A Blacklist Process Model to Enhance the IDS Using Similarity Measurements

Enas Ayman Al-Utrakchi, Mohammad Rasmi Al-Mousa
Zarqa University, Jordan

Abstract

Nowadays, Intrusion Detection Systems (IDSs) are used as main security perspective in order to detect any breaches and to measures security level for most of the organizations. However, limitations exist within IDS frequently rendering them inaccurate in detecting attacks; the primary reason for this has been false negative or false positive alarms. The accuracy of the IDS enhanced through increasing of the true positive alerts which will effect to decreasing the false negative and false positive alarms of the intrusions'. The purpose of this paper is to introduce a new blacklist model in order to enhance the accuracy of the IDS. Based upon experimental testing of sample attacks, this paper proposes a new process model for updating a blacklist based on the extraction of patterns through comparison to known malicious activities and code, using similarity measurement against a predefined database and network traffic, after analysis and classification of anomalies. The proposed model shows continued promise in making analysis of network traffic more efficient, and IDS critically more accurate through increasing the true positive alarms.

1. Introduction

Maintaining the network security systems and protecting information from intruders, and threats as spyware is important [22]. In general, rapidly using of Internet Technology (IT) and the ever increasing skill of attackers, previously unknown threats to network stability continue to abound. Because of this, the protection of Information Systems (IS) against malicious activities and attacks in networks has never been more critical. There already exist many software tools to strengthen and maintain network security, such as Anti-Virus, Firewalls, IDS, etc... [23].

This paper focuses on Intrusion Detection System (IDS); IDS is generally structured as second stage of defense. The design intent of IDS is to detect attacks and notify administrators whenever there is suspicious or abnormal traffic.

The most prominent limitation of IDS to date has been false readings of code and activity (either false positive or false negative) which result in inaccurate attack detection events. IDS is, therefore not perfectly equipped to detect new attacks. We suggest that if IDS designers strategically integrate techniques based in similarity theory IDS should realize increased detection accuracy.

Defensive security approaches such as Intrusion Detection System (IDS) and Intrusion Protection System (IPS) were developed to detect, prevent, and establish a perspective of network attacks. Intrusion must be analyzed more intensively to generate accurate data profiles describing each attack and to establish a more suitable decision-making power within the IDS.

Intrusion analysis techniques (which were originally developed to enhance IDS) provide details about the traits of the attack and the behavior of the attacker. Analysis of attack intention is a prime example of the focus of intrusion analysis.

Typical items of interest to intrusion analysis include IP addresses, ports, type of services and protocol, etc. The most common techniques for intrusion analysis depend on determining the features from the attack path as reported by [14-18]. The drawback of these techniques is that they are not suitable for large numbers of features. It has thus far been limited to handling specific types of features, and has yet to develop to an operationally feasible state, capable of efficiently presenting all the distinctive of an attack. With that limitation in mind, these techniques do work well with specific types of attacks, such as Distributed Denial of Service (DDOS) attacks.

In general, attack analysis is a critical and challenging task in security management [10]. The limited capability of security sensors and network monitoring tools make attack observation inaccurate and render the output incomprehensible.

This paper, presents a set of processes, shown in figure 1. that apply a similarity measurement to identify new attacks and add them efficiently to a blacklist. The proposed model will be described in section 3 which defines the internal processes of the model. In section 4, we will explain the most salient ideas and issues in the current body of research. Finally, our experiments and discussion of findings are provided, based upon samples of statistical data describing known attack features.

2. Related works

This section reviews literature on the network security, attacks and threats, intrusion detection systems and attack similarity theory.

Since the advent of the Internet, the number of LANs and personal computers has increased dramatically and this, in turn has given rise to a global, virtual environment, vulnerable to substantial security risks which arrive through networks. Firewall devices which impose an access control policy between two or more networks via software or hardware were a reaction to this new threat reality. Firewalls were primarily aimed at e-mail and Web surfing and sought to control inbound and outbound access to the Internet.

Network security is critically important in today's world because it is secures all information passing through networked computers. Network security involves with all hardware and software functions, characteristics, features, operational procedures, accountability measures, access controls, administrative and management policy required to provide satisfactory level of safety for hardware, software, and information in a network [1].

Design and implementation of a network security model was presented in [2] using routers and firewall. Within that function, it also identified:

- The network security weaknesses in router and firewall network devices,
- The type of threats arriving at entry points guarded by the firewall,
- The system's responses to those threats, and
- The method to prevent the attacks and hackers from accessing the network.

The proposed model in [2] provides a checklist to use in evaluating whether a network is adhering to best practices in network security and data confidentiality. However, the model aims to protect the network from vulnerabilities, threats, attacks, configuration weaknesses and security policy weaknesses.

Generally, the authors of [3] started with the current situation of network security and analyzed its most influential elements to provide references the development of new models for securing computational property within networks.

An intrusion detection system detects various kinds of malicious network traffic and computer usage that cannot be detected by a conservative firewall. Thus, some researchers have focused on different types of attacks on IDS and given descriptions of different attacks aimed at different protocols such as TCP, UDP, ARP and ICMP [4].

Whereas, the authors of [5] has concentrated on the design and develop the Intrusion Detection System for detecting Distributed Denial of Service (DDoS) Attacks in the network using Jpcap library in Java Programming language.

A new incremental hybrid intrusion detection system was proposed by [6]. This framework combines incremental misuse detection and incremental anomaly detection. The framework can learn new classes of intrusion that had not existed in previous data used for training incremental misuse detection. The framework has lower computational complexity so it is suitable for real-time or on-line learning.

Most intrusion analysis approaches are based on alert correlation techniques. These techniques are connected to network tools for assistance (such as IDS) to understand and analyze the intrusion event. The drawback of most of these techniques is that they are developed to prevent future attacks and minimize damage [11, 12, 13]. Thus, innovative methods and techniques are needed in the analysis of the attacks to increase the accuracy of the IDS through pre-analysis and the establishment of a robust intrusion blacklist in advance – this would provide the greatest help to IDS in their decision making.

Three novel algorithms based on the threshold algorithm were introduced in [7], it exploited the semantic properties of the new similarity measures to achieve the best performance in theory and practice.

Tanimoto based a similarity measure for host-based intrusions on a binary feature set for training and classification introduced by [8]. The k-nearest neighbor (KNN) classifier has been utilized to classify a given process either as normal or as an attack.

The Agile Similarity Attack Strategy (ASAS) model proposed by [9] heuristically identifies and monitors similar evidence between a new criminal case and others. The model uses a classification method based on a relation between attack evidence priorities with evidence group values, presented as a vector. Furthermore, the model uses a cosine similarity as a distance-based similarity measure (Metric Axioms) to improve the quality of decision making.

3. Blacklist Process Model

This section presents a new proposed model called the intrusion blacklist model. The proposed model adopts network capturing tools such as Wireshark, and Snort, as Network Intrusion Detection Systems (NIDSs). Network traffic is captured in the initial phase through network capturing tools, which normally produce a huge array of security data. This study will be empirical, using Snort, a free, open source system, whose core function is to monitor network traffic and detect attempts at intrusion. Where one or more tools within Snort will be used in each process (either in

stand-alone status or collaboratively with tools from other software packages which were tested working alongside Snort), there will be a series of processes which will lead to updating the intrusion blacklist. The proposed model includes five processes, i.e. collection, analysis, detection, classification, and similarity.

In the collection process, network traffic will be compiled using Snort which in turn converts network data into a coded form, pcap or log file, during real time network data traffic capture. The capture is then routed for analysis. Packets will be analytically read using tools include libraries such as libpcap to extract features according to signature and then stored to a database. This process is important because all subsequent processes will rely upon it; furthermore the useful selection of features at this stage will affect the work of snort.

During the detection process, operations are performed to detect abnormal activity and possible attacks in packets by matching features with rules ranked within a hierarchy – this process is one of the main stage in NIDS, since it takes responsibility assigning a proper action in response to each case intrusion: either log the packet or send an alarm notification. Naturally, if the packet is normal will be ignored. Once the full process is concluded, an output file will be created logging all alerts generated by NIDS (Snort). However, the rules of the NIDS can be controlled and amended by the user to improve result over time [19].

In the classification process, the NIDS generates many alerts some of which will be irrelevant. During this stage abnormal packets will be noted if categorized as a serious attack or not. Using k-nearest neighbor classifier in this way can detect a serious attack and minimize false alarms, then store them into database [20].

Finally, the similarity analysis process will be used to evaluate similar, abnormal packets through comparison with a predefined database, using Jaccard with Euclidean distance to estimate similarity of the most recently discovered attack features.

The Jaccard coefficient proposed at 1901 is still widely used in the various fields such as ecology and biology [26]. The Jaccard index, also known as "the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. This index only uses presence-absence data"[24].

$$\text{Jaccard}'\text{sim}(A, B) = P(A \cap B) | P(A \cup B) \quad (1)$$

The Jaccard similarity uses a measure of the share properties of both Objects A and B whereas all of the Objects A and B given by 0 and 1 respectively [25]. The proposed model will be improve IDS work through data refinement and the reduction of false alarms. The output will be an updated Blacklist database whose aim is to prevent future intrusive attacks [21].

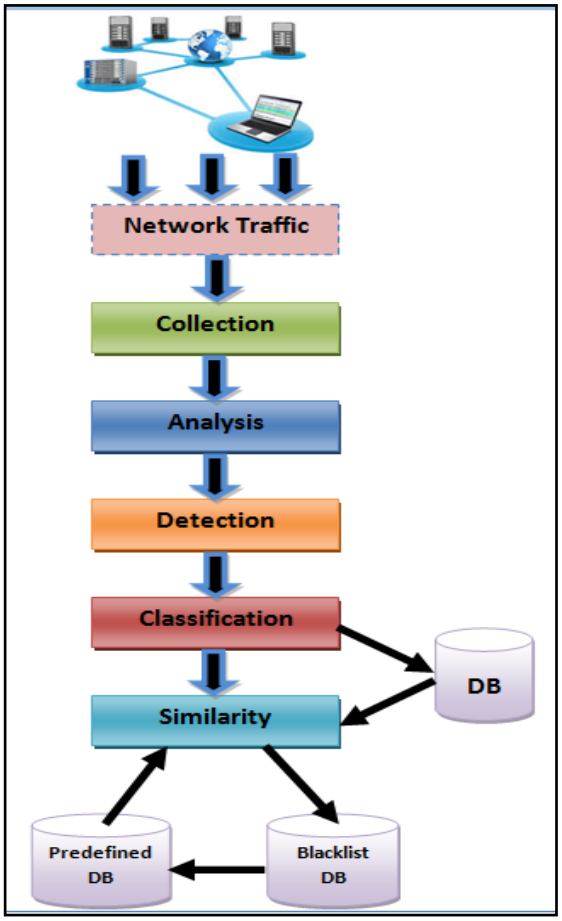

Figure 1. Blacklist Process Model

4. Experimental Results and Analysis

This section evaluates the proposed model based on its analysis of a sample of 50 different types of attack. Each attack will be analyzed and extracted to identify 10 different type of features, such as, source IP address, destination IP address, port number, vulnerability, alert priority, time of life, type of protocol, type of service, …etc.

Each feature will be numbered from 1-10, and it will be weighted for subsequent calculation based on criteria that rank the importance of each feature and the frequency of its occurrence.

Table 1 shows the similarity features' of the new attack with one of a predefined attack database (attack number 23). The similarity percentage will be calculated using similarity measurement as a real-valued function that quantifies the similarity between

the suspicious attack described in the Snort alerts database and the observed attack.

Table 1. Example of Similarity of the Attack Features

Feature #	Feature Similarity
1	12 %
2	30%
3	1%
4	0%
5	0.02%
6	0.22%
7	0%
8	0.32%
9	50%
10	2%

The proposed model compute the similarity percentage of each predefined attack then select the maximum similarity of the new attack with the other predefined attack. As an example the predefined attack number 23 has a maximum similarity with a new attack in 95.56% as shown in Table 2.

Table 2. Example of the Similarity of the new Attack

Attack Number	Similarity with New Attack
7	43.12 %
16	11.70%
23	95.56%
36	91.43%
41	8.09%

The evaluation of the results shows that the proposed blacklist model provides useful information and increases the possibility of detecting the real attack. Moreover, it helps IDSs eliminate the most similar features of the intrusions based on the similarity of attack features. Hence, it helps to improve the decision making process and the accuracy of the IDS.

5. Conclusion and future work

This paper proposes a new model to pre-analyze attacks during network traffic. The proposed model expects to make intrusion detection more accurate, an invaluable asset to IDS and all users. The proposed model will improve the quality of IDS decision making, in order to obtain clear information and achieve acceleration of the intrusion detection. Attack analysis is a critical and challenging task in network security management. Furthermore, features of attack recognition and analysis are an important research area in the field of network security. Obviously, to gain a higher ratio of intrusion detection, deeper analysis is desirable, as are more efforts to identify features of new attacks using suitable network security tools.

This paper is expected to conclude that most intrusion analysis approaches are based on alert correlation techniques which are used to understand and analyze the intrusion occurrence. Thus, the contribution of the research is anticipated to be the formulation of new, innovative methods and techniques aimed at increasing the accuracy of the IDS in order for it to be improved as a strong preliminary intrusion analysis tool capable of establishing a more reliable intrusion blacklist before actual attacks occur, and to thereby to help the IDS in its decision making.

10. References

[1] Chen S., Iyer R., and Whisnant K., "Evaluating the Security Threat of Firewall Data Corruption Caused by Instruction Transient Errors," In Proceedings of the 2002 International Conference on Dependable Systems & Network, Washington, D.C., 2002.

[2] Alabady S., "Design and Implementation of a Network Security Model for Cooperative Network," International Arab Journal of e-Technology, Vol. 1, No. 2, June 2009

[3] Zeng A., “Discussion and research of computer network security”, Journal of Chemical and Pharmaceutical Research, 2014, 6(7):780-783

[4] Anand A., Patel B., "an Overview on Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocols," International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, Issue 8, Aug. 2012

[5] G. D. K., Rao CV, Singh M. k. and Kemal M., “Network-based IDS for Distributed Denial of Service Attacks,” International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol. 3, Issue 1, Jan- Feb 2014

[6] Hadjieleftheriou M., Chandel A., Koudas N., Srivastava D., “Fast indexes and algorithms for set similarity selection queries,” in: Proceedings of the 24th International Conference on Data Engineering (ICDE ’08), pp. 267–276.2008

[7] Rasoulifard A., Bafghi A. G., and Kahani M., “Incremental Hybrid Intrusion Detection Using Ensemble of Weak Classifiers, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, vol. 6, pp. 577–584. 2008

[8] Sharma A. and Lal S. P., “Tanimoto based similarity measure for intrusion detection system,” Journal of Information Security, 2(4):195–201. 2011

[9] Rasmi M. and Jantan A., “Asas: agile similarity attack strategy model based on evidence classification for

network forensic attack analysis, AWERProcedia Information Technology & Computer Science 846-857. 2012

[10] Qin, X. & Lee, W. , "Attack Plan Recognition and Prediction Using Causal Networks". Computer Security Applications Conference, 2004, 370-379.

[11] Wei, W. & Thomas, E. D., "A Graph Based Approach Toward Network Forensics Analysis" , ACM Trans. Inf. Syst. Secur., 2008, 12, 1-33.

[12] Huang, M.-Y., Jasper, R. J. & Wicks, T. M., "A Large Scale Distributed Intrusion Detection Framework Based On Attack Strategy Analysis.", Computer Networks, 1999, 31, 2465-2475.

[13] Damiano, B., Sandro, E. & Pieter, H. H., "Panacea: Automating Attack Classification for Anomaly-Based Network Intrusion Detection Systems.", Proceedings of The 12th International Symposium On Recent Advances In Intrusion Detection. Saint-Malo, France, Springer-Verlag, 2009, 1-20.

[14] Wu, P., Zhigang, W. & Junhua, C., "Research On Attack Intention Recognition Based On Graphical Model.", Fifth International Conference On Information Assurance and Security, 2009. IAS '09.360-363.

[15] Feng, J., Yuan, Z., Yao, S., Xia, C. & Wei, Q., "Generating Attack Scenarios for Attack Intention Recognition.", International Conference On Computational and Information Sciences. Chengdu, China, IEEE Computer Society, 2011, 272-275.

[16] Hao, B., Kunsheng, W., Changzhen, H., Gang, Z. & Xiaochuan, J., "Boosting Performance In Attack Intention Recognition By Integrating Multiple Techniques.", Front. Comput. Sci China, 2011, 5, 109-118.

[17] Peng, W., Yao, S. & Chen, J., "Recognizing Intrusive Intention and Assessing Threat Based On Attack Path Analysis." International Conference: Multimedia Information Networking and Security, 2009. MINES '09, 250-253.

[18] Wang, Z. & Peng, W., "An Intrusive Intention Recognition Model Based On Network Security States Graph", 5th International Conference On Wireless Communications, Networking and Mobile Computing, 2009. WICOM '09, 1-4.

[19] Rehman R. U., "Intrusion Detection with SNORT: Advanced IDS Techniques Using SNORT, Apache, MySQL, PHP, and ACID", Prentice Hall PTR , 2003.

[20] Gabra H. N., Bahaa-Eldin A. M. and Korashy H., " Classification of IDS Alerts with Data Mining Techniques", International Journal of Electronic Commerce Studies Vol.5, No.1, pp.1-6, 2014.

[21] Soldo F., Le A. and Markopoulou A.," Predictive Blacklisting as an Implicit Recommendation System", INFOCOM, Proceedings IEEE, PP 1 – 9,2010.

[22] Al-Sammerai, N. F.(2011). Development a Network-based Intrusion Prevention System using Data Mining Approach. Master thesis, Computer Science, Amman Arab University, Jordan.

[23] Wang, S. P., &Ledley, R. S. (2012). Computer architecture and security: Fundamentals of designing secure computer systems. John Wiley & Sons.

[24] Dharavath, Ramesh, and Abhishek Kumar Singh. "Entity Resolution-Based Jaccard Similarity Coefficient for Heterogeneous Distributed Databases." InProceedings of the Second International Conference on Computer and Communication Technologies, pp. 497-507. Springer India, 2016.

[25] Niwattanakul, S., Singthongchai, J., Naenudorn, E., &Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1, p. 6).

[26] Choi, S. S., Cha, S. H., &Tappert, C. C. (2010). A survey of binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics, 8(1), 43-48.