ISSN No:-2456-2165

Hybridized Design For Feature Optimization and Reduction of Intrusion Detection Systems Alert in a Correlation Framework

Macarthy Osuo-Genseleke
Computer Science
Ignatious Ajuru University of Education
Rivers State, Nigeria

Ojekudo Nathaniel
Computer Science
Ignatious Ajuru University of Education
Rivers State, Nigeria

Abstract:- The Intrusion Detection System (IDS) produces a large number of alerts. Many large organizations deploy numerous IDSs in their network, generating an even larger quantity of these alerts, where some are real or true alerts and several others are false positives. These alerts cause very severe complications for IDS and create difficulty for the security administrators to ascertain effective attacks and to carry out curative measures. The categorization of such alerts established on their level of attack is necessary to ascertain the most severe alerts and to minimize the time required for response. An improved hybridized model was developed to assess and reduce IDS alerts using the combination of the Genetic Algorithm (GA) and Support Vector Machine (SVM) Algorithm in a correlation framework. The model is subsequently referred to as GA-SVM Alert Correlation (GASAC) model in this study. Our model was established employing the object-oriented analysis and design software methodology and implemented with Java programming language. This study will be benefitted by cooperating with networked organizations since only real alerts will be generated in a way that security procedures can be quickly implemented to protect the system from both interior and exterior attacks.

Keywords:- *Intrusion; Genetic Algorithm; Support Vector Machine; Feature selection; Optimization; Alert correlation; False alert; Real alert ; Alert Reduction.*

I. INTRODUCTION

Large computer networks are deployed by organizations within their systems. However, sophisticated network infrastructures are often accompanied by various threats. These threats include several malicious packages that affect networks efficiency as it relates to communication over the network or available network data. This prevailing problem urged academicians to improve and develop innovative techniques to discover and handle such threats [1]. This brought about cybersecurity. Cybersecurity which is well-known as internet security is applied to computer systems and networks telecommunications, entertainment, transportation, shopping, and medical organizations. Cybersecurity requires an effective defense of the system from danger and attacks [2]. Attacks are can be referred to as disturbances and offensives within a system and they threaten the safekeeping of the system by interfering with confidentiality, truthfulness, and ease of access [3]. Records of happenings on a system are logged into a file in a chronologically sorted sequence called the audit trail. Since virtually all activities are recorded, a manual assessment of these records would certainly allow intrusions to be discovered. However, the extraordinarily huge amount of audit records generated each day makes manual analysis difficult [4].

II. INTRUSION

Intrusions are usually sets of activities that try to compromise the truthfulness, privacy, or accessibility of information. This detection system (IDS) observes network movement or system records for distrustful activity and raises alerts to inform the administrator of danger [4]. This is a standard, recognized amongst the resources applied in the field of information security. Infiltration and attacks are regularly made by intruders to access their targets, knockout, or control their target either by distributing viruses, worms, or bot [5]: [1].

III. INTRUSION DETECTION SYSTEM

Intrusion Detection System (IDS) plays an important key role in information security in that, it detects intrusions by using the audit data creäted by systems. They are the valid options for establishments after applying firewall expertise in the network area [6]. They access suspected interference as soon as it happens and signals are raised. Through this means, unfriendly occurrences are perceived and reacted to, in a sensible amount of time. Devoid of IDS, an institution possibly will be confronted time after time and gain access without anyone knowing. The intrusion detection system observes packets to spot out their unpleasant activities. It produces alert once an unpleasant event is spotted and alerts the security administrator to carryout quick remedial actions against any such danger. These alerts contain a high volume of false alerts, that necessitates careful evaluation to identify and reduce the ineffective ones [7];[8]. Presently, intrusion detection systems have hitches that limit their configuration, scalability, and efficiency. Importantly, alerts generated by IDS are usually categorized into false and true [9]. As cited in [10], major problems with IDS are as follows:

- Alert overflowing: They make an available large quantity of alerts to the security analyst, who battles with the technical hitches, managing the stress
- Contextual problem: Attacks produce numerous correlated alerts. Security officers cannot with ease logically group correlated alerts from IDS
- False Alert: Present IDS generates alerts from legal activities.

IV. GENETIC ALGORITHM

It is a search technique, usually adopted to discover a suitable clarification to search difficulties. Its technicalities, natural choice, and neutral genetics are used to discover the best or near-best-case scenario solutions in an equitable quantity of time. Results achieved by a genetic algorithm is practically never best but near to optimality [11]. GA initiates problem resolution by attaining the resolution in changing fashion. GA is centered on the arrangement and improvement of cells found in living organisms. These cells comprise of chromosomes, made up of genes. It encompasses information about specific attributes of man, like eye color, etc. Recombination which is referred to as crossover happens during replication when genes of both parents form a fresh chromosome. Then freshly produced chromosome transforms as a significance of DNR. Its features might change due to inappropriate copy of the parent's genes; this aspect is the evolvement of the chromosome.

V. SUPPORT VECTOR MACHINE

It is a supervised machine-learning tool best suited for binary data grouping. SVM initially is a type of pattern classifier for categorization and regression with different kernel functions and is effectively used for pattern recognition applications. SVM is progressively useful to the improvement of intrusion detection systems for information safekeeping. One of the foremost benefits of using SVM for IDS is its speed and its competency of discovering intrusions instantaneously. SVM can learn a huge set of arrangements and be capable to scale better, for the reason that the classification complications do not rest on the dimension of the attribute space. SVMs brings up to date the training arrangements dynamically whenever there is a new arrangement during classification [12]. SVM is very good in alert reduction because it has demonstrated its robustness for pattern classification. It can be used to ascertain if two alerts made available can correlate or not and can also handle risk minimization implementations [13]. Another encouraging characteristic of SVM is, for discovering a universal minimum of definite risk by a way of organizational threat minimization, since it can specify well with kernel actions even in high-dimensional spaces under slight training illustration circumstances [12].

VI. ALERT CORRELATION

Alert Correlation is an alert controlling scheme that finds a relationship amongst alerts and endeavor to find attack patterns. It examines alert production processes and yields concise information on the security position of the network under observation. Granted that, variety of correlation schemes are recommended, there is no defined form for the implementation of its processes and evaluation. Present-day, correlation methods function on a less number of phases of the relationship course such as fusion and multistep identification of threats which signifies a categorization of actions executed by same intruder. As an effect, it not certain if and how diverse parts of the procedure adds to the over-all objective of the correlation process [14]. Correlation approaches handle the issues of an enormous size of alerts [15]. It helps the expert party to investigate and differentiate between alerts produced.

VII. ALERT CORRELATION PHASES

- *Resemblance of alert features technique*

This is responsible for associating alerts to all alert threads with related features [16], for example, source and terminus IP addresses and then associates alarms with a high grade of attribute resemblance, if matched or a novel thread is generated if no on matches [16].

- *Predefined invasion patterns*

This method exploits the fact that intrusions regularly need for diverse actions to be successful. Every attack pattern has consistent steps necessary for effectiveness [10]. Alerts are associated with pre-defined attack patterns for alarms to be correlated. It is narrowed to identified threats only and signature discoveries only and detailed by human operators.

- *Conditions and penalties of outbreaks*

This method works at an advanced stage than correlation established on characteristics resemblances, but at a lesser stage than correlation established on well-known patterns [10]. Pre-conditions are well-defined here as the essential circumstances that should be in existence for a threat to be positive, and penalties for the threat are well-defined as conditions that ought to be present after a definite attack has transpired [10]. This modus operando is not restricted to identified attack patterns and it reveals the causal association between alerts.

- *Statistical causal analysis*

This method affirms that every single threat produces a signal with numerical resemblances in their features, and threat stages also have fundamental affiliations [17]. The realistic solution for a whole correlation process is not certain here. Nonetheless, it can be employed as a fragment of the system to clean up alarms or to make available meta-alert signatures.

VIII. RELATED WORK

[18], proposed the combination of Particle Swarm Optimization and Support Vector Machine techniques for the reduction of attack features. Their aim was achieved as their system performed optimally by reducing the 41 number of attack features NSL-KDD dataset is having to 20.

[19], proposed an IDS quality framework using alert post-processing techniques to separate false alerts from true alerts. An evaluation was conducted using DARPA 2000 dataset and produced a pleasant result than many of the present techniques with a 95% alert reduction rate, 85% of completeness.

[20], recommended a rough set theory to handle redundancy alert. DARPA 98 dataset was used to evaluate their model and experiment showed that their technique efficiently reduced redundant alerts for the system administrator to shun wasting time on valueless alerts.

[21], aimed to advance network safety by curtailing the amount of dishonest positive alarms. Two major phases were considered in their study. Phase one removed repeated alerts that were achieved using a filtering algorithm. Phase two reduced dishonest alerts by eradicating the jobless alerts and this was accomplished by using association procedures and removal of recurrent item-set algorithms. Their model was evaluated and appraised on a DARPA 99 dataset. The result shows that their technique drastically reduced the volume of IDS False Positive alerts by 97.98%.

[22], proposed the fusion of data removal and alert association scheme. This work focused on the decline of untruthful positive alerts at the attack generated stage and alert handling phase. Experimental outcome reveals that these techniques were able to overwhelm the IDSs difficulties by reducing untruthful positive alerts by 70%.

[23], suggested a scheme that finds a relationship amongst attack considering their entropy. Their target is to discover underlying dealings amongst the intrusive behaviors signals, given a high opinion of attack pattern demonstration. Firstly, they decreased the number of alerts by a way of data aggregation. The researchers achieved a level of 99.98% of data decline by applying the DARPA 2000 data set to appraise the suggested system.

[24], employed data mining and neural distribution techniques for grouping and feature optimization correspondingly in the decline of alerts. The reduced feature enhanced the classification pattern of intrusion discovery scheme and decreased the untruthful and negative detection of alerts thereby, improving the efficiency of the system but, a dataset with few alert types was used to evaluate their model.

IX. METHODOLOGY

This study offers an improved model for multiple intrusion detection systems combining Support Vector Machine (SVM) and Genetic Algorithm (GA) in an alert correlation framework. The model combines numerous IDS alerts established on the investigations of the multifaceted connections amongst the alert information from numerous IDSs. SVM is relatively a unique classification procedure that has displayed higher performance than traditional learning procedures particularly in multidimensional classification, saddled with the obligation of reducing the influence of false alerts. Relating to the improvement of the discrimination accuracy of SVM in prediction, a Genetic Algorithm (GA) finds the best parameters for SVM in the solution space. The GA-SVM method used has high accuracy and low false-positive rates.

X. SYSTEM DESIGN

The design phase of the GA-SVM correlation model combines the Genetic Algorithm (GA) optimizer with Support Vector Machine (SVM) classifier for improved alert assessment and reduction. This system handles the difficulties of alert feature optimization and reduction by combining the novel machine-learning algorithm (Support vector machine (SVM) algorithm) and evolutionary computing algorithm (Genetic algorithm (GA)) in an alert correlation process. SVM is adopted for handling classification jobs, whereas GA is to handle the optimization of alert features. It has been established that SVM is a unique algorithm with good functions for classification. Nevertheless, its performance hinges on the accurate choice of the functions for the SVM. The proposed scheme offers a GA–SVM technique, with parameter optimization for alert correlation to assess and reduce IDS alerts. Our planned system is talented to achieve optimal detection, that will help in minimizing the overheads and maximizing operations of the SVM classifier in the anticipated system. The figure below shows the hybridized system architecture.

The system comprises three main phases: the pre-processing phase, the feature selection, and the optimization phase and finally, Post-processing and reduction phase.

- Pre-processing: This phase begins the starting process of the model. Here, alerts features are unified and merged from multiple IDSs, symbolic features are converted to numeric values and in an IDMEF format using normalization technique. It improves the accuracy and efficacy of alert features to be selected and optimized by genetic algorithms.

- Feature selection and optimization phase: This phase is attained using Genetic Algorithm through the methods of selection, prioritization, verification, and fusion techniques. GA performs a selection of attack features using a fitness function. Fitness function is a mathematical formula used to set criteria or rules for selection. The prioritization component assigns appropriate priority to alert features while the verification component, takes a single feature and determine the achievement of the attack that matches it. Lastly, the fusion technique combines all features from different IDS that represent independent detection by same attack feature for optimized feature sets to be achieved.

- Post-processing and reduction phase: Here, optimized feature sets are sent to SVM classifier for classification and reduction of alerts through the processes of focus recognition, uncorrelated removal, and multistep correlation. Finally, an intrusion report is generated. Focus recognition brings together optimized parameters and classifies features into different alert types. Thereafter, uncorrelated removal separates and discards all false alert types while the multistep correlation, crosschecks and validates the process carried out by uncorrelated removal for policy compliance and only real alert type to be established as intrusion report.

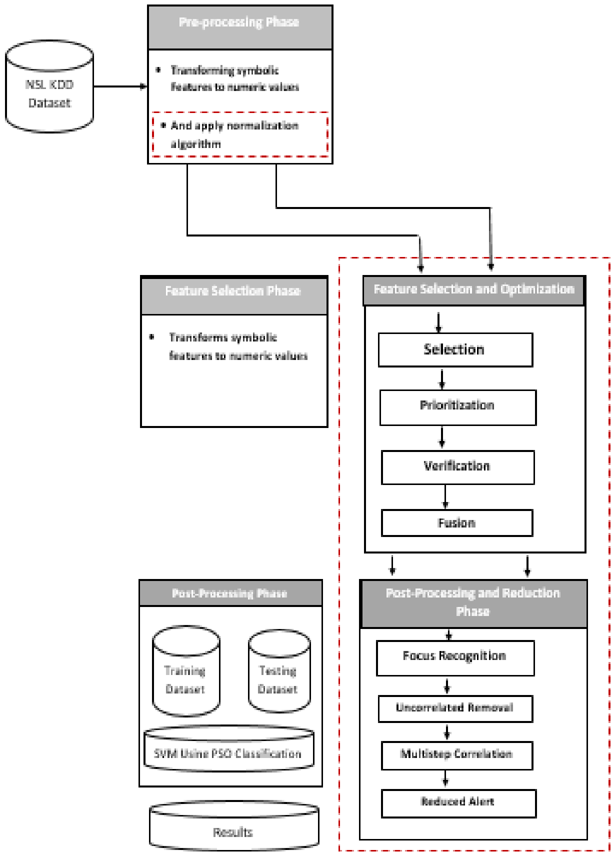

Fig 1:- GA-SVM Alert correlation system architecture

XI. RECOMMENDATION

The system was designed to accomplish comprehensive feature optimization and reduction of IDS alerts from multiple IDSs and reduce the alert for more concentrated intervention by the network admin but the computational expensive assessment of the genetic kernel questions the applicability of GA-SVM in a large-scale application. We further recommend:

- A better GA-SVM parallelization could help but would increase the complexity of the overall infrastructure.
- For a more efficient calculation and caching of genetic kernels as well as the SVM classifier, new techniques should be considered in further research.

XII. CONCLUSIONS

The study focused on IDS alert assessment and reduction. It gave specific consideration to the correlation of alerts from multiple IDS in cooperating networked organizations. The system used a combination method that makes available evolutionary optimized parameters for SVM classification of alerts using GA. The System is an enhancement from the works of [18].

This work is an improved hybridized design for the optimization of alert features and categorize the degree of attacks generated by multiple Intrusion Detection System and reduce false alerts. Using Genetic Algorithm (GA) and Support Vector Machine (SVM) in alert correlation framework is to our best knowledge the first attempt to apply GA-SVM to alert correlation for multiple IDS alert assessment, feature optimization and alert reduction.

REFERENCES

[1]. Al-Saedi, K. H., H. Al-Khafaji, A. Almomani, S. Manickam, and S. Ramadass (2011). An approach to assessment of network worm detection using threatening-database mining. Journal of Applied Science, 12(5), 2676-2683.

[2]. Tjhai, G. C., S. M. Furnell, M. Papadaki, and N. L. Clarke. (2010). A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm. Journal of Computer Security. 29(6), 712-723.

[3]. Bace, R., and P. Mell (2001). Intrusion Detection Systems. National Institute of Standards and Technology (NIST) - Technical Report. New York.

[4]. Manish, K., and M. Hanumanthappa (2015). Distributed Intrusion Detection System Using IDMEF. International Journal of Advanced Trends in Computer Applications (IJATCA), 12(5),1-5.

[5]. Elshoush, H. T., and I. M. Osman. (2012). Alert correlation in collaborative intelligent intrusion detection systems - A survey. Applied Soft Computing Journal. 11(3), 4349-4365.

[6]. Mon, M. Z. and P. Thandar (2014). Correlation Model of IDS Attack Alert on Network Attack Alerting System. International Journal of scientific Engineering and Technology Research. 3(13), 2950-2954.

[7]. Porres, I., and M. D. M. Fernandez. (2008). An Evaluation of current Intrusion Detection System. First Asian Conference on Intelligent Information and Database Systems, IEEE Computer Society. 10(4), 465- 470.

[8]. Njogu, H., and L. Jiawei, (2010). Using alert cluster to reduce IDS alerts. Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, New York. 9-11.

[9]. Karim, A., S. Manickam, S. Ramadass, W. Al-Salihy, and A. Almomani, (2013). Research Proposal: An Intrusion Detection System Alert Reduction and Assessment Framework Based On Data Mining. Journal of Computer Science. 9 (4), 421-426.

[10]. Robiah, Y., R. S. Siti, and S. Shahrin (2008). Intrusion Alert Correlation Technique Analysis for Heterogeneous Log. International Journal of Computer Science and Network Security(IJCSNS), 8 (9), 132-138.

[11]. Goldberg, D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. IEEE Security and Privacy Magazine. 3(2), 23-28.

[12]. Shon, T., Y. Kim, C. Lee, and J. Moon. (2005). A Machine Learning Framework for Network Anomaly Detection using SVM and GA. IEEE Transaction on Dependable and Secure Computing. 4(3), 234-240.

[13]. Ankita B. P., and S.S. Dhande. (2017). Study of Alert Correlation Technique. International Journal of Advanced Research in Computer and Communication Engineering.6(3), 640-644.

[14]. Vigna, G., and R. Kemmerer. (1999). NetSTAT: A Network-Based Intrusion Detection System. Journal of Computer Security. 7(1), 37-71.

[15]. Omar, A., E. Homam, M. Ahmed, and R. Sureswaran, (2009). False positive reduction in Intrusion Detection System: A survey. Proceedings of 2nd IEEE International Conference on Broadband Network and Multimedia Technology. 463-466.

[16]. Valdes, A., and K. Skinner. (2001). Probabilistic Alert Correlation. International Workshop on Recent Advances in Intrusion Detection (RAID), 54-68.

[17]. Qin, X., and W. Le. (2003). Statistical Causality of INFOSEC Alert Data. Proceedings of 6th International Symposium on Recent Advances in Intrusion Detection. 73-93

[18]. Chakir, E., M. Mohamed, and Y. I. Khamlichi (2018). An Effective Intrusion Detection Model Based On Svm with Feature Selection and Parameter Optimization. Journal of Theoretical and Applied Information Technology. 96 (12). 3873-3885.

[19]. Ali, M. R., M. S. I. Ahmed, and H. H. Almistarihi, (2019). A Quality Framework to Improve IDS Performance Through Alert Post Processing. International journal of intelligent Engineering and systems, 12(5), 149-160.

[20]. Ru, Z., T. Guo and J. Liu, (2018). An IDS Alerts Aggregation Algorithm Based on Rough Set Theory. IOP Conf. Series: Materials Science and Engineering 322-335.

[21]. Isam, K. T., N.U. Osman, O. Bayat, and K. H. Al-saedi (2017). Improving IDSs alerts to improve high quality network security by using data mining technique. Aurum Journal of Engineering systems and architecture. 1(2),17-29.

[22]. Anthony, R. A and S. Siddarama. (2016). The False Positive Alert Reduction Using Datamining Techniques in Intrusion Detection System. International Research Journal of Computer Science (IRJCS), 3(6), 16-21.

[23]. GhasemiGol M. and A. Ghaemi-Bafghi (2015). E-correlator: an entropy-based alert correlation system. Security and Communication Networks. 8(5):822–836.

[24]. Jaykaran, H. (2014). A Review of Intrusion Detection Technique Based On Classification and Feature Optimization. International Journal for Scientific Research & Development. 1(12), 2805-2809.