graphviz.org