Machine Learning Methods For Cyber Security

RASHMI DEEP, M.E. Research Scholar, Department of Computer Science and Engineering, Chandigarh University Gharuan, Mohali, Punjab, Pin Code-140413, India , rashmideep21@gmail.com

Dr. VINAY GOYAL, Professor, Department of Computer Science and Engineering, Chandigarh University Gharuan, Mohali, Punjab, Pin Code-140413, India, hod.cse@cumail.in

Abstract:

With the growing use of internet by people, the protection of sensitive data has become something which we take into consideration and it is a very severe issue. Because in today's era we heard the term cyber security by every person and related to this the cyber security conditions are not very good.

So, to overcome these issues we introduced the Machine Learning methods for the analysis of network for the Intrusion Detection. Because data security is very important so the method of Machine Learning are introduce. In this review we discuss how machine learning can be use to notice the security threat.

Introduction:

the expansion in-deep alliance of the web and social life, the Internet is switching in a route that how a person attain and work, infact it also disclose us to developing severe security threat . Cyber security is method and action design to safeguard the computer hazards , network and information from cyber attacks such as alteration of data and unauthorized access of a hacker. Network security consisting many parameter such as network security system or we can also say that a computer security system. It contains firewalls, software for the protection of computers and intrusion detection systems. IDSs is used to find and identify the unusual behavior of the system. The unusual behavior are describe as modification of information, extension of data, copying of information.

Basically we have 3 main type of network analysis for intrusion detection system:

1. Misuse based- This analysis is is sometimes reffers as a signature based analysis.

2. Anomaly-based

3. Hybrid-based

In the verification method based on improper use, we try to observe and detect known attacks using the signature method.

Anomaly methods are generally used for simple network and actions of the system. It define abnormalities as a normal behavior. Basic limitation in anomaly-based approach is possible to give a false alarm at the very high rate.

Hybrid – based method integrate misuse and anomaly detection method . In this we expand the detection rate of the well known threat and it also cut down the false positive rate of unknown threat. Various machine learning methods uses Hybrid anomaly. The main idea of this paper is who wants to learn about intrusion detection in machine learning. This paper gives a deep focus on machine learning methods. We get knowledge regarding the previous paper that what type of challenges facing during the detection methods at the time of intrusion detection method. If we talk about wired network then the unauthorized user have to go through the multi layer of the firewall and OS(operating system) security or achieve physical entry at the structure. But in the opposite regarding wired network, in wireless network unauthorized user can attack at any point or at any node. So the wireless networks are highly vulnerable to any massive attack and to prevent those attacks are more difficult than the wired networks. Because as we know wired networks are much more secure than the wireless networks.

Interrelation and variation in machine learning and deep learning:

Artificial intelligence is newly introduced technical science that evolves new theories, methods and techniques that replicate, increases, develop person brilliance. In this area of science builds computer vision, robotics, expert system and natural language processing. AI can replicate the action and procedure of human response and thoughts. But thoughts

like a human mind can also improve person brilliance.

Machine Learning is a part of the AI and very deeply related to computational statics which targets on predicting output with using computers. Sometimes ML may be unsupervised for establishing the new relation for numerous entities and to evolve some new rarity.

ML basically focuses on:

1. Classification
2. Regression

Which is basically related to past known features of the basics data.

Deep learning is a recent concept in M(machine learning). It is basically used in the development of the Neural structure that replicate the person mind for training new data. It imitate person mind mechanism for analysis of data like texts, sound and images.

Variation between ML and DL:

1. **Data Dependency**: Performance measurement is main difference between the DL and ML as the data increases in higher volume. Deep learning uses where large amount of data processing are needed.
2. **Hardware Dependency**: Many type of matrix operation are needed for the algorithm so the GPU is highly used to suppress the matrix operations. GPU is the hardware which is used in the deep learning.
3. **Feature processing**: It reduces the complexity of data by using the feature extractor. It generate pattern to make learning algorithm work more efficient.
4. **Execution Time**: DL algorithm takes more times for execution rather than any other algorithm due to the existence of many type of attribute on which the DL algorithm have to work. So it takes longer process time for execution

Security Set Information for web:

Data creation is the basic part of the network secrecy gurantee. The right choice and fair use of the data is necessary for developing Information security research.

Information can be gathered in two forms:

1. Directly
2. Current public data set

Direct access gather the data of required cyber data using some software tools such as Wireshark or Win Dump to acquire network packet. With the help of these approaches we gather short term of minimal amount of data addition, purchase and storage costs will be expand. In this section we also mention some security data set those are admittance on the internet.

1.**DARPA ID'S Data Set:** It works under "DARPA and AFRL/SNHS", which gathered and published by cyber system and technology group(CSTG) of MIT Lincoln Lab. It classify intrusion detection system.

2.KDD CUP 99 Data Set: It is most generally used training set. The base of this data set depends on DARPA 1988 data set.

3.NSL-KDD Data Set: This is the extended version of the "KDD CUP 99" information set. It improvises limitation of KDD CUP 99.

ML Algorithm For Cyber Security: It enlighten the application of ML algorithm in network security.

1.SUPPORT VECTOR MACHINE(SVM): SVM is powerful and detailed method in all type of machine learning algorithm. It works on the idea of decision boundaries. It is the integration of support vector classification(SVC) and support vector regression(SVR).

SVC works on the idea through decision Bounds. It holds binary and multiclass classification.

2.K-NEAREST NEIGHBOUR: It works on a distance function which measures variation and collateral among particulars. The standard euclidean distance d(x,y) among identical x and y defined as:

$$d(x,y) = \sqrt{\sum (x-y)2}$$

In which x shows the kth features of item x. y is featured of item y.
and n whole number of feature in data set.

3.Decision Tree: In this classifier technique it consist tree structure where every internal node shows a analysis on one tract and every devision represent analysis output. Every leaf nodes represent category.

It is a prediction model it show a drawing among item traits and item codes. Every divergence root shows a attribute value of the item shows the route from first(root) node to last(leaf) node. As a resultant of decision tree only single output occurs at last.

Discussion And Future Scope:

In this paper we get to known about number of intrusion detection which are related to machine learning. The limitation in this area:

1.Scanty of Dataset

2.Non uniformity of Evaluation Metrices.

3.Low acknowledgment to deployment efficiency

Conclusion:

In this article we present a review ML methods related to networking security. It shows the recent application of ML of intrusion detection. Till now there has no such appropriate method evolved for intrusion detection. Every method has its own limitation and advantages. Data sets for intrusion detection system are highly notable for training and to test the system.

Refrences:

[1] S. Aftergood, ``Cybersecurity: The cold war online,'' Nature, vol. 547, no. 7661, pp. 30_31, Jul. 2017.

[2] A. Milenkoski, M. Vieira, S. Kounev, A. Avritzer, and B. D. Payne, ``Evaluating computer intrusion detection systems: A survey of common practices,'' ACM Comput. Surv., vol. 48, no. 1, pp. 1_41, 2015.

[3] C. N. Modi and K. Acha, ``Virtualization layer security challenges and intrusion detection/prevention systems in cloud computing: A comprehensive review,'' J. Supercomput., vol. 73, no. 3, pp. 1192_1234, 2017.

[4] E. Viegas, A. O. Santin, A. França, R. Jasinski, V. A. Pedroni, and L. S. Oliveira, ``Towards an energy-ef_cient anomalybased intrusion detection engine for embedded systems,'' IEEE Trans. Comput., vol. 66, no. 1, pp. 163_177, Jan. 2017.

[5] A. Patcha and J.-M. Park, ``An overview of anomaly detection techniques: Existing solutions and latest technological trends,'' Comput. Netw., vol. 51, no. 12, pp. 3448_3470, Aug. 2007.

[6] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, ``A survey of intrusion detection techniques in Cloud,'' J. Netw. Comput. Appl., vol. 36, no. 1, pp. 42_57, 2013.

[7] S. Revathi and A. Malathi, ``Adetailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion

detection,'' in Proc. Int. J. Eng. Res. Technol., 2013, pp. 1848_1853.

[8] D. Sahoo, C. Liu, and S. C. H. Hoi. (2017). ``Malicious URL detection using machine learning: A survey.'' [Online]. Available: https://arxiv.org/abs/1701.07179

[9] A. L. Buczak and E. Guven, ``A survey of data mining and machine learning methods for cyber security intrusion detection,'' IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1153_1176, 2nd Quart., 2016.

[10] M. Soni, M. Ahirwa, and S. Agrawal, ``A survey on intrusion detection techniques in MANET,'' in Proc. Int. Conf. Comput. Intell. Commun. Netw., 2016, pp. 1027_1032.

[11] R. G. Smith and J. Eckroth, ``Building AI applications: Yesterday, today, and tomorrow,'' AI Mag., vol. 38, no. 1, pp. 6_22, 2017.

[12] P. Louridas and C. Ebert, ``Machine learning,'' IEEE Softw., vol. 33, no. 5, pp. 110_115, Sep./Oct. 2016.

[13] M. I. Jordan and T. M. Mitchell, ``Machine learning: Trends, perspectives, and prospects,'' Science, vol. 349, no. 6245, pp. 255_260, 2015.

[14] Y. LeCun, Y. Bengio, and G. Hinton, ``Deep learning,'' Nature, vol. 521, pp. 436_444, May 2015.

[15] G. E. Hinton, ``Deep belief networks,'' Scholarpedia, vol. 4, no. 5, p. 5947, 2009.

[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ``Gradient-based learning applied to document recognition,'' Proc. IEEE, vol. 86, no. 11, pp. 2278_2324, Nov. 1998.

[17] L. Deng and D. Yu, ``Deep learning: Methods and applications,'' Found. Trends Signal Process., vol. 7, nos. 3_4, pp. 197_387, Jun. 2014.

[18] I. M. Coelho, V. N. Coelho, E. J. da S. Luz, L. S. Ochi, F. G. Guimarães, and E. Rios, ``A GPU deep learning metaheuristic based model for time series forecasting,'' Appl. Energy, vol. 201, no. 1, pp. 412_418, 2017.