International Journal of Engineering & Technology, 7 (4.28) (2018) 626 -630

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Adequacyscrutinyof Intrusion Detection Techniques Over Discrete Dataset's

Neha Singh [1], Dr. Deepali Virmani[2]

[1]*Research Scholar,Guru Gobind Singh Indraprastha University ,Delhi, India*
[2]*Associate Professor Bhagwan Parshuram Institute of Technology ,Delhi, India*

Abstract

Intrusion detection is one of the major issues in wireless sensor networks. There is a drastic change in algorithms dealing with intrusion detections due to rapid change in types of intrusion. The paper presents a scrutiny of various intrusion detection techniques over various datasets and illustrates how these algorithms have evolved with time. The scrutiny is based on various parameters such as various types of technologies used to propose new systems and number of publications over a period of time, accuracy of intrusion detection rate of various data sets. The analysis concludes that their is a significant increase in new intrusion detection systems over time. The analysis also concludes that researchers are gradually shifting from old data sets to new data set to validate their systems. There is a rapid growth in systems using machine learning for intrusion detection systems.

Keywords: *Intrusion; wireless sensor network; intrusion detection techniques; data mining; machine learning; fuzzy rules.*

1.Introduction

Intrusion can be defined as any activity that hinders the objectives of network security. Due to increase in security breaches in the recent years, intrusion detection system (IDS) has gained lot of attention. IDS aims at identifying any unauthorized activity so as it can be prevented to harm the network or the communication. Nowadays, intrusion detection not only aims at providing authentication, encryption or any other prevention mechanism like in the traditional network but to determine the novel attacks in ad hoc networks and classifies them based on various criteria.

Intrusion detection techniques can be broadly classified as knowledge (Signature)-based detection, anomaly detection or hybrid detection methods with each having its pros and cons [23]. Signature based detection methods uses the knowledge of known attacks to construct detection rules that are further used to tell whether the network security is compromised or not. In this, a characteristic profile of an intrusion is developed and stored in database offline before the system begins operation. It is the simplest detection mechanism that compares the current activity with the knowledge stored in the intrusion database and thus has low false positives and high response speeds. However, the major drawback of this technique is that it cannot detect any novel attack for which there is no knowledge stored in the database

In contrast, anomaly detection method [35] builds a model of what is considered as a normal behavior of the network and detects various deviations from normal behavior. Anomaly detection methods thus have an edge over the knowledge-based methods, they have the capability of detecting attacks. Such methods however suffer from high rate of false alarms because any unseen activity even though legitimate will be classified as malicious. In order to account for this problem many IDS systems are based on data mining and machine learning techniques [7, 29]. Data mining techniques used with anomaly detection methods judge the nature of intrusion and helps to lower the false positives thus giving high accuracy.

[6] In order to avoid the shortcomings of signature and anomaly based methods various hybrid approaches have been developed and are found to have better results. In spite of these anomaly based methods on data mining are common and used in practice. Therefore the aim of this survey is to conduct a comparative analysis of various intrusion detection schemes based on data mining

The rest of the paper is organized as follows. Section 2 presents related work in the field of intrusion detection system in wireless sensor network. In Section 3, a comparative analysis of various Intrusion Detection System techniques is presented. Section 4 shows the analysis of comparative study. Finally, in Section 5, the concluding remarks and future work is presented.

2. Related Work

Paper[10] provides a self learning technique of Intrusion detection system. The system classifies an unknown data pattern into an attack or non attack. It compares and considers variation from attack free scenario. The data set used is NSL-KDD. Their system yields an accuracy of 91%. Paper [15] is a survey paper that focuses on IDS using self-organizing map. They concluded that techniques based on self-organizing maps have poor detection rate for U2R and R2L attacks. The survey also concluded that HSOM and GHSOM are advance model of SOM with better performance. Paper [11] have proposed a hierarchy anomaly intrusion detection model. The model combines fuzzy c-means. It is based on genetic algorithm and SVM. The results proved the proposed model could effectively detect majority of network attack types. NSLKDD dataset is used as data set and accuracy of the system is 98.86%. In Paper[33], 253 log files of computer-science students and 206 log files of non-computer-science students was collected from computer center of TungHai University. The author proposes

Intrusion Detection and Identification System (IDIS). They built a profile for each user in an intranet. This keeps a track of usage habits as forensic features. A recognition accuracy of 98.99% was achieved. Paper[24]presented a malicious node detection scheme which is neighbor-based. Their simulation shows that most of the malicious nodes reporting against their own readings are correctly detected. Paper[8] gives a detailed overview of the attacks and some methods for secured data transmission. Paper[9] explains the vulnerability of attacks in WSN . The author has suggested a trust model that is fuzzy logic based and multi-attribute. The model consists of elapsed time at node, correctness,fairness as trust metrics and message success rate. The final trust value of every node is computed by fuzzy computational theory. Paper[17] proposed EDADT (Efficient Data Adapted Decision Tree) algorithm for Intrusion Detection. Their algorithm was able to solve four issues:High Level of Human Interaction, Classification of Data, Effectiveness of Distributed Denial of Service Attack, and Lack of Labeled Data. KDD Cup99 Dataset was used and yields an accuracy of 98.12%. Paper [36] provides a distributed fault detection algorithm. The authors have also employed a sliding window for eliminating the delay involved in time redundancy scheme. Their system successfully identifies sensor nodes with permanent faults and the performance is negligible degraded while tolerating many transient faults. Paper[3] proposes an Internal Intrusion Detection with an extra feature of protection. It detects insider attacks by usingforensic techniques and data mining. The author creates users personal profiles for tracking users habits. The author concludes that IIDPS has an accuracy of 94.29%. Paper[2] proposes an adaptive and online Network basedIntrusion Detection System. The system is based on Clustering and Extreme Learning Machines. Their system detects unknown and known attacks and updates itself according to new trends. The performance is evaluated using NSL-KDD data set. Paper[32] provides a review of intrusion detection systems from machine learning. An extensive survey covers 55 studies between period 2000-2007. Paper[18] proposes a Stop Transmit and Listen (STL) scheme to find a malicious node. In this scheme, every node has built-in time limit to stop transmission. If a malicious node sends data where rest every node has stopped transmitting, it is caught by their neighbors. Paper[22] presents a dynamic clustering technique to detect the malicious nodes. It overcomes the drawbacks of conventional Watchdog Mechanism. The author have claimed to improve network security and saves energy. Paper[34] proposes a hybrid classifier for intrusion detection with the capability of multiple-level detection. The classifier combines unsupervised Bayesian clustering and supervised tree classifiers. The Performance is measured using KDDCup 99 Data set.

3.Comparative Study

In this paper, a detailed analysis of the papers published in the field of Intrusion Detection System for Wireless Sensor networks is shown. Several factors are taken into consideration for the comparative analysis. They are:

1. Name of the Proposed Algorithm
2. Year of Publication
3. Data set used
4. Features
5. Accuracy

The table is as follows:

Table 1: A Comparative analysis of different ID

S.No.	Year	Technique	Data Set	Feature	Accuracy	Reference
1.	2018	The proposed enhancement of the algorithm is done by adding the SOM training process.	KDD'99 dataset	Unlabeled intrusions are detected by training SOM.	90%	[1]
2.	2017	Novel IDS proposed, based on ELM	NSL-KDD data set	The capability of detecting novel and known attacks is offered, The system is updatedas in when new trends and patterns in data comes.	67%	[2]
		Internal Intrusion Detection and Protection System (IIDPS)	Recorded user Data.	Detect insider attacks by using forensic techniques and data mining.	94.29%	[3]
		A multi-level hybrid intrusion detection model that uses sup- port vector machine and extreme learning machine	KDD Cup 1999 dataset	Designs a model to deals with real intrusion detection problems Classify network data into normal and abnormal behaviors.	95.75%	[4]
		SSL algorithm for improving the classifier performance	NSL-KDD dataset	The classifier performance is improvedfor IDSs.	84.12%	[5]
3.	2016	Feature Reduction technique	NSL-KDD	Provide IDS based on self-learning technique.System classifies pattern as an attack or non-attack.	91%	[10]
		a new hierarchy anomaly intrusion detection model that propose SVM and combines the fuzzy c-means (FCM) based on genetic algorithm	NSLKDD dataset	It provides solution for false alarm and detection rate. Effectively detect majority of network attack.	98.86%	[11]
4.	2015	a method based on the combination of Decision Tree (DT) algorithm and Multi-Layer Perceptron (MLP)	KDD CUP 99	Identifies various attacks. The accuracy is high and reliability is increased.	97.93% for decision tree 99.71% for MLP	[12]
		Proposed approach	KDD-Cup 99 dataset	Combination of nearest	97.12%	[13]

		combines cluster centers and nearest neighbors for effective and efficient intrusion detection, namely CANN.		neighbors and cluster centers .		
		Combination between GFS and the OVO learning scheme	KDD Cup 99	A smoother borderline is enabled between the concepts, and a higher interpretability of the rule set is allowed.	99.81%	[14]
5.	2014	IDS using Self Organizing Map.	KDD cup 1999	Have less computation time and better detection rate.	99.6%	[15]
		A new clustering approach for anomaly intrusion detection by using the approach of K-medoids method of clustering and its certain modifications	kddcup99	Provides high detection rates Provides less false negative rate	96.38%	[16]
		In this EDADT (Efficient Data Adapted Decision Tree algorithm	KDD Cup 99 dataset	IDS and data mining concepts are integrated to identify the hidden and relevant data. Reduced alarm rate and better accuracy	98.12%	[17]
6.	2013	Algorithm incorporates Information Gain Ratio (IGR) and K-mean algorithm to SVM for intrusion detection	NSL-KDD	The paper efficiently characterizes normal traffic and the abnormal traffic is distinguished using Support vector machine.	99.37%	[19]
		Outlier detection algorithm	our proposed approach when applied on both the real dataset from Intel Berkeley Research lab and synthetic dataset.	Provides outlier detection and data clustering simultaneously. Distinguishes between error due to faulty sensor and an error due to an event	100% detection rate	[20]
		Decision trees have been adopted as classification algorithm in the detection process of the Central Agent and their behavior has been analyzed		The IDS is hybrid, lightweight, distributed. Uses both anomaly-based and misuse-based detection techniques.	Classification And Regression *Tree* (CART) 99.57%	[21]
7.	2011	Proposed algorithm combines fuzzy set theory with GNP,	KDD99Cup and DARPA98 databases from MIT Lincoln Laboratory	Proposed method deals with the mixed database. Many important class-association rules are extracted to contribute for enhancing detection ability.	94.4%	[25]
		Hybrid learning approach through combination of K-Means clustering and Naïve Bayes classification.	KDD Cup ’99 dataset.	All data into groups are clustered before applying a classifier.	99.6%	[26]
		This study proposed an SVM-based intrusion detection system, which combines a hierarchical clustering algorithm, a simple feature selection procedure, and the SVM technique	KDD Cup 1999 training set	Detects DoS and Probe attacks.	95.72%	[27]
8.	2010	we propose a new approach, called FC-ANN, based on ANN and fuzzy clustering,	KDD CUP 1999 dataset	Less false positive rate,high detection rate and stronger stability.	96.71%	[28]
9.	2009	The proposed classification algorithm uses fuzzy association rules for building classifiers	KDD-99 dataset	Classification engine uses Association Based Classification	80.6%	[30]
		Grey self-organizing map (GSOM) model is proposed	dataset from the DARPA’98 training data	Analyze DOS attack Proposed GSOM model Uses grey relational coefficients and its	97%	[31]

				correlative functions.		
10.	2008	Intrusion Detection and Identification System (IDIS)	Collected 253 computer-science students' log files and 206 non-computer-science students' log files from computer center of TungHai University as the experimental data.	A profile is built for each user in an intranet. Keeps a track of IDIS. Identifies the underlying user in intranet	98.99%	[33]
		Proposed algorithm combines the supervised tree classifiers and unsupervised Bayesian clustering to detect	KDDCUP99	Detects intrusions with an extremely low false-negative rate	For known attacks - 99.19% For unknown attacks -80.63%	[34]
11.	2005	MMIDS (Multi-step Multi-class Intrusion Detection System	KDD CUP 1999	Detection of new unknown attacks. Provides information about detected types of attack Cost-effective maintenance	98.24%	[37]

4.Analysis

According to our analysis, we have chosen various parameters such as accuracy over data set used, techniques used and years for analysis of papers. All these parameters are explained below.

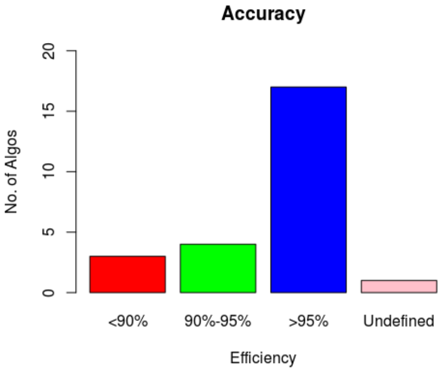

Fig.1: Analysis of accuracy of algorithms in various papers

According to Fig.1, 64% of the papers have an accuracy of more than 95% for KDD cup99 dataset.Only 12% of the papers have an accuracy of less than 90%.

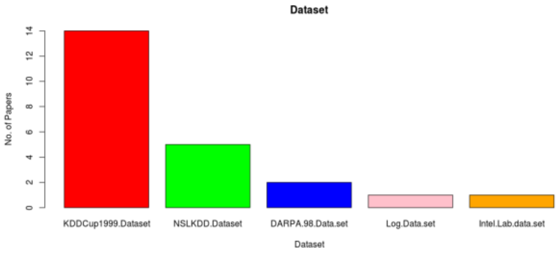

Fig. 2: Analysis of Data set used in various papers

According to our analysis, there are majorly five types of datasets used for research in testing an Intrusion Detection System for Wireless Sensor networks. Out of these five, a majority of papers have used KDDCUP99 data set. Few papers have used NSLKDD dataset. Only two papers have used DARPA9 data set and just 1 paper each have used Log data set and Intel labs data set.

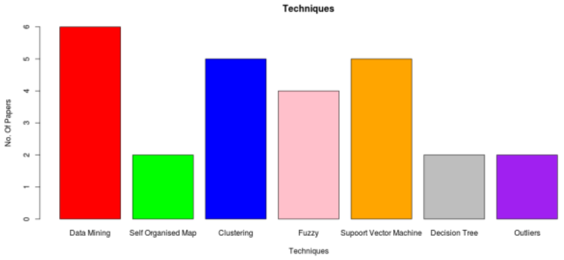

Fig. 3: Analysis of various techniques used in papers

According to the analysis, it is shown that data mining is the most suitable technique.

Fig 3 gives an overview of major techniques used for Intrusion Detection System in Wireless Sensor networks in our survey. Out of studied papers, Data mining technique is used in around 26% of the papers. Clustering and support vector machine are at the same level with around 19% share of each technique. Fuzzy rules are just behind clustering and support vector machine with a share of 15%. Self organized map, decision tree and outliers lag behind with a share of 7% for every technique.

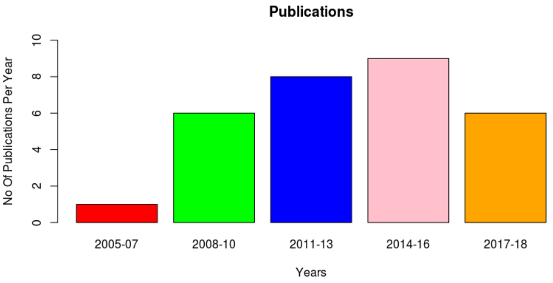

Fig 4: Analysis of number of publications for a period of 3 consecutive years

Fig 4 shows that the number for publications in the field of Intrusion detection Techniques in Wireless Sensor Networks have linearly increased over the past decade.

5.Conclusion

Intrusion detection systems in wireless sensor networks have been a very fascinating area of research for researchers. There is a rapid change in technology for dealing with intrusions. The paper presented a comprehensive survey of 25 various intrusion detection techniques over a decade. The analysis is done on

accuracy of various techniques, data sets used, various types of technologies used to propose new systems and number of publications over a period of time. There is a rapid growth in systems using machine learning for intrusion detection systems.
This analysis concludes that there is a significant increase in new intrusion detection systems over time. Data Mining is the best technique for the analysis of intrusion. KDD cup99 dataset is giving more accurate result. The analysis also concludes thatresearchers are gradually shifting from old data sets to new data set to validate their systems.

6.Future Scope

There is a need to shift from old data sets to new data sets thus opening a new research field to develop new data sets for validating intrusion detection systems.

References

[1] Borah, S., Panigrahi, R., & Chakraborty, A. (2018). An Enhanced Intrusion Detection System Based on Clustering. In *Progress in Advanced Computing and Intelligent Engineering* (pp. 37-45). Springer, Singapore.

[2] Roshan, S., Miche, Y., Akusok, A., & Lendasse, A. (2017). Adaptive and online network intrusion detection system using clustering and Extreme Learning Machines. *Journal of the Franklin Institute*.

[3] Leu, F. Y., Tsai, K. L., Hsiao, Y. T., & Yang, C. T. (2017). An internal intrusion detection and protection system by using data mining and forensic techniques. *IEEE Systems Journal*, *11*(2), 427-438.

[4] Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A. (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Systems with Applications*, *67*, 296-303.

[5] Ashfaq, R. A. R., Wang, X. Z., Huang, J. Z., Abbas, H., & He, Y. L. (2017). Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, *378*, 484-497.

[6] Otoum, S., Kantarci, B., & Mouftah, H. T. (2017). Detection of Known and Unknown Intrusive Sensor Behavior in Critical Applications. *IEEE Sensors Letters*, *1*(5), 1-4.

[7] Mrugala, K., Tuptuk, N., & Hailes, S. (2017). Evolving attackers against wireless sensor networks using genetic programming. *IET Wireless Sensor Systems*.

[8] Padmaja, P., & Marutheswar, G. V. (2017, January). Detection of Malicious Node in Wireless Sensor Network. In *Advance Computing Conference (IACC), 2017 IEEE 7th International* (pp. 193-198). IEEE.

[9] Prabha, V. R., & Latha, P. (2017). Fuzzy Trust Protocol for Malicious Node Detection in Wireless Sensor Networks. *Wireless Personal Communications*, *94*(4), 2549-2559.

[10] Mahapatra, B., & Patnaik, S. (2016). Self Adaptive Intrusion Detection Technique Using Data Mining Concept in an Ad-hoc Network. *Procedia Computer Science*, *92*, 292-297.

[11] Tang, C., Xiang, Y., Wang, Y., Qian, J., & Qiang, B. (2016). Detection and classification of anomaly intrusion using hierarchy clustering and SVM. *Security and Communication Networks*, *9*(16), 3401-3411.

[12] Esmaily, J., Moradinezhad, R., & Ghasemi, J. (2015, May). Intrusion detection system based on Multi-Layer Perceptron Neural Networks and Decision Tree. In *Information and Knowledge Technology (IKT), 2015 7th Conference on* (pp. 1-5). IEEE.

[13] Lin, W. C., Ke, S. W., & Tsai, C. F. (2015). CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. *Knowledge-based systems*, *78*, 13-21.

[14] Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. *Expert Systems with Applications*, *42*(1), 193-202.

[15] Choksi, K., Shah, B., & Kale, O. (2014). Intrusion detection system using self organizing map: a surevey. *Int. J. Engin. Res. Appl*, *4*(12), 11-16.

[16] Ranjan, R., & Sahoo, G. (2014). A new clustering approach for anomaly intrusion detection. *arXiv preprint arXiv:1404.2772*.

[17] Nadiammai, G. V., & Hemalatha, M. (2014). Effective approach toward Intrusion Detection System using data mining techniques. *Egyptian Informatics Journal*, *15*(1), 37-50.

[18] Sathyamoorthi, T., Vijayachakaravarthy, D., Divya, R., & Nandhini, M. (2014). A simple and effective scheme to find malicious node in wireless sensor network. *International Journal of Research in Engg. And Tech*, *3*(2).

[19] Jha, J., & Ragha, L. (2013). Intrusion detection system using support vector machine. *International Journal of Applied Information Systems (HAIS)-ISSN*, 2249-0868.

[20] Fawzy, A., Mokhtar, H. M., & Hegazy, O. (2013). Outliers detection and classification in wireless sensor networks. *Egyptian Informatics Journal*, *14*(2), 157-164.

[21] Coppolino, L., D'Antonio, S., Garofalo, A., & Romano, L. (2013, October). Applying data mining techniques to intrusion detection in wireless sensor networks. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on* (pp. 247-254). IEEE.

[22] Virmani, D., Soni, A., Hemrajani, M., & Chandel, S. (2013). Dynamic Clustering to Detect Malicious Nodes in Wireless Sensor Network.

[23] Srivastava, M., & Mishra, S. (2012). EXISTING TRENDS IN INTRUSION DETECTION - A COMPARATIVE ANALYSIS. *International Journal of Information Technology and Knowledge Management*,*5*(1), 79-83.

[24] Yim, S. J., & Choi, Y. H. (2012). Neighbor-based malicious node detection in wireless sensor networks. *Wireless Sensor Network*, *4*(09), 219.

[25] Mabu, S., Chen, C., Lu, N., Shimada, K., & Hirasawa, K. (2011). An intrusion-detection model based on fuzzy class-association-rule mining using genetic network programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(1), 130-139.

[26] Muda, Z., Yassin, W., Sulaiman, M. N., & Udzir, N. I. (2011, July). Intrusion detection based on K-Means clustering and Naïve Bayes classification. In *Information Technology in Asia (CITA 11), 2011 7th International Conference on* (pp. 1-6). IEEE.

[27] Horng, S. J., Su, M. Y., Chen, Y. H., Kao, T. W., Chen, R. J., Lai, J. L., & Perkasa, C. D. (2011). A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert systems with Applications*, *38*(1), 306-313.

[28] Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert systems with applications*, *37*(9), 6225-6232.

[29] Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, *12*(2), 159-170.

[30] Tajbakhsh, A., Rahmati, M., & Mirzaei, A. (2009). Intrusion detection using fuzzy association rules. *Applied Soft Computing*, *9*(2), 462-469.

[31] Wang, C. D., Yu, H. F., & Wang, H. B. (2009). Grey self-organizing map based intrusion detection. *Optoelectronics Letters*, *5*(1), 64-68.

[32] Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2009). Intrusion detection by machine learning: A review. *Expert Systems with Applications*, *36*(10), 11994-12000.

[33] Leu, F. Y., & Hu, K. W. (2008). A real-time intrusion detection system using data mining technique. *Journal of Systemics, Cybernetics and Informatics*, *6*(2), 36-41.

[34] Xiang, C., Yong, P. C., & Meng, L. S. (2008). Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees. *Pattern Recognition Letters*, *29*(7), 918-924.

[35] Davanzo, G., Medvet, E., & Bartoli, A. (2008, September). A comparative study of anomaly detection techniques in Web site defacement detection. In *IFIP International Information Security Conference* (pp. 711-716). Springer, Boston, MA.

[36] Lee, M. H., & Choi, Y. H. (2008). Fault detection of wireless sensor networks. *Computer Communications*, *31*(14), 3469-3475.

[37] Lee, H., Song, J., & Park, D. (2005). Intrusion detection system based on multi-class SVM. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 511-519.