DATASET GENERATION AND NETWORK INTRUSION DETECTION BASED ON FLOW-LEVEL INFORMATION

AHMED ABDALLA MOHAMEDALI ABDALLA

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Electrical Engineering)

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

17th SEPTEMBER 2015

To My Mother for all what she gave to me and for what she is still giving.

I dedicated this humble work to you.

ACKNOWLEDGEMENT

First of all, I would like to express my deepest thanks to ALLAH (SWT), who guided, helped, and supported me.

I wish to send my serious gratefulness and thankfulness to my research supervisors, Associate Professor Dr. Sulaiman Mohd Nor and Dr. Shaikh Nasir bin Shaikh Husin for encouragement, motivation, full support in academic, social, and technical issues. Truly, I have been overwhelmed by their patience, personal kindness, valuable comments, and advices during my study. Without their continuous guidance and help, this thesis would not have been completed.

I have greatly appreciated the opportunity from UTM to purse my PhD, I thank very much all UTM community including each people at (faculties, library, staff, employees, students and labourers). Special thankfulness to every member of people in FKE, CICT, SPS.

Finally, I would take pleasure in gratitude my mother, father and all the members of my family, friends and relatives for their constant dear one, help and support which motivate me to face the life difficulties.

Ahmed Abdalla

ABSTRACT

The growth of the Internet and networking has made securing networks against attacks a very challenging task. For high-speed networks, flow meta-data inspection can replace conventional Deep Packet Inspection but with the cost of low precision of identifying attacks since the former deals with an aggregated version of the traffic. The first part of this research addresses the problem of the lack in benchmarking datasets for developing new Network Intrusion Detection Systems (NIDSs) or comparing existing NIDSs. The aim in the second part is to design a near real-time NIDS without degrading the detection accuracy when compared to conventional misuse packet-based approaches. To achieve the first objective, a NIDS dataset creation framework had been developed. Based on that framework, a flow-level NIDS dataset had been created. The traces were collected from campus main routers in NetFlow format while malicious traffic of different attack scenarios was generated by Nmap and BoNesi tools. In the second part a flow-based software-based system were developed to detect and identify network volume-level attacks in near real-time. Attack detection is based on statistical time-aggregated features of the exported NetFlow version of the traffic to detect several scan and Denial-of-Service (DoS) attacks. A validation for the designed system is done using Defense Advanced Research Projects Agency (DARPA) datasets. The timeline performance outperformed all relevant software-based systems and suited for up to one gigabit per second links with an average detection delay of less than one minute. The proposed method achieved 95% True Positive Rate (TPR) and almost zero False Positive Rate (FPR). Compared to relevant methods when operated in the same conditions, the proposed method improved the TPR by 4% and improved FPR by 1%. In addition, the capability of flow-based approach in detecting packet-level attacks was experimentally demonstrated. The results against Snort were benchmarked and 75% TPR and almost zero FPR were achieved.

ABSTRAK

Pertumbuhan Internet dan rangkaian telah menjadikan keselamatan rangkaian terhadap serangan satu tugas yang sangat mencabar. Untuk rangkaian kelajuan tinggi, penelitian meta-data boleh mengganti penelitian tahap paket konvensional tetapi dengan kos ketepatan rendah semasa mengenalpasti serangan memandangkan tahap-aliran berkait dengan agregasi trafik. Bahagian pertama kajian ini bertujuan menyumbang dalam menyelesaikan masalah keperluan set data penandaan aras yang piawai untuk membangun Sistem Pengesan Penceroboh Rangkaian (NIDS) baru atau membanding dengan yang sedia ada. Tujuan di bahagian kedua adalah untuk merekabentuk NIDS masa nyata berhampiran tanpa merendahkan ketepatan pengesanan apabila dibandingkan dengan pendekatan konvensional penyalahgunaan berasaskan paket. Bagi mencapai objektif pertama satu rangka kerja bagi mewujudkan dataset NIDS telah dibangunkan. Berdasarkan rangka kerja tersebut dataset tahap-aliran NIDS telah dibangunkan. Surih dikutip dari penghala utama kampus dalam format *NetFlow* manakala trafik hasad dari senario serangan berbeza telah dihasilkan oleh peralatan *Nmap* dan *BoNesi*. Dalam bahagian kedua, satu sistem berasaskan perisian tahap-aliran telah dibangunkan bagi mengesan dan mengenalpasti serangan rangkaian tahap-jumlah dalam masa nyata berhampiran. Pengesanan serangan adalah berdasarkan kepada ciri-ciri statistik masa terkumpul trafik *NetFlow* versi yang dieksport bagi mengesan beberapa imbasan dan serangan Penafian Perkhidmatan (DoS). Pengesahan sistem yang direkabentuk dibuat menggunakan set data Agensi Projek Penyelidikan Termaju Pertahanan (DARPA). Prestasi had masa sistem telah mengatasi sistem lain berasaskan perisian dan hanya sesuai sehingga kelajuan satu gigabit sesaat dengan purata lengah kurang dari satu minit. Kaedah ini berjaya mencapai 95% Kadar Positif Benar (TPR) dan hampir sifar Kadar Positif Palsu (FPR). Berbanding dengan kaedah berkaitan apabila beroperasi dalam keadaan yang sama, kaedah ini telah memperbaiki TPR sebanyak 4% dan memperbaiki FPR sebanyak 1%. Sebagai tambahan, kemampuan kebolehan pendekatan berasaskan aliran di dalam mengesan serangan tahap paket telah dapat ditunjukkan secara eksperimen. Keputusan berbanding dengan *Snort* telah ditanda aras dan telah mendapat 75% TPR dan hampir sifar FPR.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview and Motivation

The global growth of the Internet and networking has made securing networks and information one of the most challenging tasks in the field of network communication. Today, intrusion attacks are generating significant worldwide epidemic to network security environment and bad impact involving financial loss. Intruders have the capability to infect thousands of hosts and networks within few minutes before human action takes place.

Studies on the field by Dell™ SonicWALL™ Threat Research Team demonstrated that network attacks are increasing exponentially every year. For example, over 20.1 million unique malware samples were collected in 2013, compared to 16 million in 2012 (DellSonicWALL, 2013). In the year 2013 more than 1.06 trillion intrusion related incidents had been detected and prevented and more than 1.78 billion malwares downloads had been blocked. According to the same report, there were approximately 4,429 new vulnerabilities reported from Common Vulnerabilities and Exposures (CVE) and 3,644 were related with network attacks in 2013 (DellSonicWALL, 2013). Even a small security breach on commercial dealings such as online e-payment and money transactions may cause huge unrecoverable losses to companies and individuals. Successful attacks to hijack network-based sensitive services in hospitals or airports may lead to disastrous results.

In response to those increasing cyber-attacks, the cost of securing networks information is also increasing. According to the industry survey on cyber security conducted by New York State Department of Financial Services (Andrew M. Cuomo, 2014), 77% of institutions involved in the survey experienced an increase in their total information security budget in the past three years. Almost no institutions reported a reduction in spending in the past three years.

In order to cope with the enormous increasing security threads that face network communication, various third party techniques or middle-boxes techniques have evolved such as firewalls, Network Intrusion Detection Systems (NIDSs). Network Intrusion Detection Systems (NIDS) proved to be an efficient technique that can process large volume of networks traffic and detect intrusions in their early stages in order to limit their catastrophic damages.

Intrusion can be defined as any activity that violates confidentiality, authority, integrity or availability of information system (Lazarevic et al., 2005). Although in the field of information security, intrusion and attack are used interchangeably, there is a little difference between the two terms. An attack is an intrusion attempt, and an intrusion results from an attack that has been (at least partially) successful (Barrus and Rowe, 1998). Intrusion Detection is defined by The National Institute of Standards and Technology (Bace and Mell, 2001) as “the process of monitoring the events occurring in a computer or network and analyzing them for signs of intrusions”. An Intrusion Detection System (IDS) can be defined as a combination of software or/and hardware components that monitors computer and network systems and raises an alarm when intrusion happens (Lazarevic et al., 2005).

There are two main categories of IDSs; host-based IDSs (HIDS) which are designed for single host, and network-based IDSs (NIDS) which monitor the whole traffic of a network. Generally, the concepts of IDS is more related to NIDS than HIDS since the latter can be considered as an ordinary antivirus system limited to one host whilst the prior is an entity that can monitor a whole enterprise network or even an ISP network.

1.2 Background

1.2.1. A Simple NIDS Architecture

In general, the components which make up any NIDS can be viewed as shown in Figure 1.1 (Lazarevic et al., 2005).

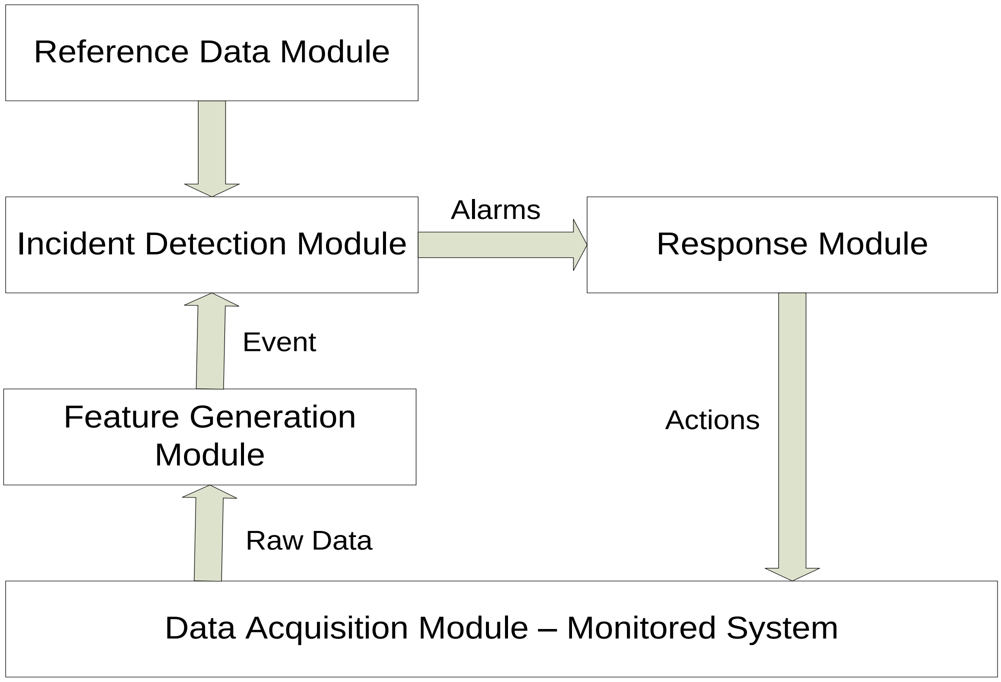
Reference Data Module
Incident Detection Module
Alarms
Response Module
Event
Feature Generation Module
Actions
Raw Data
Data Acquisition Module – Monitored System

Figure 1.1 Simple architecture of NIDS (Lazarevic et al., 2005)

The purpose of each module is as follows:

i. Data Acquisition Module: Responsible for collecting network traffic data.

ii. Feature Generation Module: Responsible for extracting a set of selected traffic features (packet header features, payload features, flow features)

iii. Incident Detection Module: Responsible for identifying intrusion and generating alarms by comparing the data generated from the feature generation module with that of the reference module. Generally there are two methods of this detection; misuse-based detection and anomaly-based detection.

iv. Response Module: Once an alert is received, this module is responsible for initiating actions in response to a possible intrusion.

v. Reference Data Module: This module contains the reference data that is to be used by the incident detector to compare with. If the detection method is misuse-based then this data would be based on the description of all known intrusion (intrusion database), and if the detection method is anomaly-based then it would be based on description of normal attack free network operation (network profile).

1.2.2. IDS Taxonomy

Several taxonomies had been proposed for categorizing NIDS, but the most common and acceptable one is that one used by Axelsson (2000) and Debar et al. (1999). This classification categorizes IDSs according to the following aspects:

i. Information source: NIDS can be packet-level based or flow-level. Packet-level NIDS can be further divided to packet header-based or payload-based.

ii. Analysis Strategy (Detection Method): This categorization distinguishes NIDS according to the nature of reference data used for identifying intrusions which is either misuse-based or anomaly-based.

iii. Architecture (Locus of Detection): NIDS can be centralized when it is placed in single position or distributed when there are several points for monitoring.

iv. Response to intrusion: NIDS can be either passive when it is just to raise an alarm on detecting intrusion or active when a further action is to be done in response to that intrusion.

v. Time Aspects: NIDS can work in real-time and detect intrusions while they are taking place or by batch (non real-time).

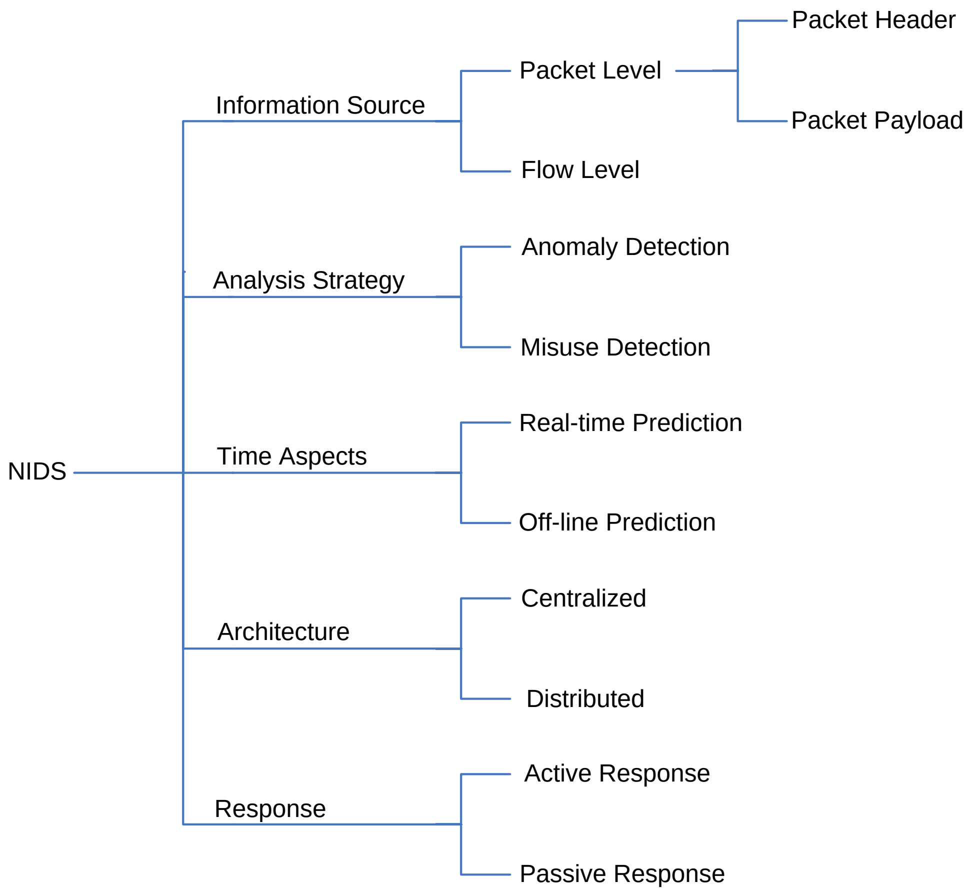

Figure 1.2 NIDS classification aspects

1.2.3. NIDS Required Characteristics

The required NIDS characteristics had been specified by many researchers (Lazarevic et al., 2005, Catania and Garino, 2012). These characteristics are required from NIDS to achieve security goals. We can summarize these characteristics as follows:

i. High Detection Performance: Detection of all attacks without false alarms

ii. Low Processing time: Attack is to be detected as soon as possible

iii. Adaptability: NIDS should be able to readapt itself to deal with novel attacks and changing environments

iv. Fault tolerance: robustness, resistance to attacks, quick recovery from successful attacks

v. Minimum Resource consumption: storage resources and processing capabilities needed.

1.3 Problem Statement

As stated in the previous section, NIDS have a wide range of approaches to be categorized accordingly (Figure 1.2). However, there are two main different approaches now for NIDS. The conventional and most common implemented approach nowadays is the misuse-based NIDS with deep packet inspection (DPI), which tries to detect attacks based on previously prepared database of all known intrusion signatures. This approach proved to be efficient and accurate for detecting known intrusions but cannot detect novel and new attacks or even new variants of known attacks. Another drawback of misuse-based NIDS is that it uses DPI. This approach of auditing data by inspection inspects every incoming packet in the traffic. Hence, it cannot cope with high speed networks. Besides, it seems that researches on misuse-based NIDS have reached a saturation point making it harder to make further enhancements.

The other approach is the anomaly-based NIDS, which is based on detection of deviations of a normal attack-free model of the protected network and flagging these deviations as intrusions. This approach can detect known and novel attacks as well. Although this latter approach seems promising and in spite of lot of researches on it, it is still immature and suffers from serious problems that makes it impractical for real life situations. The major problem of anomaly-based NIDS is that it suffers from high false alarm rates. This problem arises from the difficulty of specifying normal and abnormal thresholds for any network and which itself can be considered as a second problem. A third problem of that approach is the limited capacity and precision of identifying an attack when an alarm is raised. This problem is a consequence of anomaly-based detection nature which reacts with any network event as either normal or abnormal (anomaly) and nothing else. If we consider the fact that there are lots of network anomalies that are neither attacks nor intrusions, then definitely there would be a high rate of false alarms.

With respect to the information inspected (audit data), some NIDS inspect the complete traffic (Deep Packet Inspection) for known attack signatures (byte patterns) whilst others consider data from packets headers for attacks that violates network protocols rules. Meanwhile, other NIDSs depend on extensive summaries and aggregations of traffic flow (flow-level) for intrusion that attack network resources. Each of these levels of audit data inspection concentrates on particular category of intrusive attacks and, therefore, is strong on some parts but may be poor on other parts. Since flow-level NIDS deals with aggregations and summarization of traffic, it can inspect high speed traffic in a near real-time mode. However, it suffers from low precision of specifying attacks in the traffic. This problem is a result of the absence of the actual data in the detection and the use of metadata instead. The whole flow is to be flagged when an attack is detected. The exact malicious packet(s) cannot be specified nor the exact attack time nor any other further information that can help to describe the attack.

Any NIDS needs to go through a knowledge acquisition process to build a reference model in the system before the actual detection process can take place. To build this reference model, the system must be supplied with previously known categorized dataset so that the system would acquire the necessary information to detect attacks and differentiate between malicious and non-malicious incidents in network traffic. Unfortunately, creation of optimum NIDS dataset that would acquire all required characteristics is very hard and costly. In addition, some of these required characteristics seem to be contradicting with each other. Hence, combining these characteristics in a dataset is almost impossible. There are a number of challenges and difficulties in creating ideal datasets. For all that, researchers on NIDS generally suffer from a lack of standard datasets for developing and training new systems. The problem is even severe when looking for benchmarking common public datasets for validating, evaluating and comparing between existing systems and methods.

Most NIDS in the research community today have achieved a very high degree of accuracy in detecting intrusions. However, the majority of these systems face difficulties when they are to be deployed and implemented in real world. With the growing connection speed by the establishment of end users broadband Internet, the data volumes in the backbone networks has increased steadily. Monitoring of links that may reach few gigabits per second capacities by NIDS may require more computational resources than that available on commodity computer hardware. This problem of resource demand is commonly solved by engineering dedicated hardware platforms such as on Field-Programmable Gate Array (FPGA) that can process detection tasks more efficiently. However, specially crafted hardware comes at a significant higher cost and may not be worthwhile in every situation.

Our research problem is divided into two main parts. The first part aims to deal with the problem of the lack in standard benchmarking datasets for developing and training new systems or comparing existing ones. The second part tries to find software-based solutions for flow-level NIDS to achieve accurate near real-time detection. As a part from the second problem, the research tries to answer some questions concerning the capability and efficiency of flow-level attributes to detect network traffic attacks in spite of the absence of actual monitored data, and to find its efficiency in terms of accuracy and timeline performance.

1.4 Research Objectives

The objectives of this research can be summarized as follows:

1. **To create a labeled flow-level dataset for anomaly-based NIDS using a general framework that would be available publicly.**

   A labeled flow-level NIDS dataset is a set of flow records with each flow record labeled as either benign or malicious. A flow record is the set of attributes that abstracts and summarizes all the information which concerns a single session between two communicating entities in a network. The dataset is to satisfy a set of conditions and guidelines in order to acquire validity for anomaly detection.

2. **To design an effective near real-time anomaly-based NIDS to detect brute-force attacks based only on software solutions.**

Effective here means the proposed solution will not degrade the detection accuracy and will not increase the false alarm rate compared to existing conventional DPI or relevant flow-based methods. The system is supposed to detect network intrusion attacks when they are taking place not more than few minutes after the attack has commence. Detection scope of the system consists of all attacks that consume significant network bandwidth in any phase of the attack. The detection system is a set of developed software that does not need special hardware platform to run on and meant to be placed in the ingress/egress point of Universiti Teknologi Malaysia (UTM) campus network. Hence, it has to cope with high speed traffic of possibility to few hundreds megabit per second.

3. **To investigate the efficiency of flow-level traffic inspection to detect packet content-based attacks.**

The capability, accuracy and timeline performance of flow-level features to detect attack contained in payload packets will be investigated and evaluated.

1.5 Research Scope

This research focuses on developing a system that can monitor the traffic of server farm in UTM campus network in real-time and setting alarm when observing intrusive activities. Hence, the research deals with traffic bandwidth not exceeding one gigabit per second, as this is a little higher than the core bandwidth of UTM campus backbone.

All proposed solutions are software-based and are to be run within general purpose computer machines that are commercially available in the market. No special hardware platform such as FPGA would be considered.

Although the study includes misuse detection approach, it will limit inspection level to flow features and will not consider any of payload inspection or pattern signature as information source for intrusion detection. Inspection of flow-level data is limited to network and transport layers. Hence, attacks and techniques in lower layers (physical layers) and upper layers (application layers) will not be considered.

All proposed solutions and systems are passive with respect to the detected intrusions. Their role is solely to flag an alarm when any intrusive activity is detected with no further action to be performed.

The study is about Network IDS. Hence, all techniques reviewed, studied or analyzed will depend only on network traffic data (packets and flow data). Host IDS is out of scope of this study. Intrusions would be considered an attack when it uses network resources to scan, propagate, or penetrate network information resources. Intrusions affect information sources through other means (e.g. physical access) will not be considered as network attack.

1.6 Significance of the Research

The most significant contribution of this study is the development of a general framework to create NIDS datasets. That framework would be available publicly for research community. This would help other researchers to develop their own systems and methods and ease the comparison and validations for these new systems and methods.

Another outcome of the study is a cost effective solution to the problem of real-time traffic monitoring in any enterprise networks with up to one gigabit per second bandwidth. It provides a user-friendly NIDS that make the security task easier and provide an early alarm security system for the monitored network through developing an efficient NIDS as a step to replace other costly commercial software products that are currently being used.

1.7 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2, an overview on the evolution of NIDS and its state-of-the-art that highlights gaps and defines future trends is given. This is followed by an overview on network traffic generating, capturing and processing tools and the significant theories and methods used in NIDS in general. A deep flow-level NIDS literature review is done followed by another review on NIDS datasets.

Chapter 3 shows the overall research methodology, explains and discusses general dataset creation framework, the framework of flow-level detection of brute-force attacks, explains the framework of flow-level detection of packet-level attacks. It also discusses with details all experimental setups, dataset creation steps, validation methods and rational constructions behind the selected methodology.

Chapter 4 presents the implementation details of creating brute-force attacked dataset from captured traffic traces, statistics and various distributions of the created dataset and discusses the corresponding results.

Chapter 5 presents the implementation details of designing flow-level NIDS to detect brute-force attacks and all partial and final results concerning evaluation of accuracy and performance and validation of the detection method and discusses results.

Chapter 6 presents the implementation details of investigating the efficiency of flow-level data to detect packet-level attacks and a discussion to justify the results.

Chapter 7 concludes all the study and states the degree of objective fulfillment, highlights research contributions and makes suggestion for future work.

REFERENCES

Aldous, D. (1991). The Continuum Random Tree I. *The Annals of Probability*, 19(1), 1-28.

Allen, J., Christie, A., Fithen, W., McHugh, J., Pickel, J. and Stoner, E. (2000). *State of the Practice of Intrusion Detection Technologies*. Technical Report CMU/SEI-99-TR-028 ESC-99-028. Software Engineering Institute, CarnegieMellon University, Pittsburgh, Pennsylvania.

Aslam, T. (1995). *A Taxonomy of Security Faults in the Unix Operating System.*Doctoral Dissertation.Department of Computer Sciences,Purdue University, Lafayette, Indiana.

Attanasio, C. R., Markstein, P. W. and Phillips, R. J. (1976). Penetrating an Operating System: AStudy of VM/370 Integrity. *IBM Systems Journal,* 15, 102-116.

Axelsson, S. (2000a). The Base-rate Fallacy and the Difficulty of Intrusion Detection. *ACM Transactions on Information and System Security,* 3, 186-205.

Axelsson, S. (2000b). *Intrusion Detection Systems: A Survey and Taxonomy.*Technical Report 99-15, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden.

Bace, R. and Mell, P. (2001). *NIST Special Publication on Intrusion Detection Systems.*Technical ReportSP 800-31, National Institute of Standards and Technology, Gaithersburg, Maryland.

Barrus, J. and Rowe, N. C. (1998) A Distributed Autonomous-agent Network Intrusion Detection and Response System. *Proceedings of the Command and Control Research and Technology Symposium,*Naval Postgraduate School Monterey,USA.

Bhuyan, M. H., Bhattacharyya, D. and Kalita, J. K. (2014). Network Anomaly Detection: Methods, Systems and Tools. *IEEE Communications Surveys and Tutorials,* 16(1), 303-336.

Bodt, B. A. (2002). Computer Intrusion Detection and Network Monitoring. *Technometrics,* 44(3), 294-295.

Breiman, L. (2001). Random Forests. *Machine Learning,* 45(1), 5-32.

CAIDA. (2015). *Center for Applied Internet Data Analysis,* [Online]. Available: http://www.caida.org/home/.

Catania, C. A. and Garino, C. G. (2012). Automatic Network Intrusion Detection: Current Techniques and Open Issues. *Computers and Electrical Engineering*, 38(5), 1062-1072.

Cheng, C.M., Kung, H. R. and Tan, K.S. (2002). Use of Spectral Analysis in Defense Against DoS Attacks. *Proceedings of IEEE Global Telecommunications Conference*, 3,2143-2148.

Cisco. (2013). *Cisco IOS NetFlow* [Online]. Available: http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html.

Cisco. (2015). *SNORT* [Online]. Available: https://www.snort.org/.

Cisco. (2015). *NetFlow Export Datagram Format* [Online]. Available: http://www.cisco.com/c/en/us/td/docs/net_mgmt/netflow_collection_engine/3-6/user/guide/format.html.

Claise, B. (2008). *Specification of the IP flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information*. RFC 7011, IETF.Available: https://tools.ietf.org/html/rfc7011.

Computer Knowledge. (2015). *Computer Knowledge- Virus Tutorial* [Online]. Available: http://www.cknow.com/cms/vtutor/cknow-virus-tutorial.html.

Copeland III, J. A. (2007). Flow-based Detection of Network Intrusions. U.S. Patent No. 7,185,368. Washington, DC: U.S. Patent and Trademark Office.

Dagon, D., Gu, G., Lee, C. P. and Lee, W. (2007). A Taxonomy of Botnet Structures. *Proceedings of the Annual Computer Security Applications Conference*, *23,* 325-339.

Debar, H., Dacier, M. and Wespi, A. (1999). Towards a Taxonomy of Intrusion-Detection Systems. *Computer Networks,* 31, 805-822.

DellSonicWALL (2013). Dell Network Security Threat Report 2013. US-TD593-20140123. U.S.A: Dell.Available:http://marketing.sonicwall.com/documents/dell-network-security-threat-report-2013-whitepaper-30197.pdf

Denning, D. E. (1987). An IntrusionDetection Model. *IEEE Transactions on Software Engineering*, 13(2), 222-232.

Diadem Firewall European Project. (2015). *Diadem Firewall*[Online]. Available: http://www.diadem-firewall.org/index.php.

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J. and Tan, P.N. (2002). Data Mining for Network Intrusion Detection. *Proceedings of NSF Workshop on Next Generation Data Mining*, 21-30.

Dreger, H., Feldmann, A., Paxson, V. and Sommer, R. (2004). Operational Experiences with High Volume Network Intrusion Detection. *Proceedings of the Conference on Computer and Communications Security*. ACM, 11, 2-11.

Erbacher, R. F. (2001). Visual Behavior Characterization for Intrusion Detection in Large Scale Systems. *Proceedings of the International Conference on Visualization, Imaging, and Image Processing*, 54-59.

Fide, S. and Jenks, S. (2006). *A Survey of String Matching Approaches in Hardware*. Technical ReportTR SPDS 06-01, Department of Electrical Engineering and Computer Science, University of California, Irvine, USA.

Fioreze, T., Oude Wolbers, M., van de Meent, R. and Pras, A. (2007). Finding Elephant Flows for Optical Networks. *Proceedings of the 10th IFIP/IEEE International Symposium on Integrated Network Management*, 627-640.

Fisher, D. H., Pazzani, M. J. and Langley, P. (Eds.).(2014). *Concept Formation: Knowledge and Experience in Unsupervised Learning*.San Mateo: Morgan Kaufmann.

Frank, J. (1994). Artificial Intelligence and Intrusion Detection: Current and Future Directions. *Proceedings of the 17th National Computer Security Conference*, 10, 1-12.

Galtsev, A. A. and Sukhov, A. M. (2011). Network Attack Detection at Flow Level. *Proceedings of the 11th International Conference and 4th International Conference on Smart Spaces and Next Generation Wired/Wireless Networking*, 326-334.

Gao, M., Zhang, K. and Lu, J. (2006). Efficient Packet Matching for Gigabit Network Intrusion Detection Using TCAMs. *Proceedings of the 20th International Conference of Advanced Information Networking and Applications*, 249-254.

Gao, Y., Li, Z. and Chen, Y. (2006). A DoSResilient Flow-level Intrusion Detection Approach for High-speed Networks. *Proceedings of26thIEEE International Conferenceon Distributed Computing Systems*, 39-39.

Garner, S. R. (1995). Weka: The Waikato Environment for Knowledge Analysis. *Proceedings of the New Zealand Computer Science Research Students Conference*, 57-64.

Gates, C., McNutt, J. J., Kadane, J. B. and Kellner, M. I. (2006). Scan Detection on very Large Networks Using Logistic Regression Modeling. *Proceedings of the11th IEEE Symposium on Computers and Communications*, 402-408.

Ghorbani, A. A., Lu, W. and Tavallaee, M. (2009). *Network Intrusion Detection And Prevention: Concepts and Techniques*, New York:Springer.

Gil, T. M. and Poletto, M. (2001). MULTOPS: A DataStructure for Bandwidth Attack Detection. *Proceedings of 10th USENIX Security Symposium*,23-38

GitHub. (2015). *BoNeSi - the DDoS Botnet Simulator* [Online]. Available: https://github.com/markus-go/bonesi.

Gogoi, P., Bhuyan, M. H., Bhattacharyya, D. and Kalita, J. K. (2012). Packet and Flow Based Network Intrusion Dataset. *Proceeding of the 5th International Conference on Contemporary Computing*, Springer, 322-334.

Andrew M. Cuomo, Lawsky,B. M. (2014). *Report on Cyber Security in the Banking Sector*, Department of Financial Services,New York State,USA.Available:http://www.dfs.ny.gov/about/press2014/pr140505_cyber_security.pdf

Hansman, S. and Hunt, R. (2005). A Taxonomy of Network and Computer Attacks. *Computers and Security,* 24(1), 31-43.

Hofmeyr, S. A. and Forrest, S. (1999). *An Immunological Model of Distributed Detection and its Application to Computer Security.* DoctoralDissertation,Department of Computer Science,University of New Mexico, New Mexico.

Hoque, N., Bhuyan, M. H., Baishya, R. C., Bhattacharyya, D. and Kalita, J. K. (2014). Network Attacks: Taxonomy, Tools and Systems. *Journal of Network and Computer Applications,* 40, 307-324.

Hu, W., Liao, Y. and Vemuri, V. R. (2003). Robust Anomaly Detection Using Support Vector Machines. *Proceedings of the International Conference on Machine Learning*, 282-289.

Igure, V. and Williams, R. (2008). Taxonomies of Attacks and Vulnerabilities in Computer Systems. *IEEECommunications Surveys and Tutorials*, 10(1), 6-19.

Information Security Center of eXcellence. (2015). *ISCX Dataset* [Online]. Available: http://www.iscx.ca/datasets.

irchelp.org. (2015). *Trojan Horse Attacks* [Online]. Available: http://www.irchelp.org/irchelp/security/trojan.html.

Jones, A. K. and Sielken, R. S. (2000). *Computer System Intrusion Detection: A Survey*. Technical Report, Computer Science Department, University of Virginia, Charlottesville.

Joshi, M. V., Agarwal, R. C. and Kumar, V. (2002). Predicting Rare Classes: Can Boosting Make any Weak Learner Strong? *Proceedings of the eighth ACM International Conference on Knowledge Discovery and Data Mining*, 297-306.

Jung, J., Paxson, V., Berger, A. W. and Balakrishnan, H. (2004). Fast Portscan Detection Using Sequential Hypothesis Testing. *IEEE Symposium on Security and Privacy*, 211-225.

Kashyap, H. J. and Bhattacharyya, D. (2012). A DDoS Attack Detection Mechanism Based on Protocol Specific Traffic Features. *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology*. ACM, 194-200.

Ke-xin, Y. and Jian-qi, Z. (2011). A Novel DoS Detection Mechanism. *Proceedings of the International Conference on Mechatronic Science, Electric Engineering and Computer*, 296-298.

Kendall, K. (1999). *A Database Of Computer Attacks For The Evaluation Of Intrusion Detection Systems.*Master's Thesis, Department of Electrical Engineering and Computer Science,Massachusetts Institute of Technology, Cambridge.

Kim, M.-S., Kong, H.-J., Hong, S.-C., Chung, S.-H. and Hong, J. W. (2004). A Flow-based Method for Abnormal Network Traffic Detection. *Proceedings of IEEE/IFIPNetwork Operations and Management Symposium*, 599-612.

Kotsiantis, S., Zaharakis, I. and Pintelas, P. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Frontiers in Artificial Intelligence and Applications,* 160, 3-24.

Kotsokalis, C., Kalogeras, D. and Maglaris, B. (2001). Router-based Detection of DoS and DDoS Attacks. *Eighth Workshop of the HP OpenView University Association, Berlin, Germany.*

Kruegel, C. and Toth, T. (2000). *A Survey On Intrusion Detection Systems.* Technical Report TUV-1841-00-11, University of Technology, Vienna,Austria.

Kumar, S. (1995). *Classification and Detection of Computer Intrusions.*Doctoral Dissertation, Department of Computer Sciences,Purdue University, Lafayette, Indiana.

Lai, H., Cai, S., Huang, H., Xie, J. and Li, H. (2004). A Parallel Intrusion Detection System For High-speed Networks. *Proceedings of the Second International Conference on Applied Cryptography and Network Security*, 439-451.

Lakhina, A., Crovella, M. and Diot, C. (2004a). Characterization of Network-wide Anomalies in Traffic Flows. *Proceedings of the 4th ACM SIGCOMM conference on Internet Measurement*, 201-206.

Lakhina, A., Crovella, M. and Diot, C. (2004b). Diagnosing Network-wide Traffic Anomalies. *Proceedings of the ACM SIGCOMM conference on Computer Communication Review*, 219-230.

Lakhina, A., Crovella, M. and Diot, C. (2005). Mining anomalies Using Traffic Feature Distributions. *Proceedings of the ACM SIGCOMM conference on Computer Communication Review*, 217-228.

Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E. D. and Taft, N. (2004c). Structural Analysis of Network Traffic Flows.*ACM Sigmetrics,*32(1), 61-72.

Larochelle, D. and Evans, D. (2001). Statically Detecting Likely Buffer Overflow Vulnerabilities. *Proceedings of the 10thUSENIX Security Symposium*, 32, 177-190.

Lau, F., Rubin, S. H., Smith, M. H. and Trajkovic, L. (2000). Distributed Denial of Service Attacks. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics,*3, 2275–2280.

Lawrence Berkeley National Laboratory and ICSI. (2005). *LBNL/ICSI Enterprise Tracing Project* [Online]. Available: http://www.icir.org/enterprise-tracing/.

Lazarevic, A., Kumar, V. and Srivastava, J. (2005). Intrusion Detection: A Survey. *Managing Cyber Threats*, 19-78.

Lee, W., Wang, C. and Dagon, D. (2007). *Botnet Detection: Countering the Largest Security Threat.*New York: Springer.

Lee, W. and Xiang, D. (2001). Information-theoretic Measures for Anomaly Detection. *Proceedings of the IEEE Symposium on Security and Privacy*, 130-143.

Li, B., Springer, J., Bebis, G. and Gunes, M. H. (2013). A Survey of Network Flow Applications. *Journal of Network and Computer Applications,* 36, 567-581.

Li, Z., Gao, Y. and Chen, Y. (2005). Towards a High-speed Router-based Anomaly/Intrusion Detection System.*Proceedings of the Special Interest Group on Data Communication ,* 15-16.

Lincoln Laboratory, M. I. o. T. (2000). *DARPA Intrusion Detection Evaluation*[Online]. Available: http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/2000data.html.

Lindqvist, U. and Jonsson, E. (1997). How to Systematically Classify Computer Security Intrusions. *Proceedings of the IEEE Symposium on Security and Privacy*, 154-163.

Lippmann, R. P., Fried, D. J., Graf, I., Haines, J. W., Kendall, K. R., McClung, D., Weber, D., Webster, S. E., Wyschogrod, D. and Cunningham, R. K. (1998). Evaluating Intrusion Detection Systems: The 1998 DARPA Offline Intrusion Detection Evaluation. *The International Journal of Computer and Telecommunications Networking,*34, 579-595.

Long, N. and Thomas, R. (2001). Trends In Denial Of Service Attack Technology. *CERT Coordination Center.* [Online]. Available: http://resources.sei.cmu.edu/asset_files/WhitePaper/2001_019_001_52491.pdf

Lough, D. L. (2001). *A Taxonomy of Computer Attacks with Applications to Wireless Networks*.Doctoral Dissertation, Faculty of the Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Makhtar, M., Neagu, D. C. and Ridley, M. J. (2011). Comparing Multi-class Classifiers: on the Similarity of Confusion Matrices for Predictive ToxicologyApplications.*Proceedings of the IEEE 12th International Conference on Intelligent Data Engineering and Automated Learning*, 252-261.

McHugh, J. (2000). Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory. *ACM transactions on Information and system Security,* 3, 262-294.

Mell, P., Hu, V., Lippmann, R., Haines, J. and Zissman, M. (2003). An Overview of Issues in Testing Intrusion Detection Systems. Technical Report NIST Interagency Reports 7007, Department of Commerce, National Institute of Standards and Technology, USA.

Mirkovic, J., Prier, G. and Reiher, P. (2002). Attacking DDoS at the Source. *Proceedings of the 10th IEEE International Conference on Network Protocols*, 312-321.

Mirkovic, J. and Reiher, P. (2004). A taxonomy of DDoS Attack and DDoS Defense Mechanisms. *ACM SIGCOMM Computer Communication Review,* 34, 39-53.

Moore, D., Shannon, C., Brown, D. J., Voelker, G. M. and Savage, S. (2006). Inferring Internet Denial ofService Activity. *ACM Transactions on Computer Systems,* 24, 115-139.

Morin, B. and Mé, L. (2007). Intrusion Detection and Virology: An Analysis of Differences, Similarities and Complementariness. *Journal in computer virology,* 3, 39-49.

Munz, G. and Carle, G. (2007). Real-time Analysis of Flow Data for Network Attack Detection. *Proceedings of the 10th IFIP/IEEE International Symposium on Integrated Network Management*, 100-108.

Münz, G., Weber, N. and Carle, G. (2007). Signature Detection in Sampled Packets. *Workshop on Monitoring, Attack Detection and Mitigation*, Citeseer, Toulouse, France.

Muraleedharan, N., Parmar, A. and Kumar, M. (2010). A Flow based Anomaly Detection System Using Chi-square Technique. *Proceedings of the2nd International IEEEConference on Advance Computing*, 285-289.

Nguyen, H. A., Tam Van Nguyen, T., Kim, D. I. and Choi, D. (2008). Network Traffic Anomalies Detection and Identification with Flow Monitoring. *Proceedings of the 5thIEE/IFIP International Conference on Wireless and Optical Communications Networks*, 1-5.

NLS-KDD. (2015). *The NSL-KDD Data Set* [Online]. Available: http://iscx.cs.unb.ca/NSL-KDD/.

nmap.org. (2015). *Nmap Security Scanner* [Online]. Available: http://nmap.org/.

Park, J., Tyan, H.-R. and Kuo, C.-C. (2006). Internet Traffic Classification for Scalable QoS Provision. *Proceedings of the International IEEE Conference on Multimedia and Expo*, 1221-1224.

Park, K. and Lee, H. (2001). On the Effectiveness of Route-based Packet Filtering for Distributed DoS Attack Prevention in Power-law Internets. *ACM SIGCOMM Computer Communication Review*,31(4), 15-26.

Paxson, V. (1999). Bro: A System for Detecting Network Intruders in Real-time. *Computer Networks,* 31, 2435-2463.

Portnoy, L., Eskin, E. and Stolfo, S. (2001). Intrusion Detection with Unlabeled Data Using Clustering. *In Proceedings of ACM CSS Workshop on Data Mining Applied to Security*, Citeseer.

Quinlan, J. R. (1993). *C4. 5: programs for machine learning*, San Francisco: Morgan Kaufmann.

Quittek, J., Zseby, T., Claise, B. and Zander, S. (2004). Requirements for IP Flow Information Export (IPFIX). RFC 3917, IETF. Available: https://www.rfc-editor.org/rfc/rfc3917.txt

Reich, Y. and Fenves, S. J. (1991). The Formation and Use of Abstract Concepts in Design. *Concept formation: knowledge and experience in unsupervised learning*. San Mateo: Morgan Kaufmann.

Rhodes, B. C., Mahaffey, J. A. and Cannady, J. D. (2000). Multiple Self-organizing Maps for Intrusion Detection. P*roceedings of the 23rdNational Information Systems Security Conference*, 16-19.

Robertson, S., Siegel, E. V., Miller, M. and Stolfo, S. J. (2003). Surveillance Detection In High Bandwidth Environments. *Proceedings of the DARPA Information Survivability Conference and Exposition*, 130-138.

Roesch, M. (1999). Snort Lightweight Intrusion Detection for Networks. *Proceedings of the 13th USENIX conference on System Administration*, 229-238.

Satten, C. (2008). *Lossless Gigabit Remote Packet Capture With Linux* [Online]. Available: http://staff.washington.edu/corey/gulp/.

Schaffrath, G. and Stiller, B. (2008). Conceptual Integration of Flow-based and Packet-based Network Intrusion Detection. *Resilient Networks and Services.* Springer.P*roceedings of the2nd International Conference on Autonomous Infrastructure, Management and Security*, 190-194.

Shannon, C. E. (2001). A Mathematical Theory Of Communication. *ACM SIGMOBILE Mobile Computing and Communications Review,* 5, 3-55.

Shiravi, A., Shiravi, H., Tavallaee, M. and Ghorbani, A. A. (2012). Toward Developing a Systematic Approach to Generate Benchmark Datasets for Intrusion Detection. *Computers and Security,* 31, 357-374.

Sinclair, C., Pierce, L. and Matzner, S. (1999). An Application of Machine Learning to Network Intrusion Detection. P*roceedings of the15th IEEE Annual Conference on Computer Security Applications*, 371-377.

Snort Incorporation. (2013). *SNORT Users Manual 2.9.5* [Online]. Available: http://manual.snort.org/.

Softflowd. (2013). *Softflowd* [Online]. Available: http://www.mindrot.org/projects/softflowd/.

Sommer, R. and Paxson, V. (2010). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. *Proceedings of theIEEE Symposium on Security and Privacy*, 305-316.

sourceforge.net. (2014). *NFDUMP* [Online]. Available: http://nfdump.sourceforge.net/.

sourceforge.net. (2015). *NfSen - Netflow Sensor* [Online]. Available: http://nfsen.sourceforge.net/.

Sperotto, A., Sadre, R., van Vliet, F. and Pras, A. (2009). A Labeled Dataset for Flow-based Intrusion Detection. *Proceedings of the 9th IEEE International Workshop on IP Operations and Management,*39-50.

Sperotto, A., Schaffrath, G., Sadre, R., Morariu, C., Pras, A. and Stiller, B. (2010). An Overview of IP Flow-based Intrusion Detection. *IEEE Communications Surveys and Tutorials,* 12(3), 343-356.

Staniford, S., Hoagland, J. A. and McAlerney, J. M. (2002). Practical Automated Detection of Stealthy Portscans. *Journal of Computer Security,* 10, 105-136.

Stoecklin, M. P., Le Boudec, J.-Y. and Kind, A. (2008). A Two-layered Anomaly Detection Technique Based on Multi-modal Flow Behavior Models. *Proceedings of the 9th InternationalConference on Passive and Active Network Measurement,* 212-221.

Stolfo, S. J., Fan, W., Lee, W., Prodromidis, A. and Chan, P. K. (2000). Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project. *Proceedings of theDARPA Information Survivability Conference and Exposition*, 130-144.

Stone, R. CenterTrack: (2000). An IP Overlay Network for Tracking DoS Floods. *Proceedings of USENIX Security Symposium*, 21, 114.

Strayer, W. T., Lapsely, D., Walsh, R. and Livadas, C. (2008). Botnet Detection Based on Network Behaviorin Series: Advances in Information Security, Springer, 1-24.

Tavallaee, M., Bagheri, E., Lu, W. and Ghorbani, A.-A. (2009). A Detailed Analysis of the KDD CUP 99 Dataset. *Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications*, 53-58.

Tcpdump/Libpcap. (2015). *TCPDUMP & LIBPCAP* [Online]. Available: http://www.tcpdump.org/.

The Shmoo Group, the DefCon. (2015). *Capture the Capture the Flag Data* [Online]. Available: http://cctf.shmoo.com/.

University of California, (1999). *KDD Cup 1999 Data* [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

Wagner, A. and Plattner, B. (2005). Entropy Based Worm and Anomaly Detection in fast IP Networks. *Proceedings of the14thIEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise*, 172-177.

Wang, H., Zhang, D. and Shin, K. G. (2002a). Detecting SYN Flooding Attacks. *Proceedings of the21stIEEE Annual Joint Conference on Computer and Communications Societies*, 1530-1539.

Wang, H., Zhang, D. and Shin, K. G. (2002b). SYN-dog: Sniffing SYNFlooding Sources. *Proceedingsof the 22ndIEEE International Conference on Distributed Computing System*, 421-428.

Wang, Y. and Global, I. (2009). *Statistical Techniques for Network Security: Modern Statistically-based Intrusion Detection and Protection*, Hershey: IGI Global.

Wireshark.org. (2015). *Wireshark* [Online]. Available: https://www.wireshark.org/.

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G. and Cunningham, S. J. (1999). Weka: Practical Machine Learning Tools and Techniques with Java Implementations.*Proceedingsof theInternational Workshop on Emerging Knowledge Engineeringand Connectionnist-based Information Systems*, 99, 192-196.

Yau, D. K., Lui, J., Liang, F. and Yam, Y. (2005). Defending Against Distributed Denial of Service Attacks with Max-Min Fair Server-centric Router Throttles. *IEEE/ACM Transactions on Networking,* 13, 29-42.

Zhang, Z., Li, J., Manikopoulos, C., Jorgenson, J. and Ucles, J. (2001). HIDE: a Hierarchical Network Intrusion Detection System using Statistical Preprocessing and Neural Network Classification. *Proceedingsof the IEEE Workshop on Information Assurance and Security*, 85-90.

Zhao, Q., Xu, J. and Kumar, A. (2006). Detection of Super Sources and Destinations in High-speed Networks: Algorithms, Analysis and Evaluation. *IEEE Journal on Selected Areas in Communications,* 24, 1840-1852.