WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Thread based Machine Learning Framework for Cyber Security Operations Center

M. Sathana [#1], M.Hemamalini [#2]

[#1] M.Sc, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai-606603
[#2] Assistant professor, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai-606603

ABSTRACT

In order to ensure a company's Internet security, SIEM (Security Information and Event Management) system is in place to simplify the various preventive technologies and flag alerts for security events. Inspectors (SOC) investigate warnings to determine if this is true or not. However, the number of warnings in general is wrong with the majority and is more than the ability of SCO to handle all awareness. Because of this, malicious possibility. Attacks and compromised hosts may be wrong. Machine learning is a possible approach to improving the wrong positive rate and improving the productivity of SOC analysts. In this article, we create a user-centric engineer learning framework for the Internet Safety Functional Center in the real organizational context. We discuss regular data sources in SOC, their work flow, and how to process this data and create an effective machine learning system. This article is aimed at two groups of readers. The first group is intelligent researchers who have no knowledge of data scientists or computer safety fields but who engineer should develop machine learning systems for machine safety. The second groups of visitors are Internet security practitioners that have deep knowledge and expertise in Cyber Security, but do Machine learning experiences do not exist and I'd like to create one by themselves. At the end of the paper, we use the account as an example to demonstrate full steps from data collection, label creation, feature engineering, machine learning algorithm and sample performance evaluations using the computer built in the SOC production of Seyondike.

Keywords: SOC, SIEM, SCO

1.INTRODUCTION:

By and by frameworks associated by the web, for example, the equipment, programming and information can be shielded from cyberattacks utilizing cybersecurity. Cybersecurity is a lot of advancements and procedures intended to secure PCs, networks, projects, and information from assaults and unapproved access, change, or obliteration. As dangers become increasingly refined the latest advancements, for example, Machine learning (ML) and profound learning (DL) are utilized in the cybersecurity network to use security capacities. These days, cybersecurity is an invigorating issue in the internet and it has been relying upon computerization of various application spaces, for example, accounts, industry, clinical, and numerous other significant zones This paper manages past work in machine learning (ML) and profound learning (DL) techniques for cybersecurity applications and a few utilizations of every strategy in cybersecurity tasks are depicted. The ML and DL techniques shrouded in this paper are pertinent to distinguish cybersecurity dangers, for example, programmers and predators, spyware, phishing, and network interruption location in ML/DL. In this manner, incredible noticeable quality is set on an exhaustive portrayal of the ML/DL techniques, and references to original works for every ML and DL strategy are given What's more, examine the difficulties and chances of utilizing ML/DL for cybersecurity.

2.LITERATURE SURVEY:

[1]The input to an algorithm that learns a binary classifier normally consists of two sets of examples, where one set consists of positive examples of the concept to be learned, and the other set consists of negative examples. However, it is often the case that the available training data are an incomplete set of positive examples, and a set of unlabeled examples, some of which are positive and some of which are negative. The problem solved in this paper is how to learn a standard binary classifier given a nontraditional training set of this nature. Under the assumption that the

labeled examples are selected randomly from the positive examples, we show that a classifier trained on positive and unlabeled examples predicts probabilities that differ by only a constant factor from the true conditional probabilities of being positive. We show how to use this result in two different ways to learn a classifier from a nontraditional training set. We then apply these two new methods to solve a real-world problem: identifying protein records that should be included in an incomplete specialized molecular biology database.

[2]This survey paper describes a focused literature survey of machine learning (ML) and data mining (DM) methods for cyber analytics in support of intrusion detection. Short tutorial descriptions of each ML/DM method are provided. Based on the number of citations or the relevance of an emerging method, papers representing each method were identified, read, and summarized. Because data are so important in ML/DM approaches, some well-known cyber data sets used in ML/DM are described. The complexity of ML/DM algorithms is addressed, discussion of challenges for using ML/DM for cyber security is presented, and some recommendations on when to use a given method are provided.

[3] In our proposed paper, several classification techniques and machine learning algorithms have been considered to categorize the network traffic. Out of the classification techniques, we have found nine suitable classifiers like BayesNet, Logistic, IBK, J48, PART, JRip, Random Tree, Random Forest and REPTree. Out of the several machine learning algorithms, we have worked on Boosting, Bagging and Blending (Stacking) and compared their accuracies as well. The comparison of these algorithms has been performed using WEKA tool and listed below according to certain performance metrics. Simulation of these classification models has been performed using 10-fold cross validation. NSL-KDD based data set has been used for this simulation in WEKA.

[4] Network attacks have become more pervasive in the cyber world. There are various attacks such as denial of service, scanning, privilege escalation that is increasing day by day leading towards the requirement of a more robust and adaptable security techniques. Anomaly detection is the main focus of our paper. Support Vector Machine (SVM) is one of the good classification algorithm applied specially for intrusion detection. However, its performance can be significantly improved when it is applied in integration with other classifiers. In this paper, we have performed a comparative analysis of SVM classifier's performance when it is stacked with other classifiers like BayesNet, AdaBoost, Logistic, IBK, J48, RandomForest, JRip, OneR and SimpleCart.

[5] (DNN). In the proposed technique, in-vehicle network packets exchanged between electronic In this paper, we propose a novel intrusion detection technique using a deep neural network control units (ECU) are trained to extract low- dimensional features and used for discriminating normal and hacking packets. The features perform in high efficient and low complexity because they are generated directly from a bitstream over the network. The proposed technique monitors an exchanging packet in the vehicular network while the feature are trained off-line, and provides a real-time response to the attack with a significantly high detection ratio in our experiments.

3. PROPOSED SYSTEM:

User-centric cyber security helps enterprises reduce the risk associated with fast-evolving end-user realities by reinforcing security closer to end users. User-centric cyber security is not the same as user security. User-centric cyber security is about answering peoples' needs in ways that preserve the integrity of the enterprise network and its assets. User security can almost seem like a matter of protecting the network from the user — securing it against vulnerabilities that user needs introduce. User-centric security has the greater value for enterprises. cyber-security systems are real-time and robust independent systems with high performances requirements. They are used in many application domains, including critical infrastructures, such as the national power grid, transportation, medical, and defense. These applications require the attainment of stability, performance, reliability, efficiency, and robustness, which require tight integration of computing, communication, and control technological systems. Critical infrastructures have always been the target of criminals and are affected by security threats because of their complexity and cyber-security connectivity. These CPSs face security breaches when people, processes, technology, or other components are being attacked or risk management systems are missing, inadequate, or fail in any way. The attackers target confidential data. Main scope of this project in reduce the unwanted data for the dataset.

4. METHODOLOGY:

3.1 Cyber threatanalysis is a process in which the knowledge of internal and external information vulnerabilities pertinent to a particular organization is matched against real-world cyber-attacks. With respect to cyber security, this threat-oriented approach to combating cyber-attacks represents a smooth transition from a state of reactive security to a state of proactive one. Moreover, the desired result of a threat assessment is to give best practices on how to maximize the protective instruments with respect to availability, confidentiality and integrity, without turning back

to usability and functionality conditions. CYPER ANALYSIS.A threat could be anything that leads to interruption, meddling or destruction of any valuable service or item existing in the firm's repertoire. Whether of "human" or "nonhuman" origin, the analysis must scrutinize each element that may bring about conceivable security risk.

3.2 If a dataset in your dashboard contains many dataset objects, you can hide specific dataset objects from display in the Datasets panel. For example, if you decide to import a large amount of data from a file, but do not remove every unwanted data column before importing the data into Web, you can hide the unwanted attributes and metrics, To hide dataset objects in the Datasets panel, To show hidden objects in the Datasets panel, To rename a dataset object, To create a metric based on an attribute, To create an attribute based on a metric, To define the geo role for an attribute, To create an attribute with additional time information, To replace a dataset object in the dashboard.

5. SVM ARCHITECTURE;

Support Vector Machine (SVM) This is a supervised learning method used in machine learning for classifying objects. Linear classifiers determine if an object is that object or is not that object by finding a hyperplane, or demarcating line, that clearly segments the objects from each other. By doing so, the algorithm can determine the classification of the object by determining on which side of the hyperplane the object falls. By altering the kernel function, it is possible to find a hyperplane to determine non-linear classifications by creating hyperplane lines that appear to weave through the data set. This determination is based upon Gaussian radial basis and tangents. However this

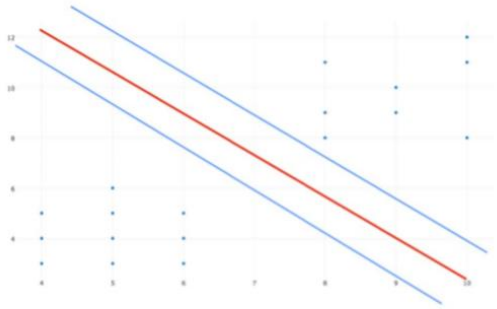

FIGURE 1 AN EXAMPLE OF A LINEAR SVM WITH HYPERPLANE

For the proposed model, the SVM makes a binary classification of the data as either an attack or normal traffic. In order to train this model, a new attribute to the dataset has been added indicating whether that instance is an attack or normal. Weka does not allow for multiple labels on a dataset, thus filtered classifiers were utilized in order to ignore the additional labels during training and testing to prevent the model from learning on one of the labels.

6. CONCLUSION

We provide a user-centered computer learning system that affects large data from various security logs, awareness information, and inspector intelligence. This method provides complete configuration and solution for dangerous user detection for the Enterprise System Operating Center. Select machine learning methods in the SOC product environment, evaluate efficiency, IO, host and users to create user-centric features. . Even with simple mechanical learning algorithms, we prove that the learning system can understand more insights from the rankings with the most unbalanced and limited labels. More than 20% of the neurological model of modeling is 5 times that of the current rule-based system. To improve the detection precision situation, we will examine other learning methods to improve the data acquisition, daily model renewal, real time estimate, fully enhance and organizational risk detection and management. As for future work, let's examine other learning methods to improve detection accuracy.

7. OUTPUT

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing

are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

Select methods for presenting information.Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.Convey information about past activities, current status or projections of theFuture.Signal important events, opportunities, problems, or warnings.Trigger an action.Confirm an action.

RESULTS:

FIG 1: Thread Prediction Login

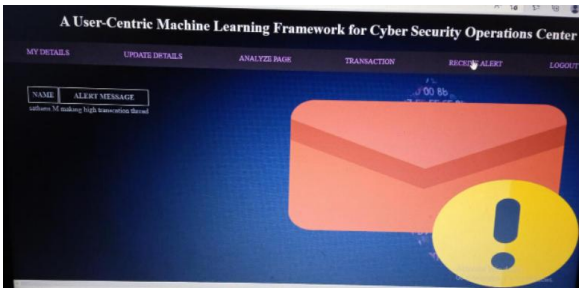

FIG 2: Thread Prediction data

FIG 3: Crime Performance

FIG 4: Thread User Validation

REFERENCES

[1] Corporate.target.com, "2014 Annual Report | Target Corporate", 2016. [Online]. Available: https://corporate.target.com/annual-reports/2014. [Accessed: 11- Oct- 2015].

[2] J. McHugh, "Testing Intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory", ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262-294, 2000.

[3] A. J. M. Abu Afza, M. S. Uddin, "Intrusion detection learning algorithm through network mining." In Computer and Information Technology (ICCIT), 2013 16th International Conference on, pp. 490-495, IEEE, Mar. 2014.

[4] J. P. Anderson, "Computer security threat monitoring and surveillance," Technical report, James P. Anderson Company, Fort Washington, Pennsylvania, vol. 17, 1980.

[5] D. Denning, S. Ak, M. Heckman, T. Lunt, M. Morgenstern, P. Neumann and R. Schell, "Views for Multilevel Database Security", IEEE Transactions on Software Engineering, vol. -13, no. 2, pp. 129-140, 1987.

[6] D. E. Denning, "An Intrusion-Detection Model," 1986 IEEE Symposium on Security and Privacy, 1986. [7] J. Quinlan, "Induction of decision trees", Mach Learn, vol. 1, no. 1, pp. 81-106, 1986.

[8] J. Bell, Machine learning: hands-on for developers and technical professionals. Indianapolis: John Wiley & Sons, 2015.

[9] W. Lee, S. J. Stolfo, and P. K. Chan, "Learning patterns from unix process execution traces for intrusion detection," AAAI Workshop on AI Approaches to Fraud Detection and Risk Management, pp. 50–56, Jul. 1997.

[10] T. Lane and C. E. Brodley, "Temporal sequence learning and data reduction for anomaly detection," ACM Transactions on Information and System Security TISSEC ACM Trans. Inf. Syst. Secur., vol. 2, no. 3, pp. 295–331, Jan. 1999.

[11] W. Lee and S. J. Stolfo, "Data mining approaches for intrusion detection," Usenix security, Jan. 1998.

[12] S. Forrest, et al., "A sense of self for unix processes," Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on. IEEE, pp. 120–128, 1996.

[13] S. Forrest, S. A. Hofmeyr, and A. Somayaji, "Computer immunology," Communications of the ACM, vol. 40, no. 10, pp. 88–96, Oct. 1997.

[14] C. Warrender, S. Forrest, and B. Pearlmutter, "Detecting intrusions using system calls: alternative data models," Proceedings of the 1999 IEEE Symposium on Security and Privacy (Cat. No.99CB36344), pp. 133–145, 1999.

[15] B. C. Rhodes, J. A. Mahaffey, and J. D. Cannady, "Multiple selforganizing maps for intrusion detection," Proceedings of the 23rd national information systems security conference , pp. 16–19, Oct. 2000.

[16] J. Cannady, "Artificial neural networks for misuse detection," National information systems security conference, pp. 368-81, 1998.

[17] K. Julisch, "Mining alarm clusters to improve alarm handling efficiency," Seventeenth Annual Computer Security Applications Conference IEEE, pp. 12–21, Dec. 2001.

[18] F. Cuppens and A. Miege, "Alert correlation in a cooperative intrusion detection framework," Proceedings 2002 IEEE Symposium on Security and Privacy, IEEE, pp. 202–215, 2002.

[19] Y.-A. Huang and W. Lee, "A cooperative intrusion detection system for ad hoc networks," Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks - SASN '03 ACM, pp. 135–147, Oct. 2003.

[20] Z.-S. Pan, et al., "Hybrid neural network and C4. 5 for misuse detection," Machine Learning and Cybernetics, 2003 International Conference on. IEEE, vol. 4, pp. 2463–3467, Nov. 2003.