UNIVERSITI PUTRA MALAYSIA

ANOMALY BEHAVIOR DETECTION USING FLEXIBLE PACKET FILTERING AND SUPPORT VECTOR MACHINE ALGORITHMS

MOHAMMED N. ABDUL WAHID

FSKTM 2016 12

ANOMALY BEHAVIOR DETECTION USING FLEXIBLE PACKET FILTERING AND SUPPORT VECTOR MACHINE ALGORITHMS

By

MOHAMMED N. ABDUL WAHID

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

April 2016

COPYRIGHT

DEDICATION

To My Dearest and First Teachers: My father and Mother. To My Lovely Wife "Ban" and to my beloved daughter 'May'

I will always be grateful for your endless love, unlimited support and deep faith on me

Mohamed Abdulwahid Alimam

Abstract of thesis presented to Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

ANOMALY BEHAVIOR DETECTION USING FLEXIBLE PACKET FILTERING AND SUPPORT VECTOR MACHINE ALGORITHMS

By

MOHAMMED N. ABDUL WAHID

April 2016

Chairman : Azizol Abdullah, PhD
Faculty : Computer Science and Information Technology

Many of the present network visitors' analysis have the capability to capture extraordinary forms of traffics. The main method is related to processing and filtering data packets using different types of packet filtering on network system and, more specifically, capturing and filtering data packets transmitted on high speed communications links for errors and attackers' detection and signal integrity analysis. Many anomaly detection experiments have been conducted in order to investigate the performance of network traffics filtering methods that analyzing bandwidth, speed, errors and attackers. These experiments are performed and examined under different network environments such as methods in Traffic Analysis and Monitoring (TAaM) and Entropy and Support Vector Machine (EaSVM). Both methods used DARPA 98-99 dataset and Lincoln Labs data. However, these methods are limited to analyze the entire traffic as one entity, which makes them unable to quantify network anomalies. Furthermore, Network traffic prediction algorithms based on SVM such as EaSVM have commented about the fundamental difficulties in achieving an accurate declaration that defines anomaly which suppose to solve the problem of the high rate of false positive alarm and finding excellent ways that guarantees to clear up pending issues of the network traffic normality such as the alluvial data noise of the TAaM method.

Filtering traffics and detecting anomalies in real time environments using single machine is the motivation behind this research. A unique method that combines the Flexible Packet Filtering (FPF) with Support Vector Machine (SVM) algorithm is proposed to classify the behavior of the network traffics. The methodology is to use the maximize margin of SVM algorithm to alert for the presence of attack, and the proposed flexible packet filtering, which is a combination of both static and dynamic packet filtering was the method that has been followed to filter network traffics based on anomaly behavior. The User Profile Filter (UPF) was proposed to aid the SVM algorithm to classify the captured traffics into normal and abnormal behavior. The Network Traffic Analysis is the tool that allows users to monitor and view the network traffics details. The proposed FPFaSVM method and TAaM depending on the network

traffic analyzer to capture and analyze the network traffics, and a special technique that detects anomalies while monitoring network traffics have been proposed by both methods using DARPA 99 dataset and real environments. The Entropy and SVM (EaSVM) is relying on DARPA 99 dataset for analyzing the captured traffics based on anomalies and the use of SVM is to classify the entropy values of the data traffic into normal and abnormal behavior for more accurate results. FPF of SVM have merged the analyzed results of flexible packet filtering with support vector machine algorithm to get better classification of the captured network traffics and to detect anomalies. The proposed packet filtering (FPF) will isolate the captured traffics based on their source using traffic source separation 'TSS' strategy, during the separating operation the traffic signature will be examined with the stored signatures of the system database using Traffic Signature Matching (TSM).

The experiment results shows that by using a User Profile Filter (UPF) that will be based on SVM and examining the traffic signature, the total of error received from the traffic classifier has been reduced to 0.5%, and the traffic capturing speed has been increased by 10% as well as the total bandwidth captured per minute in comparing with TAaM and EaSVM.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

ANOMALI GELAGAT PENGESANAN MENGGUNAKAN ALGORITMA PENAPISAN PAKET FLEKSIBEL DAN MESIN PENYOKONG VEKTOR

Oleh

MOHAMMED N. ABDUL WAHID

April 2016

Pengerusi : Azizol Abdullah, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Ramai daripada analisis tingkah laku rangkaian ini mempunyai keupayaan untuk menangkap jenis lain daripada yang lain daripada trafik. Kaedah utama yang berkaitan dengan pemprosesan dan data penapisan paket pada sistem rangkaian dan, lebih khusus, menangkap paket data yang dihantar pada pautan komunikasi berkelajuan tinggi untuk mengesan kesilapan dan penyerang dan analisa integriti isyarat. Penapisan paket dan pengesanan anomalii dalam persekitaran masa nyata dengan satu mesin menggunakan algorithma FPF dan SVM adalah motivasi bagi penyelidikan ini. Penggunaan algoritma SVM sebagai pengkelas bagi tingkahlaku tarfik rangkaian mempunyai banyak isu-isu yang mencabar yang perlu di ambil kira untuk menghasilkan keadah yang baik. Jadi, penggunaan margin yang maksimum algorithma SVM bagi memaklumkan kehadiran serangan dengan cadangan penapisan paket yang fleksibel yang menapis trafik rangkaian berasaskan tingkah laku anomali adalah idea bagi penyelidikan ini. Banyak eksperimen pengesanan anomali telah dijalankan untuk menyiasat prestasi rangkaian memperdagangkan kaedah penapisan yang menganalisis jalur lebar, kelajuan, kesilapan dan penyerang. Eksperimen ini dilakukan dan diperiksa di bawah persekitaran rangkaian yang berbeza seperti TAaM dan EaSVM. Kedua-dua kaedah yang digunakan DARPA 98-99 set data dan data Lincoln Labs. Walau bagaimanapun, kaedah ini adalah terhad untuk menganalisis keseluruhan trafik sebagai satu entiti, yang membuat mereka tidak dapat mengukur anomali rangkaian. Tambahan pula, trafik rangkaian algoritma ramalan berdasarkan SVM seperti EaSVM telah mengulas tentang kesukaran asas dalam mencapai pengisytiharan tepat yang mentakrifkan anomali yang sepatutnya menyelesaikan masalah kadar yang tinggi penggera positif palsu yang dianggap sebagai salah satu isu yang paling mencabar di anomali rangkaian. Oleh itu, idea kajian ini adalah untuk mencadangkan penapisan paket fleksibel yang merupakan gabungan kedua-dua paket statik dan dinamik menapis dengan margin dimaksimumkan sokongan algoritma mesin vektor untuk mengesan tingkah laku anomali dan mengelaskan trafik ke dalam tingkah laku normal dan tidak normal.

Rangkaian Analisis Trafik adalah alat yang membolehkan pengguna untuk memantau dan melihat rangkaian memperdagangkan maklumat lanjut. Kaedah yang dicadangkan FPF dan Taam bergantung kepada penganalisis trafik rangkaian untuk menangkap dan menganalisis trafik rangkaian, dan teknik khas yang mengesan anomali manakala trafik pemantauan rangkaian telah dicadangkan oleh kedua-dua kaedah menggunakan DARPA 99 set data dan persekitaran sebenar. The Entropy dan SVM (EaSVM) adalah bergantung kepada kaedah Taam untuk menganalisis trafik yang ditangkap berdasarkan anomali menggunakan set data yang sama dan penggunaan SVM adalah untuk mengelaskan nilai entropi trafik data untuk hasil yang lebih tepat. FPF daripada SVM telah bergabung keputusan dianalisis penapisan paket fleksibel dengan sokongan algoritma mesin vektor untuk mendapatkan klasifikasi yang lebih baik daripada trafik rangkaian ditangkap dan untuk mengesan anomali. Penapisan paket yang dicadangkan (FPF) akan mengasingkan trafik ditangkap berdasarkan sumber mereka menggunakan pemisahan sumber trafik 'TSS' strategi, semasa operasi yang memisahkan tandatangan trafik akan diperiksa dengan tandatangan disimpan dalam pangkalan data sistem menggunakan Traffic Signature Matching. Keputusan eksperimen menunjukkan bahawa dengan menggunakan Profail Pengguna Penapis (UPF) yang akan berdasarkan SVM dan memeriksa tanda tangan lalu lintas, jumlah ralat yang diterima daripada pengelas lalu lintas itu telah dikurangkan kepada 0.5%, dan kelajuan menangkap lalu lintas telah meningkat sebanyak 10% dan juga jumlah bandwidth yang ditangkap seminit dalam membandingkan dengan TAaM dan EaSVM.

ACKNOWLEDGEMENTS

First and foremost, praise is for Allah S.W.T for giving me the strength, guidance and patience to complete this thesis. Many blessing and peace be upon Prophet Mohammad SallaAlahu Alaihi Wasallam, the prophet that has been sent mercy to the world.

I would like to express my sincere gratitude to my supervisor Dr. Azizol Abdullah for the continuous support of my study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me all the time during my research and his advices improved my writing skills while preparing this thesis.
I would like to thank the supervisory committee members starting with Assoc. Prof. Dr. Nur Izura Udzir and Assoc. Prof. Dr. Zuriati Ahmed Zukarnain for their encouragement and insightful comments as well as Dr. Jalil Desa for his helpful information and advices.

I am very grateful to the faculty of Computer Science and Information Technology and the staff of postgraduate office, library, and University Putra Malaysia, for providing the best research environment. Thanks to every person who has supported me to successfully produce my thesis.

I am very grateful to my family: My Father Nazeh Alimam, My Mother Dr. Fayhaa, and for my friends for their unflagging love and support throughout my life. I have no suitable words that can fully describe my feelings to them except, I appreciate all of you.

Words fail me to express my appreciation to my lovely wife Ban whose dedication, love and persistent confidence in me, has taken the load off my shoulder. I owe her for being unselfishly let her intelligence, passions, and ambitions collide with mine.

Finally, I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

Mohamed Nazeh

April 2016

I certify that a Thesis Examination Committee has met on 12 April 2016 to conduct the final examination of Mohammed N. Abdul Wahid on his thesis entitled "Anomaly Behavior Detection Using Flexible Packet Filtering And Support Vector Machine Algorithms" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Md Nasir bin Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Abd. Rahman bin Ramli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Eyas El-Qawasmeh, PhD
Professor
King Saud University
Saudi Arabia
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 August 2016

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Azizol hj Abdullah, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Zuriati Ahmed Zukarnain, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Nur Izura Udzir, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- This thesis is my original work;
- Quotations, illustrations and citations have been duly references;
- This thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- Intellectual property from the thesis and copyright of the thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- Written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceeding, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- There is no plagiarism or data falsification / fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Mohammed N. Abdul Wahid, GS29205

Declaration by Members of Supervisory Committee

This is to confirm that:

- The research conducted and the writing of this thesis was under our supervision;
- Supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of
Chairman of
Supervisory
Committee: Dr. Azizol hj Abdullah

Signature: _____
Name of
Member of
Supervisory
Committee: Assoc. Prof. Dr. Zuriati Ahmed Zukarnain

Signature: _____
Name of
Member of
Supervisory
Committee: Assoc. Prof. Dr. Nur Izura Udzir

TABLE OF CONTANTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Program Interface
BSS	Blind Source Separation
COI	Community of Interests
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
D-DoS	Distributed-Denial of Service
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
DS	Distribution System
EaSVM	Entropy and Support Vector Machine
EWMA	Exponential Weighted Moving Average
FPF	Flexible Packet Filtering
FTP	File Transfer Protocol
GLR	Generalized Likelihood Ratio
HTTPS	Hypertext Transfer Protocol Secure
ICMP	Internet Control Massage Protocol
IDS	Intrusion Detection System
IGMP	Internet Group Management protocol
IP	Internet Protocol
IPS	Intrusion Prevention System
IS	Information Server
ISP	Internet Service Provider
LAN	Local Area Network
LibPcap	Library Packet Capture
MAC	Media Access Control
MIB	Management Information Base
MRTG	Multi Router Traffic Graph

NFS	Network File System
PC	Personal Computer
PCA	Principle Component Analysis
PIX	Private Internet Exchange
QoS	Quality of Service
RBFNN	Radial-Basis Function Neural Network
RFC	Request For Comment
RM	Remote Monitoring
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SP	Services Provider
SQL	Structured Query Language
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
TSS	Traffic Source Separation
TTL	Time To Leave
UDP	User Datagram Protocol
UPF	User Profile Filter
WLAN	Wireless Local Area Network
WWAN	Wireless Wide Area Network

CHAPTER 1

INTRODUCTION

1.1 The Motivation behind Network Security

Network protection turns into probably the most preferred property that obtained extra attention for protecting the identification of 1's companions in verbal exchange. This is a primary property. Consider an example; there are two companies in the existing market. They both compete in the market. If they share texts for various works and such exchange of thoughts may be lead to various expertises and one word may be beneficial for the other. The same is with the various users using internet. They don't want their surfing activities be checked continuously and such information kept by an unauthorised third party. Whereas, this all is done to check if any sort of anonymous activities aren't done. Thus, it has both positive consequence for the one whereas not for the other. This study involves web browsing, or more generally for areas where low latency, interactive communication is needed (electronic mail protocols, for illustration, in general do not require this). These protocols ought to be effective as good as cover the identities of the two communicating parties (what URL a exact user is accessing). The proposed filtering mechanism is a web-based network traffic analyser used to generate instant statements on network visitors and customers. Furthermore, exact subtle elements on information assessment, information elucidation, and graphical presentation of result to connect them, and creates diagrams and surveys that backing in making sense of and investigating the captured traffics. It empowers clients to watch transfer speed and traffics in an interface certain level. The selectable chart permits zooming in on the separating guests, likewise demonstrates the information focuses, which gives the guests traffic IN and site traffics OUT essential focuses practically identical to speed, volume of traffics and use off the entire transfer speed and protocol.

This study describing the essence of network visitors traffics analysis, and its capacity to capture and reveal all of the traffics (incoming and outgoing), and in addition how it's in a position to observe any style of suspicious movements which are coming from the network community similar to intruders or some other undesirable applications prompted by means of misused or anomalies. Many of the present network visitors analysis has the capability to capture extraordinary forms of traffics [10, 11, and 14] and exhibit the full details on the distinctive interface, these traffics are subdivided into two types:

1. Ordinary traffics: these traffics are essentially the most coming traffics from the community.
2. Abnormal traffics: These kinds of traffics are seldom when detected from the network considering the fact that the existence of firewall and other safety contraptions comparable to IDS, IPS.

The second sort of traffics are also known by using misuse or anomaly habits and each one among them has its own algorithms that supposed to be utilized on top of the structure of the network traffic analysis to be equipped to seize and notice like these types of attacks or threats.

The most usually deployed approaches by which the cyber attacks are detected and such detection helps in preventing various cyber terrorism. This all is done with the help of signature-primarily based technique for detection [84]. Such ways can handiest notice the before diagnosed assaults which have a signature corresponding to it, given that the database of signature is revised manually for every new style of assault it really is determined. By such barriers, the interest in intrusion detection processes is gaining a positive response and the process is installed on knowledge mining [10, 22 and 50]. Data mining based intrusion detection structures usually fall into one in all two classes; misuse detection and anomaly detection.

1. In misuse detection, every and every instance in a facts set is labeled as 'typical' or 'intrusive' and a studying algorithm is informed over the categorized records. those techniques are ready to routinely retrain intrusion detection models on specific enter understanding that comprise new varieties of assaults, so long as they were categorised successfully. A key competency of misuse detection procedures with their excessive degree of accuracy of detecting diagnosed attacks and their versions. Their apparent concern is the disability to word assaults whose instances have now not but been found.
2. Anomaly detection techniques, however, assemble fashions of traditional knowledge and are aware deviations from the usual mannequin in located records. Anomaly detection applied to intrusion detection and pc safety has been a lively subject of have a look at for the reason that it changed into as soon as at first proposed with the aid of denning [17].

Anomaly detection algorithms have the competencies of discovering new sorts of intrusions as deviations from normal utilization [4, 10, and 11]. In this quandary, given a suite of ordinary know-how to educate from, and given a contemporary piece of scan expertise, the reason of the intrusion detection set of rules is to evaluate whether or not to take a look at data belong to "common" or to an anomalous behaviour. However, anomaly detection schemes undergo from a high value of false alarms [1, 3, 4 and 10] and so forth. This happens certainly considering that previously unseen method behaviours are moreover well-known as anomalies, and for that reason flagged as talents intrusions.

1.2 Analysing Traffics and Offences Detection

Generally, when attack by virus or an assault by hacker occurs. Generation of similar samples or the "signature" of packets is generated. These packets can be easily picked up by a network analyser. This analyzer can further alert the administrator by a mail. Maximum analysers permit you to add alarms, while a distinct pattern is observed. Various analysers also can be taught to deliver an e-mail or net page. At the time, various conditions are also fulfilled. By this assume that virus and its signature were visible previously and incorporated the analyser's packet filters' list (a filter out specifies the set of standards beneath which an analyser will capture packets or prompt an alarm or some other movement).

Intrusion detection or intrusion prevention includes various ways by which the attacks on the computer and infrastructure of network are removed. This is done by anomaly

detection. This method is an important method as it keeps a check and tells of intentional and non-intentional attacks [4, 7, 50 and 81], faults and defects. A few articles recognition on a delegated comparative take a look at multiple schemes for detection which on the same hand provide a unique network intrusion [3]. a few present supervised and unsupervised anomaly detection schemes and their variants are evaluated at the DARPA 1998-1999 records set of network connections [16, 18 and 19] and actual community facts utilising cutting-edge trendy assessment strategies as excellent as making use of a couple of targeted metrics which might be correct whilst detecting assaults that involve a massive number of connections.

This study is just excessively precarious, making it impossible to establishment anyway it is convenient to be taught and start working with. It can put in on a home windows or Linux machine, and utilize only a web program to passage the customer interface. Subsequent to walking the application inside seconds, guest's charts are plotted and audits are mechanically produced through "Yield and assessment work". To have hitter making sense of, more data should be expressed in little print concerning the consequences of the site guests assessment starting from the source that the guests has been produced from with the exception of it goes to the last stride of the methodology. The quick advancement of the web in estimation, intricacy and guests frames has made group administration a troublesome test [4, 5, and 6]. The capability of an observing framework to outfit right know-how with respect to the nature and sort of the group movement can't be over underlined. Understanding about who is delivering likely the most guests, what conventions are being used, and the assortments of ambushes that have been caught amid the observing and the spot is the guests starting from or where is the excursion spot of the activity can be extremely vital for altering clog issues [79, 81 and 84] and distinctive issues affected through assailants. Numerous system executives spend different times, hoping to recognize what is debasing the execution of their group. A typical procedure to blockage obstacle is to enhance system framework, i.e. [39] substitute servers with extreme end servers and widen the transmission capacity. This arrangement was steeply-evaluated, fleeting and does not scale [10]. As soon in light of the fact that the redesign is done the clog fundamental issue will toughen for an even as and later consistently crumble on the grounds that the clients trade their behaviour taking into account the change. The substitute answer for this worry is to introduce a versatile system guests observing and examination approach [22, 63] keeping in mind the end goal to comprehend the elements of the activity and adjustments in the web and aggregate parity of the group. Moreover to realize the wellbeing status of the group, observing of system leisure activity additionally has the upsides of distinguishing DoS and data transfer capacity robbery assaults. With a perspective to propensities examination of broad scope of group practices, it's indispensable to assemble group guests on an unfaltering basis then again than as an onetime occasion which least difficult catches transient practices traffics that presents knowledge into system issues. Accumulating long run group site guest's information will outfit profitable aptitude for bettering and making sense of the real system progression [3].

New viruses and worms have one-of-a-kind signatures counting on the vulnerabilities they may be trying to make the maximum, but as soon as strategies were successfully breached, there are pretty small problems of topics that hackers sincerely need to do together with your network, the pinnacle ones being:

- use your structures in a denial of service (dos) on a third party user. A well designed network analyser can readily set up such strategies via the use of the site visitors they generate.
- use your technique as an FTP server to distribute “warez” and one of kind unlawful files.

The administration can configure an analyser to look for FTP visitors or visitor’s volume the place it is surprising. The very nature of viruses and worms is to produce distinguished levels of network visitors. immoderate frequency of broadcast packets or unique servers generating an distinct wide variety of packets are logged in the analyser’s file of long run site visitors, permitting the administrator to follow up on suspicious web site visitors patterns.

1.3 Problem Statement

Mostly all the current Network Traffic Analysis has the ability to chase different types of traffics. The network traffics are situated on anomaly detection approaches are limited to research the whole site visitors as one or single entity [1, 4, 79, 81], by this, they don't even quantify the network anomalies, and their validities are affected whilst many anomalous occasions get up concurrently.

The network site filters must be equipped to dynamically elicit the noise data that usually travels with the network traffics from LAN to another. Mostly all anomaly detection algorithms require a suite of in primary terms traditional expertise to teach the model [4, 10, 20, 21 and 50] and in order that they implicitly expect that anomalies may be dealt with as patterns now not positioned before. Also the anomaly investigator system should no longer alert administrator for the presence of assault except it is validated via manner of a couple of statement. the primary difficulties in attaining an correct announcement of an intrusion to remedy the quandary of the immoderate false positive alarm [1, 4 10, 20, 21, 22 and 50] because of the difficulties to set any predefined rules for creating a desire on effectively attack traffics considering there is no principal difference among traditional and attack normal network site visitors.

The venture of establishing effective and robust network site visitor’s prediction and monitoring methodologies is commonly tricky if we consider the following factors:

(a) customers can exchange their conduct slowly with this method and evolution of time (e.g.) In a network, the guest user can just give out changes and editions, and for this reason, any algorithm associated needs to be equipped of dynamically adapting those alterations and evolutions;

(b) The analysis and prediction about the expected visitors will have to be centred on usual knowledge, whilst all data which might be noise/anomalies and may just have an impact on the accuracy and correctness of the ordinary conduct prediction must be excluded or labelled as average.

1.4 Research Objectives

This research emphasises on the design and progress and superior mechanism that can be utilised to beef up the accuracy and the prediction of network site visitor's normality. In practical, this study makes a speciality of anomaly detection situated on flow monitoring and accordingly of the total anomaly detection methodology, above all in circumstances the place high burstiness is present. To begin with, this study proposing a mechanism that presents visitors source separation and filtering centred on 'frequency area' to investigate the captured network traffics. This approach is headquartered on the remark or the dataset that the quite a lot of network site visitors components, are higher recognized, represented and isolated within the frequency area [16]. Mainly, when isolating the traffics into two main accessories: the baseline element and the quick term aspect.
So, the intention is:

1. To propose a flexible packet filtering that analyse and realize suspicious events with instant identification of the attacker traffic source and network protocol that committed the assault.
2. To propose user profile filter (UPF) in packet filtering to be able to be centred on Support vector computer algorithm to take care of noise issues. UPF will aid in performing anomaly detection on unlabelled information via observing a person profile and evaluating the routine during an intrusion to the routine for the period of the common use and then every person profile will likely be report on the check list database. So, here the target is to use the UPF so one can filter and examine every single profile relying upon the documents of the common consumer conduct to increase the detection of novel attacks and reducing the false alarm rate as a lot as viable for the ambiguity conduct at the same time monitoring the traffics.
3. To propose an enhanced algorithm of SVM as a classifier for the proposed packet filtering to restrict the false alarm price utilising the maximize margin of SVM algorithm. Also with the aid of enhancing the basic suggestion of network visitors evaluation [22, 50 and 63] and applying new techniques into the structure of packet filtering, similar to traffic signature matching (TSM) and site visit the traffic source separation (TSS) to minimize the excessive fee of false alarm and confirming the incidence of assault.

The algorithm of SVM must be ready to periodically and dynamically adapting the alterations of user conduct and the maximize margin of SVM will help in lowering the false alarm brought about by means of these alterations. This algorithm will be applied over the UPF and TSS to be certain that the captured traffics small print are nontoxic and the visitors reputation is demonstrated whether it belongs to normal or abnormal habits 'anomaly habits'.

An additional objective which assisting this analyzer is to overcome the weaknesses that have located within the earlier system corresponding to when observe suspicious endeavour or irregular traffics the action towards those traffics is constrained through notifying administration for the presence of assault. As well as the unconfirmed type of that assault. And this function considered as a testing method for the proposed packet filtering.

1.5 Research Scope

The essential mission of the network visitors assessment is the capability of taking pictures and monitoring all the network traffics (incoming and outgoing) for local area network (LAN) and extensive field network (WAN). This research is focusing on LAN and left the WAN for the long term work. The proposed FPF has used the attributes and the parameters of every static and dynamic packet filtering to analyse and recognize error and suspicious activities referring to intruders. The FPF will decode the packet header and payload to measure traffic bandwidth, time and packet loss using packets decode [63] and may have a look at the website online site visitor's four metrics (traffic 4 metrics which are, total packet, total byte, D-socket and D-port) to discover transferring mistakes and/or types of attacks. The class of the site network visitors is founded on SVM set of rules. SVM that has the record of the customer of the local users profile (UPF) will classify the filtered know-how of the FPF the use of the maximize margin to confirm the conduct of the LAN visitors. This examine is completed through manner of specializing in many problems which might be regarding monitoring and detecting anomalies tremendously on the use of flexible packet filtering that clear out the captured network traffics. The proposed packet filtering may also isolate the captured traffics founded on their source using traffic source separation 'TSS' approach, for the duration of the separation operation the visitors signature will be tested with the saved signatures of the system database using traffic signature matching (TSM).

The proposed method is relying on the network website visitors' analyser to seize and examine the network traffics and DARPA99 dataset for checking out and contrast disorders. These studies have proposed a precise approach that detects anomalies at the same time as monitoring network traffics measurements; this technique is the flexible packet filtering of support vector machine. so, this research have merged the analysed consequences for each of the flexible packet filtering and the support vector set of rules that used to get the less complicated type of the captured network traffics and to find out anomalies. The contrast has been executed with TAaM and EaSVM [4, 10, 50, 63, and 64] using identical dataset and same environments.

1.6 Contributions

The main contributions of this thesis can be described as the subsequent:

1. To come out with a method such a Flexible Packet Filtering for the growth and detection of all the network anomalies by combining the contributions of the static and the dynamic packet filtering into Flexible packet, filtering of the network traffics analysis.
2. A method that will take care of the data noise problems and associate in reducing alarms that caused by traffic miss-analysing called User Profile Filter (UPF).
3. Improved algorithm of SVM is proposed to shape the attribution of the network traffics analysis application to be operating upon the flexible packet filtering to provide effective anomaly detection and visitors type with minimizing the rate of fake alarm.

4. A contemporary mechanism will likely be applied on top of the network analyser to make it prepared for dynamically adapting to the exchange of client habits as properly as new checking out technique that would expose any analyser software to be confirmed in the direction of assaults or a few different threats that clients would face while browsing the net.

1.7 Thesis outline

Chapter one shows a clear history about network visitors monitoring to have higher working out for the progress of network traffic analysis. This chapter discusses the existed problems in the network site visitors evaluation method and any other boundaries as well because the proposed ambitions that resolve these weaknesses to discover anomalies and delivering more comfy environment.

Chapter Two discusses the literature evaluate part and defines the way that has been adopted to gather the resources and expertise to start working with, and explaining the use the packet sniffer to sniff the community traffics, and the varieties of network site visitors filter that are to be had to be used to observe anomalies. The related works part in this chapter shows what different researchers have been performed on this area. And the part of the question of this assignment is to be trained how the process is supposed to work and to find out the results from making use of it.

Chapter Three discussed the study methodology that explaining the methods followed to implement the software and the Bi-directional verbal exchange. Simulation instruments stated easy methods to behavior protocol decodes and analysis (packet decode), and shrewd sniff to start capture traffics. The important section in this chapter is easy methods to follow the bendy packet filtering with aid vector machine. The explanation of find out how to use packet generator on the way to help to generate the irregular traffic will likely be offered as well on this chapter. Additionally the performance metrics with other parameters that used within the constitution of the process is explained.

Chapter four containing very most important know-how; opening with approach design and implementation of network visitors traffic analysis, and can show the procedure interface to view the traffics important points. After that, the dialogue in regards to the operating environment and the process requisites is also mentioned on this chapter to define probably the most suitable specifications the sort of hardware and application specifications to furnish better environment to begin working with.

Chapter five discussing the outcome that have obtained from this process equivalent to protocols bandwidth fee graphs, and outcome evaluation between the general procedure with the proposed method and a further one is for making use of packet generator. All of them are explained intensive in separate section in that chapter.

Chapter Six is the conclusion and future work of this research. As researcher discussed the primary features which have concluded from this study and every other ideas for the longer term work that maybe the factor of curiosity of other resear

REFERENCES

[1] A. Dainotti and A. Pescapé, “Issues and future directions in traffic classification,” IEEE Network, vol. 26, pp. 35-40, January 2012.

[2] A.Hyvaärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[3] Allen W.H., Marin G.A., Rivera, L.A., Automated detection of malicious reconnaissance to enhance network security, South east Con Proceedings, IEEE, 2005.

[4] Altyeb Altaher, Sureswaran Ramadass, Bhavani Thuraisingham, Mohammad Mehedy, 28-30- Oct, 2011, On Line anomaly Detection based on relative entropy’” Broadband Network and Multimedia Technology (IC-BNMT), 2011 4th IEEE International Conference, pp 33-36. **

[5] Anderson D, et al. Detecting unusual program behavior using the statistical component of the Next-generation Intrusion Detection Expert System (NIDES). Computer Science Laboratory SRI-CSL 95-06, 1995.

[6] Anderson, D., Frivold, T. & Valdes, A. 1995, 'Next-generation Intrusion Detection Expert (NIDES) A Summary,' SRI INTERNATIONAL Computer Science Laboratory.

[7] Anderson, J. P. 1980, 'Computer Security Threat Monitoring and Surveillance,' James P. Anderson Company.

[8] B.J. Oommen, R.L. Kashyap, A formal theory for optimal and information theoretic syntactic pattern recognition, Patter Recognition 31 (1998) 1159.

[9] Barros AK, Vigario R, Jousmaki V, Ohnishi N. Extraction of event-related signals from multi-channel bioelectrical measurements. IEEE Transactions on Biomedical Engineering 2000; 47 (5):583–8.

[10] Basant Agarwal, Namita Mittal, Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques, Procedia Technology 6 (2012) 996 – 1003. **

[11] Brauckhoff D, Tellenbach B, Wagner A, Lakhina A, May M. Impact of traffic sampling on anomaly detection metrics. In: Proceedings of internet measurement conference, Janeriro, Brazil, 2006. p. 159–64.

[12] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Springer (2006).

[13] Chandalia G, Rish I. Blind source separation approach to performance diagnosis and dependency discovery. In: Proceedings of the 7th ACM SIGCOMM internet measurement conference, California, USA, 2007. p. 259–64.

[14] Cisco NetFlow. At www.cisco.com/warp/public/732/Tech/netflow/2008-2013.

[15] Cormode G, Muthukrishnan S. An improved data stream summary: the count-min sketch and its applications. In: Proceedings of Latin American theoretical informatics, Buenos Aires, Argentina, 2004. p. 29–38.

[16] DAPRAdataset,/http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/docs/detections_1999.htmlS.

[17] Denning, D. E. 1987, 'An Intrusion-Detection Model,' Software Engineering, IEEE Transactions on, vol. 13, no. 2, pp. 222-232.

[18] Duffield N, Lund C, Thorup M. Estimating flow distributions from sampled flow statistics. In: Proceedings of the ACM SIGCOMM, Karlsruhe, Germany, 2009. p. 325–336.

[19] E. Al-Shaer, H. Hamed, Firewall policy advisor for anomaly detection and rule editing, in: Mohammed Proc. IEEE/IFIP 8th Int. Symp. Integrated Network Management, IM 2003, March 2012, pp. 17–30.

[20] Francesco Palmieri*, Ugo Fiore Universita` degli Studi di Napoli Federico II, CSI, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126, Napoli, Italy Network anomaly detection through nonlinear analysis, computers & s e c u rity 29 (2010) 737-755.

[21] G. Nychis et al.,'" An imperial evaluation of entropy based-anomaly detection". Proceedings of the 8th ACM 2008 SIGCOMM conference on Internet measurement,, ACM Press,, pp 151-156.

[22] G. Li. Wei "Intrusion Detection Based on Information Entropy of Multiple Support Vector Machine. School of mathematics and computer science, Yunnan University of Nationalities. International journal of computer and information technology. Volume 03-issue 01, January 2014.

[23] H. Bunke, J. Csirik, Parametric string edit distance and its application to pattern recognition, IEEE Transactions on Systems, Man and Cybernetics 25 (1995) 202.

[24] H. Lewis, C. Papadimitriou, Elements of the Theory of Computation, 2nd edition, Prentice Hall, New York,1997.

[25] H. L. Sun, Y. H. Jin, Y. D. Cui, S. D. Cheng, Journal of Beijing University of Posts and Telecommunications, vol. 33, no. 1, pp. 7–11, 2010.

[26] H. Yang, F. Xie,:2010" Clustering and classification based on anomaly detection" international conference on Fuzzy Systems and Knowledge Discovery P 1082-1091.

[27] H.S. Javitz, A. Valdes, The SRI IDES Statistical Anomaly Detector, in: Proceedings of the IEEE Symposium on Security and Privacy, May 1991, IEEE Press, 1991.

[28] ISCONetFlow:/http://www.cisco.com/en/US/products/ps6601/productswhitepaper09186a00800a3db9.shtmlS. 2013.

[29] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA (2004).

[30] John W, Tafvelin S. Analysis of internet backbone traffic and header anomalies observed. In: Proceedings of the 7th ACM SIGCOMM conference on internet measurement, San Diego, USA, 2007. p. 111–6.

[31] Jung J, Paxson V, Berger A, Balakrishnan H. Fast port scan detection using sequential hypothesis testing. In: Proceedings of the IEEE symposium on security and privacy, California, USA, 2004. p. 211–25.

[32] J. Zhang, Y. Xiang, and Y. Wang, "Network traffic classification using correlation information," IEEE Trans. on Parallel and Distributed Systems, vol. 24, pp. 104-117, 2013.

[33] K. Begnum, M. Burgess, A scaled, immunological approach to anomaly countermeasures, in: Proceedings of the VIII IFIP/IEEE IM Conference on Network Management, 2003.

[34] Kim S, Reddy A. A study of analyzing network traffic as images in real-time. In: Proceedings of the IEEE INFOCOM, Miami, USA, 2005. p. 2056–67.

[35] Kohler E, Li J, Paxson V, Shenker S. Observed structure of addresses in IP traffic. IEEE/ACM Transactions on Networking 2006;14(6):1207–18.

[36] Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies. In: ACM SIGCOMM; 2004.

[37] Lee , Dong Xiang, May 14-16, 2001, Information-Theoretic Measures for Anomaly Detection , Proceedings of the 2001 IEEE Symposium on Security and Privacy, p.130.
[38] Lee D, Brownlee N. Passive measurement of one-way and two-way flow lifetimes. ACM SIGCOMM Computer Communication Review 2007;37(3):17–28.
[39] Lee W, Xiang D. Information-theoretic measures for anomaly detection. In: Proceedings of the IEEE symposium on security and privacy, Oakland, USA, 2010. p. 130–43.
[40] Liu H, Feng W, Huang Y, Li X. A peer-to-peer traffic identification method using machine learning. In: Proceedings of the networking, architecture, and storage, Guilin, China, 2007. p. 155–60.
[41] Liu Y, Towsley D, Ye T, Bolot J. An information-theoretic approach to network monitoring and measurement. In: Proceedings of the 5th ACM SIGCOMM internet measurement conference, Berkeley, USA, 2005. p. 1–14.
[42] Li Wei, School of Mathematics and Computer Science, journal article; Intrusion Detection Based on Information Entropy of Multiple Support Vector Machine, International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 03 – Issue 01, January 2014.
[43] M. Roesch, Snort, Intrusion Detection System 2014, http://www.snort.org .
[44] M. Sloman, Policy driven management for distributed systems, Journal of Network and Systems Management 2 (1994) 333.
[45] M.J. Ranum et al., Implementing a generalized tool for network monitoring, in: Proceedings of the Eleventh Systems Administration Conference (LISA XI), USENIX Association, Berkeley, CA, 1997, p. 1.
[46] Mahoney MV. Network traffic anomaly detection based on packet bytes. In: Proceedings of the ACM-SAC 2003, Melbourne, USA, 2003. p. 346–50.
[47] Marcus Ranum, "False Positives: A User's Guide to Making Sense of IDS Alarms," *white paper*, ICSA labs IDSC, February 2003.
[48] McDaniel P, Sen S, Spatscheck O, Kaashoek F. Enterprise security: a community of interest based approach. In Proceedings of the 3th annual network and distributed system security symposium, California, USA, 2006. p. 1–15.
[49] Miao Xie, SongHan. BimingTian, SaziaParvin. Digital Ecosystem sand Business Intelligence Institute, Curtin University, DEBII, GPOBoxU1987, Perth,WA6845, Australia. Anomaly detection in wireless sensor networks: A survey, journal of network and computer applications 34 (2011), 1302-1325.
[50] Muamer N. Mohammed a *, Norrozila Sulaiman b, Intrusion Detection System Based on SVM for WLAN. Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, 26300, Kuantan, Malaysia, Procedia Technology 1 (2012) 313 – 317.
[51] N. Damianou, N. Dulay, E.C. Lupu, M. Sloman, and Ponder: a language for specifying security and management policies for distributed systems, Imperial College Research Report Do C 2000/1, 2000.
[52] Nychis G, Sekar V, Andersen D, Kim H, Zhang H. An empirical evaluation of entropy-based traffic anomaly detection. In: Proceedings of the 8th ACM SIGCOMM internet measurement conference, Vouliagmeni, Greece, 2008. p. 151–6.

[53] O. Hunaidi, J.H. Rainer, G. Pernica, Measurement and analysis of traffic-induced vibrations, in: Proceedings of the 2nd International Symposium on Network Traffic Noise and Vibrations, St. Petersburg, Russia, 1994, pp. 103–108.

[54] P. Barford, J. Kline, D. Plonka, A. Ron, A signal analysis of network traffic anomalies, 2002.

[55] P. D'haeseleer, An immunological approach to change detection: Theoretical results, in: 9th IEEE Computer Security Foundations Workshop, 1996.

[56] P. D'haeseleer, S. Forrest, P. Helman, An immunological approach to change detection: algorithms, analysis, and implications, in: Proceedings of the 1996 IEEE Symposium on Computer Security and Privacy, 1996.

[57] P. Hoogenboom, J. Lepreau, Computer system performance problem detection using time series models, in: Proceedings of the USENIX Technical Conference, USENIX Association, Berkeley, CA, 1993, p. 15.

[58] P.A. Porras, P.G. Neumann, EMERALD: Event monitoring enabling responses to anomalous live disturbances, in: Proc. 20th NIST-NCSC National Information Systems Security Conference, 1997, pp.353–365.

[59] P.G. Neumann, P.A. Porras, Experience with EMERALD to date, 2000, pp. 73–80.

[60] Perf SONAR: Performance focused Service Oriented Network monitoring Architecture ,http://www.perfsonar.net

[61] Q.A. Tran, H. Duan, and X. Li, 2004. One-class Support Vector Machine for Anomaly Network Traffic Detection the 2nd network Research Workshop of the 18th APAN, Cairns, Australia.

[62] Qian Quan, Che Hong-Yi, Zhang Rui, 2009,"intrupy based method for network anomaly detection" International Symposium on Dependable Computing,, pp. 189 – 219.**

[63] Qin T, Guna X, Li W, Wang P. Monitoring abnormal traffic flows based on independent component analysis. In: Proceedings of the international conference on communications, Dresden, Germany, 2009. p. 1–5.**

[64] R. C. Chen and S. P. Chen, 2013" Intrusion Detection Using a Hybrid Support Vector Machine Based on Entropy and TF-IDF. // international journal of innovative computing information and control (IJICIC), Vol. 4, no2, pp.413-424.**

[65] R. Durbin, S. Eddy, A. Krigh, G. Mitcheson, Biological Sequence Analysis, Cambridge University Press, Cambridge, 1998.

[66] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley Inter science, New York, 2001.

[67] Ringberg H, Soule A, Rexford J, Diot C. Sensitivity of PCA for traffic anomaly detection,. ACM SIGMETRICS Performance Evaluation Review 2007; 35 (1):109–20.

[68] S.-H. Han, M.-S. Kim, H.-T. Ju, J.W.-K. Hong, The architecture of ngmon: A passive network monitoring system for high-speed ip networks, in: IFIP/IEEE 13th International Workshop on Distributed Systems: Operations and Management, DSOM 2002, 2002, p. .

[69] Schweller R, Li Z, Chen Y, Gao Y, Gupta A, Zhang Y, Dinda P, Kao M, Memik G. Reversible sketches: enabling monitoring and analysis over high-speed data streams,. IEEE/ACM Transactions on Networking 2007;15(5):1059–72.

[70] Siris VA, Papagalou F. Application of anomaly detection algorithms for detecting SYN flooding attacks. Global Telecommunication Conf 2004b;29(3):2050e4.

[71] Sodiya A., Longe H. and Akinwale A., A new two-tiered strategy to intrusion detection, Information Management & Computer Security, Vol. 12 No. 1, pp. 27-44 (2004).
[72] Somayaji, S. Forrest, Automated response using system-call delays, in: Proceedings of the 9th USENIX Security Symposium, 2000, p. 185.
[73] SrinoyS., Intrusion Detection Model Based On Particle Swarm Optimization and Support Vector Machine. The IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA), Vol.1 No. 5, pp. 86-192 (2007).
[74] Tan G, Poletto M, Kaashoek F, Guttag J. Role classification of hosts within enterprise networks based on connection patterns. In: Proceedings of the 2003 USENIX annual technical conference, Washington, USA, 2003. p. 15–28.
[75] TCPDump/libpcap‘‘TCPDUMPPublic Repository’’,http://www.tcpdump.org/2009.
[76] Thottan M, Ji C. Anomaly detection in IP networks,. IEEE Transactions on Signal Processing 2011; 51(8):2191–204.
[77] Tsang, G.C.Y.; Chan, P.P.K.; Yeung, D.S.; Tsang, E.C.C., 26-29 Aug. 2004, "Denial of service detection by support vector machines and radial-basis function neural network," Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on ,vol.7, no., pp. 4263- 4268 vol.7.
[78] Valdes A. and Skinner K., Probabilistic alert correlation, Recent Advances in Intrusion Detection (RAID), Springer-Verlag, Davis, CA, 2001.
[79] Varun Chandola, Arindam Banerjee, vipin kumar, July 2009, anomaly detection a survey, journal ACM Computing Surveys (CSUR), Volume 41 Issue 3.
[80] Weng Ling , SVM Engineering Training Center, Harbin Polytechnic University, Haerbin, China. Journal of Chemical and Pharmaceutical Research, 2014, 6(7):2294-2303.
[81] W. Tafvelin S. Analysis of internet backbone traffic and header anomalies observed. In: Proceedings of the 7th ACM SIGCOMM conference on internet measurement, San Diego, USA, 2007. p. 111–6.
[82] Xiaoling Tan, Weijian Fang, Yong Qu, traffics monitoring and SVM. International Journal of Advancements in Computing Technology, Vol. 5, No. 5, pp. 183-190, 2013.
[83] X. T. Chen, J. X. Liu, Journal on communications, vol. 32, no. 4, pp. 153–157, 2011.
[84] X. Y. Yang, S. S. Yang, J. Li, packet sniffer of network analyzer project. Chinese journal of computers, vol. 34, no. 2, pp. 395–405, 2011.
[85] Ye N, Vilbert S, Chen Q. Computer intrusion detection through EWMA for auto correlated and uncorrelated data. IEEE Transactions on Reliability 2003; 52(1):75–82.
[86] Youxin LUO, Xiaoyi CHE, Lingfang LI, Journal of Convergence Information Technology, 2012, Vol. 7, No. 22, pp. 484-491.
[87] Yuan Z., GuanX., Accurate classification of the internet traffic based on the SVM method, in: Proceedings of the 42th IEEE International Conference on Communications (ICC), 2007.
[88] Zhang Y, Ge Z, Greenberg A, Roughan M. Network anemography. In: Proceedings of the 5th ACM SIGCOMM conference on Internet measurement, Berkeley, CA, USA, 2005. p. 317–30.
[89] Zou C, Gong W, Towsley D, Gao L. The monitoring and early detection of internet worms. IEEE/ACM Transactions on Networking 2005;13(5):961–74.