The Development of a Modular Software Framework for a Distributed Forensic Attack Profiling Network

George Chlapoutakis, University of Teeside
Angus Marshall, University of Teeside

September 1, 2008

Abstract

As Internet Crime in its many forms is increasingly becoming an issue Forensic Investigators are asked to deal with in the context of a criminal investigation, the issue of building a clear profile of both the nature of an attack as well as the nature of the attacker has started attracting the interest of the Forensics community. Profiling has been used to some extent in the Computer Networking and Network Security fields in an effort to understand network traffic flow issues, such as traffic congestion and the identification of instances of Network Intrusions and Denial and Distributed Denial of Service attacks.

Initial approaches to profiling network attacks and attackers have yielded both mathematical models (Marshall et al, 2007) and some experimental approaches to developing forensic profiling tools, as in the case of ForNet (Shanmugasundaram et al, 2003).

An approach is therefore proposed that combines elements of Intrusion Detection theory, Network Profiling, Data Mining, Artificial Neural Networks and BotNet technologies to form a Forensic Attack Profiling tool in the form of a modular and customizable distributed software framework. The framework will allow for the easy development of customized network sensor components and will result in the secure and real-time profiling of host and network-based attacks and the individuals behind the attacks.

1 Introduction

As Internet Crime, in its many forms, is increasingly becoming an issue Forensic Investigators are asked to deal with in the context of a criminal investigation, the issue ofbuilding a clear profile of both the nature of an attack as well as as the nature of the attacker has started attracting the interest of the Forensics community.

2 Profiling in Computer Networks

Profiling has been used to some extent in the Computer Networking and Network Security fields in an effort to understand network traffic flow issues such as traffic congestion, as in the case of Keys et al (Keys, Moore & Estan 2005) who chose to profile Internet traffic using IP traffic summaries computations which, through automatic adaptation of the summarisation algorithm, allow the proposed system to gracefully degrade, while still achieving an acceptable monitoring and profiling standard, the accuracy of its results as a response to increased workloads or Denial of Service attacks, or in order to classify traffic by applications used such as in the case of (Karagiannis, Papagiannaki & Faloutsos 2005) whose research focused on application profiling and identification from internet traffic flows by observing them at the transport layer and identifying three layers (social, functional and application level) of increasing detail.

Network profiling was also the preferred approach of Xu et al (Xu, Zhang & Bhattacharyya 2005, Xu, F. Wang & Zhang 2007) who concentrated their research on the wider area of Internet backbone traffic profiling, looking to develop extraction and communication pattern identification techniques to identify attacks and applications affecting internet traffic dynamics as well as measure and identify instances of Denial and Distributed Denial of Service attacks and their effects on computer networks.

3 Profiling in Intrusion Detection

Research in Intrusion Detection and Intrusion Detection Systems, which is the field closest to Digital Forensics, has concentrated more on correlating a set of information to detect attacks a system or network(Amoroso 1999, Bace & Mell 2001).

One of the many approaches to utilise both Self-learning and Programmed analysis that was used for both Anomaly and Misuse-based Intrusion Detection Systems, researched by Dao et al, and later on by Anderson et al (Anderson, Selby & Ramsey 2007), was the use of Profiling. Dao et al state that Profiling is defined as

> “ a classification procedure that groups pertinent information of an event or situation together so that people can make better decisions pertaining to that event or situation.” (Dao, Vemuri & Templeton 2000)

An example of the viability of this approach is Dasgupta and Brian (Dasgupta & Brian 2001) research which used Adaptive Resonance Theory (ART-2) neural networks and fuzzy logic, the first for pattern classification and the latter for decision/action resolution, to create a Profile-based Anomaly Detection system.

Other, non Artificial Neural Network approaches to profiling have also been used, such as Durgin and Zhang’s research in profile-based anomaly detection (Durgin & Pengchu 2005) who were tasked to investigate adaptive anomaly detection techniques using profile-based methods to address the problem of recognising and evaluating threats aimed at complex and heterogenous computer networks.

Papers discussing implementations of Bayesian Inference and Bayesian Forecasting models in Intrusion Detection Systems are the research conducted by Puttini et al (Puttini, Marrakchi & M 2003), where a combination between a Parametric Mixture Model and an EM-Algorithm are used as the classifiers for anomalies in data patterns, as well as the research conducted by Sebyala et al (Sebyala, Olukemi & Sacks 2002), where the use of a Naive Bayesian Network is proposed as an Anomaly Detection engine for an IDS solution to monitor proxylets in Active Networks.

4 Profiling in Digital Forensics

4.1 Intrusion Detection versus Digital Forensics

Intrusion Detection Systems have been able to integrate increasingly automated and low-level computational models for data collation and analysis through the use of Artificial Intelligence, Data Mining, Statistical and Bayesian Inferential techniques and methodologies, making them at first glance ideal as collation, analysis, profiling and scoring platforms for a Digital Forensics investigator handling a Computer or Internet-related crime.

As reported by Sommer (Sommer 1999) and later Stephenson (Stephenson

2000), however, the NSTAC Network Group Intrusion Detection Subgroup's findings identify that the lack of information integrity protection and IT Security employee training in the collection for forensically appropriate information for subsequent use in a legal investigation as the main issues that discourage the use of IDS logs as evidence. Furthermore, while Yull et al (Yuill, Wu, Gong & Huang 1999) have found that Intrusion Detection Systems can collect enough information to clearly profile and possibly identify an attacker without compromising the integrity of the Intrusion Detection System itself, thereby allowing for forensically sound log files of undamaged integridy to be produced, Levitt and Laskey's research into computational inferences for evidential reasoning (Levitt & Laskey 2000) further identifies the need for such systems to be able to clearly demonstrate evidential reasoning, clearly show the timeline in which the offence took place and the associated actions of the offender, and be able to infer alternative hypotheses during the interpretation of the evidence in a legal context.

The inability of Intrusion Detection Systems, therefore, to provide forensically sound information, allowing for evidential reasoning and able to show a clear timeline for the incident, lead to various methodologies and models that either extended existing systems or proposed new ones to satisfy the above requirements so as to be usable in a Digital Forensics investigation.

4.2 Digital Forensics Profiling: Approaches To Date

In 2001, Yasinsac and Manzano's paper (Yasinsac & Manzano 2001) on the policies required to enhance computer and network forensics identified the need for traffic-related information to be retained as the origin and destination addresses in the header of the packet can reveal possible geographical locations of the attackers. This need for Internet traffic-related information is further supported by Marshall and Tompsett's research on Internet Crime (Marshall & Tompsett 2002), who also identify the need for automation with respect to the collation and efficient analysis of such information in order to assist intelligence agents, and later on by Mohay's research on technical challenges and directions for Digital Forensics (Mohay 2005).

Furthermore, Casey (Casey 2004, Casey 2006), Mohay (Mohay 2005) as well as Hannan et al (Hannan, Frings, Broucek & Turner 2003) before them all agree that one of the most important areas of focus in Digital Forensics should be to increase the level of analysis in proportion to the speed of the tools developed for purposes of network-related evidence collection, collation and analysis.

To this effect, Shammugasundaram et al's (Shanmugasundaram, Memon, Savant & Brnnimann 2003) research in Digital Forensic Profiling proposed a distributed network logging mechanism, ForNet, to aid in the investigation of network attacks. The proposed system, primarily a hardware-based solution, uses IP traffic summaries computations and synopsis techniques used by Data Stream Management Systems to build and store summaries of network events through the SynApp component of the system, which are queried by investigators through a querying system in the Forensic Server component.

Payer's research (Payer 2004) attempts to bridge Network Intrusion Detection and Digital Forensics through the use of network stack monitoring host-based Intrusion Detection System specially hardened to ensure data integrity protection and extremely accurate network stack monitoring, the latter allowing the system to generate simple attacker profiles through use of the short-time history of currently used connections, the view that evidentiary information can be a sequential collection of normal, malformed and forged sources packets.

Almulhem and Traore's research in engineering a Network Forensic System (Almulhem & Traore 2005) proposed a system that records both captured network packets and IP-related information in both a host and network level, combining passive packet capture, deep packet analysis and the creation and use of a history of suspicious IP addresses to aid the packet capturing in marking packets as suspicious based on the history of the IP address from which the packets originate.

Finally, Kahai et al (Kahai, Srinivasan, Namuduri & Pendse 2006) propose a forensic profiling system again of a distributed sensor and centralised server architecture, with the sensor modules acting as Intrusion Detection Systems. The IDS nature of the sensor modules allows them to detect an attack on the workstation/node they are placed in, upon occurence of which the sensor modules notify the central server of the occurence of an attack, sending a log of all captured information relevant to the attack to the server.

5 The Distributed Forensic Attack Profiling Network

This research comes under the aegis of the Cyberprofiling Project and proposes an approach that combines elements of classic Forensic Profiling, Intrusion Detection theory and Network Profiling, which were discussed earlier, as well as Artificial Intelligence and Data Mining theory and methodology as well as “Black/Gray-Hat Hacking” practices such as certain elements from a type of distributed network applications that are currently used in certain forms of Internet Crime and are known as Botnets, as discussed in (Chiller, Binkley, Harley, Evron, Bradley, Carsten & Cross 2007, Ianelli & Hackworth 2005, Bacher, Holz, Kotter & Wicherski 2005),the form and functionality of which can be most advantageous for the purposes of information hiding and distributed information processing.

Through a multi-disciplinary approach to its theoretical design and construction, the resulting tool will be able to create accurate and concise Network Attacker profiles, through the profiling and study of both off-line and real-time network attacks against computer networks, that will satisfy the requirements for data and evidence integrity and evidential reasoning, thus allowing the Network Attacker Profiles to be eligible for use in primarily a UK court of law.

Furthermore, the resulting Distributed Forensic Attack Profiling Network software will, by design, strive to satisfy considerations with regards to system and data/evidence security from compromise through a combination of modular architecture and information hiding principles and methodologies, as well as provide for both computational and evidence integrity through the use of both sound Software Engineering processes and Data Integrity Protection practices.

Finally, the resulting software will strive to comply with the EU Privacy Law with respect to IP address anonymity during data collation and processing without however compromising the integrity or evidential reasoning of the information analysed and thus its value as evidence.

5.1 Design of the proposed software

The proposed approach aims to develop a modular software framework for a distributed forensics attack profiling network that will use a collection of generalised modules to assemble customised sensors that will allow for customised network traffic data collection.

The sensors themselves will allow for preprocessing and feature extraction of the data before transmitting them using technologies such as multi-step ahead neural network predictors (Mccoy, Ward, Mcloone & Delaney 2007) and data stream management (Golab & Özsu 2003) to the controller, thus effectively pre-processing and sanitising (through dimentionality reduction) the data as a preparation for further analysis, while using a combination of passive monitoring technologies and botnet-derived stealth techniques to remain hidden from users and attackers alike. The sensors' modular nature will allow for not just a higher and more complete picture of the state of the network but also, through pre-processing and data sanitisation, lighten the

load of the main analysis engine.
The analysis and correlation of the sanitised network traffic provided by the sensors will be handled centrally and will involve the use of a collection of data mining (Last, Kandel & Bunke 2004, Anderson et al. 2007) and neural network (Bishop 2005, Ripley 2005) approaches to both profile both the attack and the attacker and evaluate the likelihood of the attack being successful, thus also allowing the classification of the attacker's prowess, either through the use of a mathematical/statistical scoring model as those discussed in (Marshall & Moor 2006) or through the use of AI-derived techniques. It will also make use of data anonymisation techniques such as discussed in (Prior & Tompsett 2006) to prevent sensitive information from breaching Data Protection directives.
The software framework will also provide a mechanism by which changes in the sensor configuration will propagate to either all of the sensor network or to a specific number of sensors throught the use of a command issued from the control interface as well as a mechanism by which a sensor can be installed automatically, but with extreme attention to security and access control, to any new computer that is connected to the network. This mechanism is integral to BotNet technologies and is one of the most useful features of such tools, and thus the proposed tool's as well, as it allows for extremely easy and efficient sensor placement, management and update without loss of integrity or performance.

6 Conclusions and Further Work

The application of Profiling theory and methodologies in the Computer Science discipline has been used to a great extent in the fields of Computer Networks and Network Security, and more specifically Intrusion Detection, to ascertain either the overall condition of a computer network at any given point in time and identify issues affecting network stability and throughtput or the presense and activity of malicious or inapropriate users and content, thus indicating an imminent attack on a company or organisation's computing infrastructure. The results such research yielded, both in terms of scientific interest in the field and in terms of useable software tools to aid Computer Network administrators and Network Security engineers, has sparked the interest of the Digital Forensics field, with research starting to take place to determine the viability of applying the resulting methodologies and models to the creation of perpetrator-related profiles in the course of Digital Forensic Investigations.

The first part of this paper identified and briefly discussed in greater detail some of the uses of the Profiling theory and methodologies in the fields of Computer Networks and Network Security, further discussing the problems inhibiting the easy usage of standard Intrusion Detection Systems' profiling capabilities in the field of Digital Forensics as anything other than proof of a perpetrator's presence in a victim company's or organisation's computer system or network. In the second part of this paper a Distributed Forensic Attack Profiling Network software was proposed that will result in the production of forensically sound, and integrity-protected profiles of host and network-based attacks and the individuals or groups of individuals behind them, utilising evidential reasoning so as to satisfy the admissibility requirements for evidence in a UK court of law.

As the proposed solution is, as yet, in its infancy, further work needs to be carried out to first of all ascertain all the parameters required to produce a working profile of both a network attack and of the attacker behind it, as well as create a number of rules taken from the fields of Digital Forensics, Classic Forensics, Law and Software Engineering the combination of which will provide an adequate level of validity to both the software and the results gained from it in terms of legal admissibility in a court of law. Furthermore, both peer-accepted off-line and real-time data needs to be obtained, while developing the proposed software in order for tests of the theoretical and computational models to be performed, the results of which will determine the viability and suitability of the proposed software.

References

Almulhem, A. & Traore, I. (2005), 'Experience with engineering a network forensics system', *Information Networking* pp. 62–71. http://www.springerlink.com/content/1lkpd7771fq87ge0.

Amoroso, E. (1999), *Intrusion Detection: An Introduction to Internet Surveillance, Correlation, Trace Back, Traps, and Response*, Intrusion.Net Books, New Jersey, US.

Anderson, G., Selby, D. & Ramsey, M. (2007), 'Insider attack and real-time data mining of user behavior', *IBM Journal of Research and Development, 2007* **51**(3/4). http://www.research.ibm.com/journal/rd/513/anderson.html, Last seen 20/11/07.

Bace, R. & Mell, P. (2001), 'Nist special publication on intrusion detection systems'. http://csrc.nist.gov/publications/nistpubs/800-31/sp800-31.pdf, Last seen: 20/11/07.

Bacher, P., Holz, T., Kotter, M. & Wicherski, G. (2005), 'Know your enemy:tracking botnets', *The Honeynet Project and Research Alliance* . http://www.honeynet.org/papers/bots/, Last seen: 20/11/07.

Bishop, C. (2005), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford,UK. ISBN: 0-19-853864-2.

Casey, E. (2004), *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet: Second Edition*, second edition edn, Elsevier Academic Press. ISBN: 978-0-12-163104-8.

Casey, E. (2006), 'Investigating sophisticated security breaches', *Commun. ACM* **49**(2), 48–55.

Chiller, C., Binkley, J., Harley, D., Evron, G., Bradley, T., Carsten, C. & Cross, M. (2007), *Botnets: The Killer Web Applications*, Syngress Publishing, Inc. ISBN: 978-1597491358.

Dao, V. N. P., Vemuri, R. & Templeton, S. J. (2000), 'Profiling users in the unix os environment', *International Computer Science Conventions, Wollongong (AU)* . http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=15004770.

Dasgupta, D. & Brian, H. (2001), Mobile security agents for network traffic analysis, *in* 'Proc. DARPA Information Survivability Conference & Exposition II DISCEX '01', Vol. 2, pp. 332–340 vol.2.

Durgin, N. & Pengchu, Z. (2005), Profile-based adaptive anomaly detection for network security, SANDIA REPORT SAND2005-7293, Sandia National Laboratories, United States. http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2005/057293.pdf.

Golab, L. & Özsu, M. T. (2003), 'Issues in data stream management', *SIGMOD Rec.* **32**(2), 5–14.

Hannan, M., Frings, S., Broucek, V. & Turner, P. (2003), 'Forensic computing theory & practice: Towards developing a methodology for a standardised approach to computer misuse', *Paper presented at the 1st Australian Computer, Network & Information Forensics Conference, Perth, WA, Australia* . http://forensics.utas.edu.au/files/ACNIFC2003.pdf.

Ianelli, N. & Hackworth, A. (2005), 'Botnets as vehicle for online crime', *CERT Coordination Centre* . `http://www.cert.org/archive/pdf/Botnets.pdf`, Last seen: 20/11/07.

Kahai, P., Srinivasan, M., Namuduri, K. & Pendse, R. (2006), *Advances in Digital Forensics: Chapter 13. Forensic Profiling System*, Vol. 194/2006 of *IFIP International Federation for Information Processing*, Springer Boston, chapter 13. Forensic Profiling System, pp. 153–164.

Karagiannis, T., Papagiannaki, K. & Faloutsos, M. (2005), 'Blinc: multilevel traffic classification in the dark', *SIGCOMM Comput. Commun. Rev.* **35**(4), 229–240.

Keys, K., Moore, D. & Estan, C. (2005), A robust system for accurate real-time summaries of internet traffic, *in* 'SIGMETRICS '05: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems', ACM, New York, NY, USA, pp. 85–96.

Last, M., Kandel, A. & Bunke, H. (2004), *Data Mining In Time Series Databases*, World Scientific Publishing Co Pte Ltd. ISBN: 9812382909.

Levitt, T. & Laskey, K. (2000), 'Computational inference for evidential reasoning in support of judicial proof', *Cardozo Law Review, 2000* . `http://www.ite.gmu.edu/~klaskey/papers/Levitt_Laskey_Inf_JP.pdf`, Last seen: 20/11/07.

Marshall, A. & Moor, G. (2006), 'Criminalisation of the internet: an examination of illegal activity online.', *The 4th European Academy of Forensic Science Conference - EAFS2006* .

Marshall, A. & Tompsett, B. (2002), 'Spam 'n' chips: A discussion of internet crime', *Science & Justice* **42**(2), 117–122.

Mccoy, A., Ward, T., Mcloone, S. & Delaney, D. (2007), 'Multistep-ahead neural-network predictors for network traffic reduction in distributed interactive applications', *ACM Trans. Model. Comput. Simul.* **17**(4), 16.

Mohay, G. (2005), Technical challenges and directions for digital forensics, *in* 'Proc. First International Workshop on Systematic Approaches to Digital Forensic Engineering', pp. 155–161.

Payer, U. (2004), 'Realtime intrusion-forensics: A first prototype implementation', *TERENA Networking Conrerence, 2004* . `http://www.`

terena.nl/library/tnc2004-proceedings/papers/payer.pdf, Last seen: 20/11/07.

Prior, S. & Tompsett, B. (2006), 'Problems of privacy, security, identity, integrity, legality and confidentiality in internet crime investigation and evidence collection', *ECCE2006 - e-crime and computer evidence 2006* .

Puttini, R., Marrakchi, Z. & M, L. (2003), A bayesian classification model for real-time intrusion detection, *in* 'AIP Conference Proceedings – March 31, 2003 – Volume 659', BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 22nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, pp. pp. 150–162.

Ripley, B. (2005), *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, UK. ISBN: 0-521-46086-7.

Sebyala, A., Olukemi, T. & Sacks, L. (2002), Active platform security through intrusion detection using naive bayesian network for anomaly detection, *in* 'Proceedings of the London Communications Symposium 2002'. http://citeseer.ist.psu.edu/542374.html.

Shanmugasundaram, K., Memon, N., Savant, A. & Brnnimann, H. (2003), 'Fornet: A distributed forensics network', *MMM-ACNS 2003* pp. 1–16. http://citeseer.ist.psu.edu/shanmugasundaram03fornet.html, Last seen: 20/11/07.

Sommer, P. (1999), 'Intrusion detection systems as evidence', *Comput. Networks* **31**(23-24), 2477–2487.

Stephenson, P. (2000), 'The application of intrusion detection systems in a forensic environment', *RAID Symposium 2000* . http://www.raid-symposium.org/raid2000/Materials/Abstracts/47/47.pdf.

Xu, K., F. Wang, S. B. & Zhang, Z. (2007), 'A real-time network traffic profiling system', *37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07)* **0**, 595–605.

Xu, K., Zhang, Z. & Bhattacharyya, S. (2005), Profiling internet backbone traffic: behavior models and applications, *in* 'SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications', ACM, New York, NY, USA, pp. 169–180.

Yasinsac, A. & Manzano, Y. (2001), 'Policies to enhance computer and network forensics', *Workshop on Information Assurance and Security, United States Military Academy, West Point, NY* . http://citeseer.ist.psu.edu/yasinsac01policies.html.

Yuill, J., Wu, S. F., Gong, F. & Huang, M.-Y. (1999), Intrusion detection for an on-going attack, *in* 'Recent Advances in Intrusion Detection'.