Predicting The Cyber Security Models Using Machine Learning Algorithms

S V N SWETHA[1], K SURYA TEJA[2], G MEGHANA REDDY [3], N VINAY KUMAR[4]

UG Scholar'[2], Professor[4,]
Department Of Electronics And Computer Engineering[1,2,3,4]
Sreenidhi Institute Of Science And Technology[1,2,3,4],
Hyderabad, Telangana, India[1,2,3,4]

ABSTRACT:

In the ever growing battlefield of cyber security, it is nearly impossible to quantify the reasons why cyber security is important. Allowing malicious threats to run amok anywhere, at any time, and in any context is far from acceptable, and it especially applies to the intricate web of customer and company data that cyber security teams are striving to protect. Cybersecurity is a necessary consideration for individuals and families, as well as businesses, governments, and educational institutions. With help of Machine Learning Can Advance Cyber Security Landscape. Businesses today are gathering huge amounts of data. Data is at the heart of just about any business-critical system you can think of. This also includes infrastructure systems. Today's high-tech infrastructure, including network and cybersecurity systems are gathering tremendous amounts of data and analytics on most key aspects of mission-critical systems. While human beings still provide the key operational oversight and intelligent insights into today's infrastructure, machine learning and artificial intelligence are gaining huge momentum in most areas of today's systems, whether positioned on-premise or in the cloud.

Keywords: cybersecurity, random forest, confusion matrix, prediction

INTRODUCTION

There is heap of buzz around analytics in cyber security. Machine learning prophetic analytics provides a robust use case for network and cyber security applications. Organizations these days arE inundated with myriads of network connections and traffic flows, as well as cyber security events that require analysis and potentially, remediation. The sheer volume of traffic and events likewise because the quality of today's hybrid cloud networks makes it impractical to possess individuals making an attempt to analyse all the network and cyber security knowledge being collected and creating choices supported this knowledge. Machine learning within the realm of network and cyber security permits network and cyber security systems to try to provide some pretty superb things. Machine learning these days is ready for the foremost half accurately confirm and develop on anomalies in traffic patterns, connections, user activity, and many other aspects of network. Powerful machine learning algorithms are able to filter through traffic patterns and learn what the normal fingerprint of network activity looks like and then make decisions based on machine learning algorithms. Our Research aims to provide cyber security with help of two trending technologies ML. Their paper focuses on network intrusion detection. Through a wired network, attackers must pass multiple layers of firewall and operating system defences or gain physical access to the network. However, any node can be a target in a wireless network; thus, the network is more vulnerable to malicious attacks and is more difficult to defend than are wired networks.

PROBLEM STATEMENT

Cyber attacks has been increasing in the cyber world. There should be some advanced security measures has taken to reduce or avoid the number of cyber attacks . There are numerous attacks are DDos attacks, Man in the middle ,Data leakage, PROBE, User To Root ,Remote To Local . These attacks has been used by the hackers or intruder to gain the unauthorised access to that private network, websites , database or even in our personal computers. So the outside hacker or internal hacker are using advance techniques or methods to tickle or break any defence systems to protect the sensitive data ,or information ,financial information's .Smart intrusion defence system should be prevent or tries to manage various innovative attacks created or programmed by the hackers.

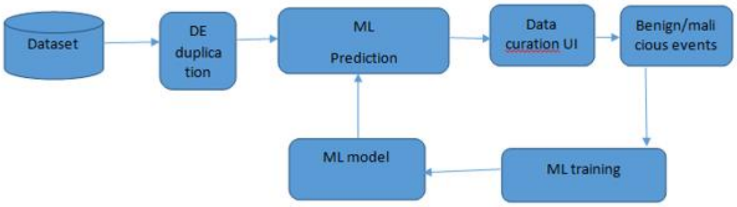
LITERATURE SURVEY

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. In recent years, Cloud computing has emerged as a new paradigm for delivering highly scalable and on-demand shared pool IT resources such as networks, servers, storage, applications and services through internet. It enables IT managers to provision services to users faster and in a cost-effective way. As a result, this technology is used by an increasing number of end users. On the other hand, existing security deficiencies and vulnerabilities of underlying technologies can leave an open door for intruders. Indeed, one of the major security issues in Cloud is to protect against distributed attacks and other malicious activities on the network that can affect confidentiality, availability and integrity of Cloud resources. In order to solve these problems, we propose a Collaborative Network Intrusion Detection System (C-NIDS) to detect network attacks in Cloud by monitoring network traffic, while offering high accuracy by addressing newer challenges, namely, intrusion detection in virtual network, monitoring high traffic, scalability and resistance capability. In our NIDS framework, we use Snort as a signature based detection to detect known attacks, while for detecting network anomaly; we use Support Vector Machine (SVM). Moreover, in this framework, the NIDS sensors deployed in Cloud operate in collaborative way to oppose the coordinated attacks against cloud infrastructure and knowledge base remains up-to-date.

PROPOSED METHOD

Today's cyber threats are too numerous and arrive too fast for purely manual defences. Machine learning provides power and speed to tackle huge volumes of attacks with countless variations. Yet the real key to leveraging AI for cyber protection is to use it with human intelligence, combining power, speed, skills and judgment. The machine learning algorithms security companies deploy generally train on large data sets to "learn" what to watch out for on networks and how to react to different situations. Machine learning is powerful in its own right, though, and approach is a natural fit for antivirus defence and malware scanning.

BLOCK DIAGRAM



EVOLUTION MODEL

Model analysis is AN integral a part of the model development method. It helps to search out the simplest model that represents our knowledge and the way well the chosen model can add the longer term. Evaluating model performance with information used for coaching isn't acceptable in data science as a result of it will simply generate overoptimistic and over fitted models. There ar 2 strategies of evaluating models in knowledge science, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of every classification model is calculable base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

ALGORITHM EXPLANATION

RANDOM FOREST

Random forest could be a style of supervised machine learning algorithmic rule supported ensemble learning. Ensemble learning could be a style of learning wherever you be part of differing kinds of algorithmic rules or same algorithm multiple times to make a additional powerful prediction model. The random forest algorithmic rule combines multiple algorithmic rule of constant kind i.e. multiple call trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithmic rule is used for each regression and classification tasks.

HOW RANDOM FOREST WORKS

The following area unit the fundamental steps concerned in activity the random forest algorithmic rule

1. Pick N random records from the dataset.

2. Build a decision tree based on these N records.

3. Choose the number of trees you want in your algorithm and repeat steps 1 and 2.

4. For classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is appointed to the class that wins the bulk vote.

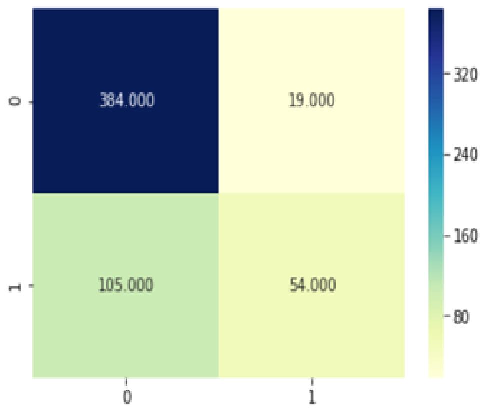
LIMITATION

Traditional Intrusion detection techniques are not efficient to tackle the hacker as well as theft sensitive data in enterprise may be in cloud or data centers.

- Hardware-based cyber security tools are not enough.
- It is impossible for any business to react quickly enough to evolving threats without relying on AI and machine learning cyber security technologies.

RESULTS

Confusion matrix of our test data



accuracy is 0.779

CONCLUSION

In this paper, we present a user-centric machine learning system which leverages big data of various security logs, alert information, and analyst insights to the identification of risky user. This system provides a complete framework and solution to risky user detection for enterprise security operation center. We describe briefly how to generate labels from SOC investigation notes, to correlate IP, host, and users to generate user-centric features, to select machine learning algorithms and evaluate performances, as well as how to such a machine learning system in SOC production environment. We also demonstrate that the learning system is able to learn more insights from the data with highly unbalanced and limited labels, even with simple machine learning algorithms. The average lift on top 20% predictions for multi neural network model is over 5 times better than current rule-based system. The whole machine learning system is implemented in production environment and fully automated from data acquisition, daily model refreshing, to real time scoring, which greatly improve and enhance enterprise risk detection and management. As to the future work, we will research other learning algorithms to further improve the detection accuracy.

REFERENCES

[1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, 2015.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 2015.

[3] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, 2015.

[4] E. Blanzieri and A. Bryl, "A survey of learning-based techniques of email spam filtering," Artificial Intelligence Review, 2008.

[5] J. Gardiner and S. Nagaraja, "On the Security of Machine Learning in Malware C8C Detection," ACM Computing Surveys, 2016.

[6] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar, "Adversarial Machine Learning," in ACM workshop on Security and artificial intelligence, 2011.

[7] F. Pierazzi, G. Apruzzese, M. Colajanni, A. Guido, and M. Marchetti, "Scalable architecture for online prioritization of cyber threats," in International Conference on Cyber Conflict (CyCon), 2017.

[8] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection," in IEEE International Conference on Platform Technology and Service (PlatCon), 2016.

[9] P. Torres, C. Catania, S. Garcia, and C. G. Garino, "An analysis of Recurrent Neural Networks for Botnet detection behavior," in IEEE Biennial Congress of Argentina (ARGENCON), 2016.

[10] G. E. Dahl, J. W. Stokes, L. Deng, and D. Yu, "Large-scale malware classification using random projections and neural networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[11] G. D. Hill and X. J. Bellekens, "Deep Learning Based Cryptographic Primitive Classification," arXiv preprint, 2017.

[12] R. Pascanu, J. W. Stokes, H. Sanossian, M. Marinescu, and A. Thomas, "Malware classification with recurrent networks," in IEEE International Conference on

Acoustics, Speech and Signal Processing (ICASSP), 2015.

[13] M. Z. Alom, V. Bontupalli, and T. M. Taha, “Intrusion detection using deep belief networks,” in IEEE National Aerospace and Electronics Conference (NAECON), 2015.

[14] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016.

[15] Y. Li, R. Ma, and R. Jiao, “A hybrid malicious code detection method based on deep learning,” International Journal of Security and Its Applications, 2015.