INTRUSION DETECTION SYSTEM USING PROCESS MINING AND DATA MINING

[1*]Meghana R. Solanki
[1]Assistant Professor
[1]Computer Department, DYPCOE, Pune, India
(*meghana.solanki@dyptc.edu.in)

Abstract: Process mining give backing to the analysis of business processes hinge on event logs. It not only determines but also improves the legitimate procedures from the log events. It is promptly accessible in the existing data framework. It gives the mislaid relation among one hand processes. It demonstrates examination as well as information arranged. It establish investigation then again execution and finally compliance. In this paper, we have reviewed the intrusion detection using process mining as well as data mining. We correlated the various algorithms which have been operated in process mining. Also we presented a comparative study of data mining techniques for intrusion detection. We have used ISCX2012 datasets a yardstick for our testing. This dataset demonstrated that relatively old methods surpass some of the techniques highly used by the society.

Key Words - Intrusion detection, Process Mining, Information Security, Audit Trails, Data mining, Machine learning, ISCX2012.

1. INTRODUCTION

Process mining approach is very fruitful to find out the intrusions in efficient manner. Previous methods which were belonging to data mining approaches are not able to determine Intrusions in adequate manner. Process mining methods correctly relate possible information to end- to- end business forms. The crossing point of Business Process Management (BPM) as well as Data Mining is Process mining. This undertaking points in the examination as well as comprehension on the most proficient method to enhance process fertility and procedure's apprehension. Calculations as well as applications in current age are the significance of process mining. The bottom line logic of process mining is [1] to eliminate information from logs which is reported by a data framework. Process Discovery, Process Conformance and Process Enhancement are types of Process Mining. Process Discovery means to uplift the scientists about the profound elements of process mining. Process Conformance takes a current procedure blueprint. Process Enhancement takes the existing process schema. Lasagna processes well as Spaghetti forms are types of Processes in Process Mining. Lasagna process is a procedure with restricted endeavors is believable to make a settled upon process demonstrated. It has a wellness of no less than 80%.Spaghetti forms are less formulated than Lasagna forms, just some of procedure mining procedures can be connected. We are breathing in an interconnected world in which around half billion mobile devices as well as connections were added in 2016-17 [2]. This high connectivity contains an impressive amount of cyber menace. In fact, cyber-attacks [3 - 4] are assumed as a fatal weapon. Intrusion detection was described as "the process of tracking the events which occurs in a computer system or network. It is used for inspecting them for signs of intrusions. Intrusion is described as attempts to negotiate the confidentiality, integrity, or to sidestep the security mechanisms of a computer or network" [5]. Based on its detection mechanism, there are two types of an Intrusion Detection System (IDS), first is a misuse based and second is anomaly based [6]. A set of signatures describing known attacks relied by Misuse based detection. A prebuilt model of normal behavior is relied by anomaly-based detection. Also it takes any breach from that model as an intrusion attempt. KDD99 is a dataset from the late nineties that can no longer echo network traffic of today's era. Thus, we apply a more recent dataset for our study which is ISCX2012.This dataset reflects the current network traffic. In summary this work has two targets one is the appraisal of data mining techniques against public, recent, and real network traffic and second is the interpretation of the impact of dataset on the performance of the investigated methods.

2. LITERATURE SURVEY

In this paper [7-8], an author showed that, procedure mining can be not only helpful strategy with quicker outcomes but also a capacity to check conformance. A subsequent network between occasions is necessary in process mining methods. In this paper [9], an author showed that, artificial intelligence methods were left due to their inability to cope with the increase of network traffic In this paper [10-11], an author showed that, the proposal works usually make the use new methods such as Random Forest, Support Vector Machine, Artificial Neural Networks and C4.5 decision tree. In this paper [12-13], an author showed that, the dataset KDD99 dataset is the most widely used benchmark in case of intrusion detection. In this paper [14], an author showed that, KDD99 has been blamed for their impractical personalities which produce amiss reflection of IDSs work in the real world. In this paper [15], an author showed that, due to changed network traffic and network attacks, KDD99 has its evaluated version NSL-KDD99 which no longer reflects network traffic today's world. In this paper [16-17], an author showed that, in this background, many datasets have been recommended for intrusion detection. Many of them are not only heavily anonymized but also payload-

stripped; this is applicable for UNIBS dataset as well as LBNL dataset. In this paper [18], an author showed that, other datasets are not pragmatic; this is the case of DEFCON datasets. The traffic for this dataset was captured through hack competition. In this paper [19], an author showed that, The Information Security Centre of Excellence (ISCX) research centre proposes ISCX2012dataset. This is a labeled dataset for intrusion detection that encompasses complete practical network traffic.

3. INTRUSION DETECTION SYSTEM

Intrusion is an act of attempt to sidestep the security mechanism of network or computer as well as to compromise with probity, familiarity and opportunity. In current digital world, defending of the network from cyber-attacks, intrusion etc is the main challenging task. We use a durable and ardent system which gives safety from unauthorized access of information. The main task of intrusion detection system is to continuously observe the network ad well as systems to signal immediately in case of any malicious codes, event or activity is raised. The intrusion detection system assembles the information. The after investigating and dealing with it makes the conclusion that whether the current or any operation is breaching the safety rules or not and generates the signal to the network controller in case of exceptional action or movement. Intrusion Detection System is shown in figure 1.

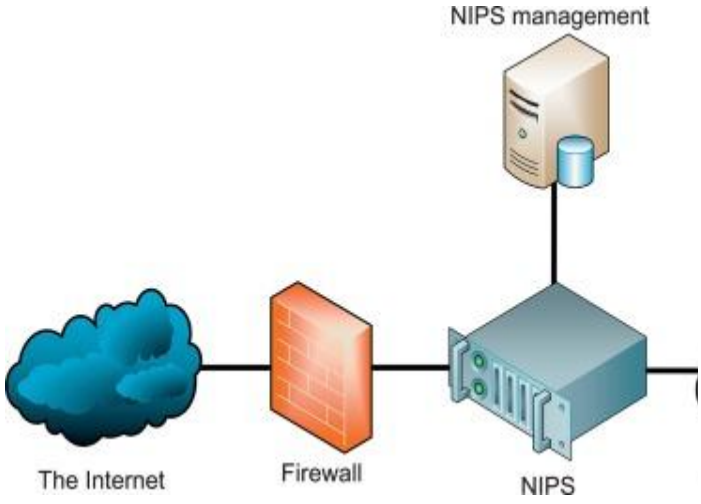

Figure 1 Intrusion Detection System

4. COMPARATIVE ANALYSIS OF PROCESS MINING ALGORITHMS

The **α-algorithm** is an algorithm which is utilized in process mining. Its main aim is to reconstruct causality from a set of sequences of events. It was first recommended by van der Aalst, Weijters and Măruşter. It automatically picks up process model. It grabs all behavior. It contributes replaying of behavior. General workflow nets may encompass several categories of constructs which is missed by the α-algorithm. It is not authoritative to identify all situations. It does not permit step siding of actions. It does not recognize the frequencies. It does not grant latent sites.

Heuristics Miner is an algorithm that acts on the Directly-Follows Graph. It provides plan to handle with noise along with to find common constructs. It is advantageous to deal with turbulence. It should be utilized with less number of events. It does not assurance of sound process models.

5. EXPERIMENTATION

In this section, we implement our examinations on three supervised learning methods. These are Random Forest, Support Vector Machine and RPART. Along with this, we add one anomaly detection method for our experiment. Specifically, One-Class Support Vector Machine trained only on regular items. The election of those methods based on their ancient management in the literature as well as the actual trends. Differently, SVM and RF are well known machine learning methods. Also they are widely used machine learning methods. RPART implements the recursive partition principal. This principle is implemented by ID3, C4.5, C5.0, and RF. Differently, OCSVM have been utilized freshly for intrusion detection. We perform our experiments in java environment since it contains packages and methods for all our selected methods.

5.1. Pre-processing

Over this experimentation, we utilize the XML version of ISCX2012. However, sonner starting our experiments, there is need of some pre-processing steps to gain a more persistent dataset. This fits not only the investigated methods but also the domain of network intrusion detection.

5.2. Evaluation metrics

We estimate the performance of classifiers according to how superiorly they segregate an item. In Table 1, we specify the context of intrusion detection, how a forecasted category of a given item falls into one of four positions. It can be a true positive, a true negative, a false positive, or a false negative prediction

Table 1 Confusion Matrix

Confusion Matrix		Predicted class	Predicted class
		Normal	Attack
Actual class	Normal	True negative (TN)	False positive (FP)
Actual class	Attack	False negative (FN)	True positive (TP)

From the confusion matrix, three metrics are driven to check out the performance of the investigated methods.

Detection rate DR= TP/ TP + FN

False alarm rate FAR= FP/ TN + FP

Accuracy = TP+TN / TP+TN+FP+FN

5.3. Results and Discussions

Table 2 shows the detection rates gained by each method against the training dataset. OCSVM has a small detection rate where as Random Forest have an acceptable detection rate. Table 2 shows the false alarm rates obtained by the reviewed methods. OCSVM has an improbable false alarm rate. Random Forest outperforms all the methods regarding false alarm rates with0.24% of false alarm rate Table 2 shows the accuracies of the reviewed method. The reviewed methods have a high accuracy except for OCSM. RF and SVM linearly gain in accuracy when we add more data to the training dataset.

Table 2 Detection Rates

Metrics	OCSVM	RPART	SVM	RF
Detection Rates	12.09%	97.63%	99.36%	99.67%
False Alarm Rates	11.52%	2.26%	1.14%	0.24%
Accuracies	78.25%	97.59%	99.31%	99.68%

6. CONCLUSION

Process mining is a demanding appliance for present day associations that need to inspect nontrivial variable patterns. Information mining methods plan to illustrate as well as comprehend reality considering notable information. It is a lower level of investigate, because these systems are nor process driven. Process mining is induced by verifiable occasion information contrary to hand-made models. An extension amongst BPM as well as Data Mining is process mining. Process mining is not limited to only process finding. A new course for investigating is opened by associating occasion log and process show. In this paper, we have reviewed the intrusion detection using process mining as well as data mining. we conducted a comparative study on some popular data mining methods applied for intrusion detection. The experiment for study has used a benchmark ISCX2012 dataset. This dataset echo the network traffic of today's era. Our experimentations show that, Random Forest and Support Vector Machine clearly overwhelm the methods used in the current trend with Random Forest taking the lead. One-Class Support Vector Machine shows a poor detection rate making it impracticable as a standalone method.

REFERENCES

[1] Van der Aalst, W.M.P & De Medeiros, A.K.A, Process miningand security: Detecting Anomalous process executions andchecking process conformance, Electronic Notes in TheoreticalComputer Science, Vol 121, No. 4, pp. 3–21, 2005.

[2] T. J. Barnett, A. Sumits, S. Jain, and U. Andra, "Cisco VisualNetworking Index (VNI) Update Global Mobile Data TrafficForecast," *Vni*, 2015. [Online]. Available:http://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/complete-white-paper-c11-481360.html.

[3]A. Ahmim and N. Ghoualmi--Zine, "A new adaptive intrusiondetection system based on the intersection of two differentclassifiers," *Int. J. Secur. Networks*, vol. 9, no. 3, pp. 125–132,2014.

[4] A. Ahmim and N. Ghoualmi Zine, "A new hierarchical intrusiondetection system based on a binary tree of classifiers," *Inf. Comput.Secur.*, vol. 23, no. 1, pp. 31–57, 2015.

[5] R. Bace and P. Mell, "NIST special publication on intrusiondetection systems," 2001.

[6] J. Peng, K.-K. R. Choo, and H. Ashman, "User profiling in intrusiondetection: A review," *J. Netw. Comput. Appl.*, vol. 72, pp. 14–27,2016.

[7] Pawar, M.V, & Anuradha J, Network Security and Types ofAttack in Network, Procedia Computer Science 48(2015),ELSEVIER, ScienceDirect, pp 503-506, 2015.

[8] Ved P Mishra & Balvinder Shukla, "Development of Simulatorfor Intrusion Detection System to Detect and Alarm the DDoSAttacks", IEEE INTERNATIONAL CONFERENCE on Infocom Technologies and Unmanned Systems (ICTUS'2017)(Trends and Future Directions), 10-12 December 2017..

[9] S. X. Wu and W. Banzhaf, "The use of computational intelligencein intrusion detection systems: A review," *Appl. Soft Comput.*, vol.10, no. 1, pp. 1–35, 2010.

[10] H. Chauhan, V. Kumar, S. Pundir, and E. S. Pilli, "A ComparativeStudy of Classification Techniques for Intrusion Detection," in *2013International Symposium on Computational and BusinessIntelligence*, 2013, pp. 40–43.

[11] R. Beghdad, "Critical study of neural networks in detectingintrusions," *Comput. Secur.*, vol. 27, no. 5–6, pp. 168–175, Oct.2008.

[12] KDD, "KDD Cup 1999 Data,http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html," *KDDCup 1999 Data*, 1999. [Online]. Available:http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[13] E. Kabir, J. Hu, H. Wang, and G. Zhuo, "A novel statisticaltechnique for intrusion detection systems," *Futur. Gener. Comput.Syst.*, 2017.

[14] M. V Mahoney and P. K. Chan, "An Analysis of the 1999DARPA/Lincoln Laboratory Evaluation Data for Network AnomalyDetection," in *Recent Advances in Intrusion Detection: 6th International Symposium, RAID 2003, Pittsburgh, PA, USA,September 8-10, 2003. Proceedings*, G. Vigna, C. Kruegel, and E.Jonsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003,pp. 220–237.

[15] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailedanalysis of the KDD CUP 99 data set," in *2009 IEEE Symposium onComputational Intelligence for Security and Defense Applications*,2009, pp. 1–6.

[16] NTW, "UNIBS: Data sharing," 2009. [Online]. Available:http://netweb.ing.unibs.it/~ntw/tools/traces/.

[17] "LBNL/ICSI Enterprise Tracing Project." [Online]. Available:http://www.icir.org/enterprise-tracing/Overview.html.

[18] DEFCON, "CTF Archive." [Online]. Available:https://www.defcon.org/html/links/dc-ctf.html.

[19] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Towarddeveloping a systematic approach to generate benchmark datasetsfor intrusion detection," *Comput. Secur.*, vol. 31, no. 3, pp. 357–374, 2012.