International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 489-494 Jan 2019-June 2019 www.csjournals.com

Improving the Accuracy of Attack Detection Using Machine Learning

Shaziya Shaheen[1], Prof.Naziya Pathan[2], Prof. Anuja Ghasad[3]
[1] M-Tech.scholar, [2]Asstt.Prof.(Guide), [3]Asstt.prof
Departement of computer science and engg.,RTM Nagpur University,Nagpur(MS),India.

Abstract: Attacks in networks are increasing day by day due to advancement in software computational architectures and the number of users exposed to hacking. Due to this fact, systems should be smart enough to identify and block these attacks, so that the system security is not hampered. Usually communication in nodes occurs with the help of data packets, and thus the analysis of these data packets leads to the identification of attack packets and thus identifying attackers in a wireless network. In this paper, we propose a machine learning based algorithm which uses agents to identify and remove attacks from a network using packet analysis. The proposed algorithm uses agent based mechanism and is very flexible. Results demonstrate that the proposed algorithm can identify and remove attacks from almost any kind of network with high level of accuracy.

Keywords: Attacks, network, machine learning, identification.

1. Introduction

Systems and networks are subject to electronic attacks. Today's information systems in government and commercial sectors are distributed and highly interconnected via local area and wide area computer networks. While indispensable, these networks provide potential avenues of attack by hackers, international competitors, and other adversaries. The increasingly frequent attacks on Internet visible systems are attempts to breach information security requirements for protection of data. Intrusion detection technology allow organizations to protect themselves from losses associated with network security problems

Intrusion detection systems (IDSs) are software or hardware systems that automate the process of monitoring the events occurring in a computer system or network, analyzing them for signs of security problems. As network attacks have increased in number and severity over the past few years, intrusion detection systems have become a necessary addition to the security infrastructure of most organizations.
Although firewalls have traditionally been seen, as the " first line of defense" against would be attackers, intrusion detection software is rapidly gaining ground as a novel but effective approach to making your networks more secure. Intrusion detection operates on the principle that any attempt to penetrate your systems can be detected and the operator alerted - rather than actually stopping them from happening. This is based on the assumption that it is virtually impossible to close every potential security breach; intrusion detection takes a very " real world" viewpoint, emphasizing instead the need to identify attempts at breaking in and to assess the damage they have caused.

The next section demonstrates various intrusion detection (IDS) techniques for wireless networks, followed by the proposed approach for improving the IDS accuracy, and finally we conclude the paper with some finer observations and some future work which can be carried out by researchers in order to further improve the reliability of the system.

2. Literature survey

Every single realized assault can be spoken to by certain examples. This [1] recognition plot looks at the examples and furthermore identifies the comparative examples (diverse variations of a similar example). It is not the same as infection recognition since it identifies comparative examples too [1]. It is an or more purpose of this plan it is solid for distinguishing the known examples and furthermore it can identify a portion of the obscure assaults, yet at the same time it can't adapt to the majority of the obscure assaults which may happen.

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 489-494 Jan 2019-June 2019 www.csjournals.com

In abuse identification, rule based methodology [2] is broadly utilized. In this procedure assaults are spoken to in various arrangements of guidelines. Principles set are made and later contrasted with various assaults with identify nearness of interruption. Standard based framework [3] requires profoundly talented programming procedures to refresh the guidelines. Along these lines, state change based plan was begun to conquer the downsides of guideline based framework. In state change, assaults are symbolized as a progression of occasions that are lead by the assailant having some underlying state to the last state. The states compare to the focused on framework that speaks to all the memory areas of the framework as appeared in Figure 1. In this situation it is accepted that aggressor must have some consent to get to the system and all entrance manual for the obtaining of some capacity that the assailant does not have preceding the assaults.

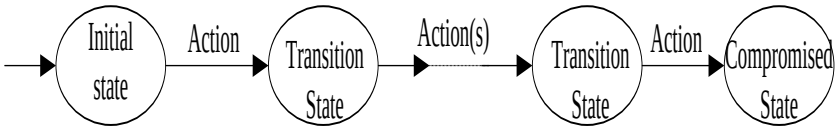


Figure 1: State Transition Diagram

Scientists in [4] saw abuse recognition as example coordinating. Their model is extremely nonexclusive and pertinent to any all around characterized configurations of info occasions, for example, review trail records, arrange parcels and so on. In this every signature spoke to as an instantiation of a Colored Petri Automaton. The thought of at least one begin and novel last states, and way between them characterize the arrangement of activities coordinated by the net. Adaptability, versatility, simplification are the primary points of interest of this coordinating plan. The major and normal disadvantage [5] of abuse discovery approaches is that they all are composed for their very own particular condition and can't function admirably for other people. To address this issue deliberation based calculation were presented. The main endeavour of deliberation based is versatile ongoing abuse location framework (ARMD). It is have based abuse identification framework, which gives language stage to marks and techniques that make an interpretation of these marks into checking program. Oddity discovery is the correlation of a conduct with some watched conduct so as to recognize interruption [6]. It is more grounded than abuse location; since it has capacity to recognize inconspicuous assault. Measurable models are one of the most punctual techniques which are utilized for interruption recognition. In this model it is accepted that aggressor conduct is unique in relation to the ordinary client, their factual techniques can be utilized to recognize typical conduct to unusual one. There are two factual models which are utilized in interruption location. Initial one is the continuous IDS having measurable segment dependent on master framework (NIDES/STAT) [7]. It breaks down conduct of the system in ordinary mode and waitlists hubs whose conduct is discovered fluctuating. The critical change or deviation from the normal conduct is hailed and treated as a potential interruption.

Sheaf [8] then again dissects client exercises as indicated by four stages. At first, it produces insights dependent on client sessions in particular session vectors. Next, it produces Bernoulli vector to describe characteristics which are not implied for that particular session. After that it doles out loads to interruptions types dependent on happened recurrence. Ultimately, it creates doubt remainder to speak to how that session is suspicious when contrasted with different sessions for explicit interruption types. AI based strategies help in free distinguishing proof and amalgamation of assemble data dependent on models, either certain or express to recognize design investigation. The said data is set apart to prepare the conduct demonstrate as needs be for applying severe request on assets to such an extent that bad conduct is recognized powerfully. Master framework approaches are broadly utilized instances of information based framework. Master framework characterizes review information as indicated by their standard sets. It includes three stages. First it recognizes distinctive classes and qualities from the prepared information based on which set of characterization rules are created and parameters and capacity are made sense of. In conclusion, the review information is ordered appropriately.

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 489-494 Jan 2019-June 2019 www.csjournals.com

Motivation behind the objective observing framework is to screen the progressions made to some particular program as opposed to distinguishing the abuse or oddity. Wherever this framework is sent, the chairman doesn't have to screen the framework persistently. For observing the alterations, trustworthiness checksum hashes can be registered either for every one of the records or for some particular document, in light of the necessity. There might be a few assailants which continue checking the framework for a significant lot of time. Motivation behind this system is to recognize such assailants [9]. Such recognition frameworks gather a lot of information from the framework to discover the connecting assaults.

System host and host based interruption discovery are able to identify interruption in the system. Be that as it may, expanded associations may give ascend in new regions and holes for interloper to assault. Not just had this heterogeneous based system IDS given discouraged execution yet in addition is troublesome for overseeing and observing vast systems. Along these lines, dispersed IDS having brought together or decentralized methodology for investigating traffic is required. In this methodology various hosts based and organize based IDSs are utilized which gather and dissect traffic on individual dimension as well as give contribution to brought together/decentralized analyzer. In this class there are two sorts of circulated frameworks. One sort contains incorporated methodology and another has decentralized [10]. DIDS (disseminated interruption identification framework) and NSTAT (Network State Transition Analysis Tool) have brought together methodology though GrIDS (Graph Based Intrusion Detection System) and EMERALD (Event Monitoring Enabling Responses to Anomalous Live Disturbances) utilize decentralized methodology.

DIDS [11] consolidates appropriated observing and information decrease with bring together framework. As it were various host-based IDS are associated through LAN having concentrated information examination as appeared in Figure 2. Appropriated screens/have gather framework data on which they are running and afterward they convert information into homogeneous arrangement, than for handling send this information to focal analyzer. This framework has ability to screen heterogeneous system PCs and it evacuates the lack of current IDS execution because of heterogeneous frameworks in the system.

NSTAT [12] is the conveyed variant of STAT, where STAT depends on state progress system. NSTAT depends on customer server engineering. Customer has two capacities: one it gathers and channels review trail information or framework data intermittently and afterward it sends this data to server for further preparing. Server side consolidations all data into single sequential structure and performs further handling. All customers performs such task like perusing logs, sifting, convert data into NSTAT group and utilized scrambled session to scatter data over the system. Where a server gets these information streams and unions it into single stream. At that point procedure this stream and perform rule coordinating, if interruption happens an activity is produced.

GrIDS [13] are the case of decentralized methodology of dispersed framework. Chart comprises of hubs and edges, speaking to areas and system traffic individually. In this youngster parent situation is executed. Kid space sends its information to its folks where information is handled by arrangements which are indicated by system overseer.

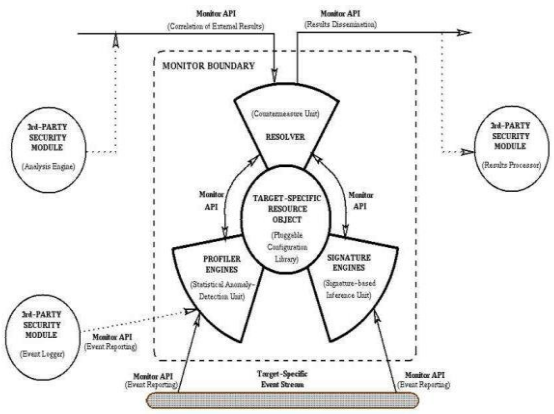


Figure 2: Generic EMERALD Monitor Architecture

EMERALD [14] display is the chain of importance/decentralized model of disseminated framework. It contains layers and each layer comprises of screens and each screen may have its very own abuse and inconsistency indicator. Figure 4 demonstrates the layer engineering of EMERALD. Layers are named as: administration (most minimal), area wide, and venture wide (most astounding). Administration layer screens single area. Space wide acknowledges contributions of administration layer and identify interruption over numerous areas, also undertaking wide acknowledges information from various area wide and attempt to distinguish interruption all through the whole framework. The next section describes the use of machine learning for identification of attacks using agents.

3. Proposed approach

The proposed approach can be demonstrated with the help of the following agents,

- Data collection
- Detection of attacks
- Responding to attacks
- Strategy evaluation

The input dataset file is given to the data collection agent, this agent processes the file and removes any unwanted entries. These entries are removed based on the port number, the protocol used by the user to access the network and the number of packets sent by the user's device. Packets with very low port numbers are usually discarded, and are not processed, as they are system generated packets to access the network status, while the packets which have very small number of packets sent or very large number of packets sent are also not processed, as they are packets for pinging the network, or packets with large file sizes. While packets have untrusted protocols are also removed by the collection agent.
Once the packets are processed by the collection agent, then these packets are given to the detection agent. The detection agent applies fuzzy C Means on the selected packets, and then evaluates if the partitioned packets are properly clustered. For cluster checking the system uses inter and intra cluster similarity metrics. If the cluster similarity is more than a pre-defined machine learning factor, then the packets are partitioned incorrectly, and re-clustering is done till the packets are properly partitioned. The final partitioned packets are selected as the packets via which attacks might be incoming to the system.
These clustered packets are then given to a response agent, which checks the packets for their similarity values, and similar signature packets are the ones which have similar values of protocols and port numbers. These packets are grouped to each other, and the IPs from these packets are blocked. These blocked IPs are then given to a strategy evaluation block, this block checks the evaluated packets, and compares them with a test dataset, if the packets are correctly identified as attackers, then they are removed, else the machine learning threshold is changed, and the detection agent is re-activated. Thus the algorithm keeps on changing the machine learning threshold in order to change the attack detection strategy, and thereby resulting in higher accuracy of attack detection in the network.

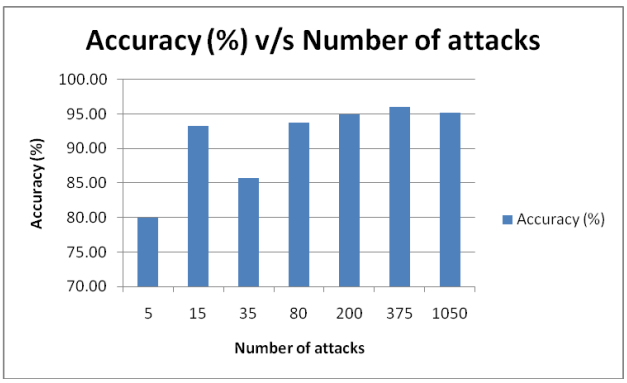
4. Results and analysis

We tested the system on Java based system with multiple configurations for the packets, the packets were taken from UCI repository and internal Windows log files. The following averaged results were obtained for the networks with different patterns,

Table 1. Comparative results

Number of entries	Number of attacks	Attacks detected	Accuracy (%)
100	5	4	80.00
300	15	14	93.33
500	35	30	85.71
1000	80	75	93.75
1500	200	190	95.00
5000	375	360	96.00
10000	1050	1000	95.24

From the table, we can observe that the accuracy for the machine learning system is very high, and increases as the number of attacks increases, thus the system adapts itself to the number of attacks, and changes the machine learning threshold so that the number of attacks are properly detected by the system. The following graph proves this point,



The simulation was done in Java environment, but is equally flexible for any other environment as well.

5. Conclusion

From the results, we can observe and conclude that the developed sytem is able to remove attacks and secure the network. Moreover due to this the overall network QoS parameters will also be improved as the attacks are being removed, which further allows for improvement in energy of the network via some parallel processing techniques to reduce the complexities of the machine learning algorithm. Thus, the system can be used for real time wireless networks in order to secure them and improve the security and reliability of the networks as well.

6. Future work

As machine learning and AI are gathering a lot of momentum, and blockchain technologies are coming up very rapidly, researchers can combine machine learning and AI with blockchain in order to improve the overall trust level of the system with the help of a fully decentralized protocol which reduces dependency on a central node, and improves the quality of the network via a better IDS system.

7. References

[1] Ozdemir, S. 2007. Secure and strong data aggregate for remote sensor frameworks. Proc. of fourth gathering on Ubiquitous handling system, Japan .
[2] Grossglauser, M., Tse, D. 2001. Transportability grows the breaking point of offhand remote frameworks, Proc. of IEEE INFOCOM 2001, pp.1360,1369.
[3] Zongheng, Z., Das, S., Gupta, H. 2004. Related K-consideration issue in sensor frameworks. Proc. of. thirteenth International Conference on Computer Communications and Networks, pp.373,378.
[4] Howard, A., Mataric, M., and Sukhatme, G. 2002. An enduring self-sending count for flexible sensor frameworks, Special Issue on Intelligent Embedded Systems Autonomous Robots, vol. 13(2), pp. 113– 126.
[5] Cardei, M., Jie W., Mingming L., Pervaiz, M. 2005. Most extraordinary framework lifetime in remote sensor frameworks with adaptable identifying ranges. Proc. of IEEE Conference on Wireless And Mobile Computing, Networking And Communications,pp.438-445.
[6] Cardei, M., Thai, M., Yingshu, L., Weili, W. 2005. Imperativeness capable target consideration in remote sensor frameworks. Proc. of INFOCOM, pp.1976,1984.
[7] Ebert, J., Willig, A. 1999. Gilbert-Elliot Bit Error Model and the Efficient Use in Packet Level Simulation, Technical Report TKN 99-2002, Germany.
[8] Wang, G., Cao, G., La Porta, T. 2006. Improvement helped sensor sending, IEEE Transactions on Mobile Computing, vol.5, pp.640-652.
[9] Taghikhaki, Z., Meratnia, N., Havinga, P. J M. 2011. Essentialness beneficial Trust-based combination in remote sensor frameworks, Proc. of IEEE INFOCOM. pp.584-589.

[10] Howard, A., Mataric, M., and Sukhatme, G. 2002. Compact sensor sort out association using potential fields: A dispersed, adaptable response for the domain consideration issue. Proc. of 6th Symposium on Distributed Autonomous Robotics Systems, pp. 299-308.
[11] Poduri, S., Sukhatme,G. 2004.Constrained consideration for convenient sensor frameworks. Proc. of Robotics and Automation Conferences. pp.165-171.
[12] Dhillon, S.S., Chakrabarty, K. 2003. Sensor position for reasonable incorporation and perception in passed on sensor frameworks, Wireless Communications and Networking, pp.1609-1614.
[13] O'Rourke, J. 1994. Computational Geometry in Column, Cambridge University Press.
[14] Cardei, M., MacCallum, D., Cheng, M. X., Min, M., Jia, X., Li, D. moreover, Du, D.- Z. 2002. Remote Sensor Networks with Energy Efficient Organization. Journal of Interconnection Networks, vol.3, pp. 213-229.

Shaziya Shaheen Wakeel Parvez (B.E) is pursuing M.Tech. degree in Computer Science and Engineering from Nuva college of Engineering,RTM Nagpur University,Nagpur,India.