A Analytical Review on The Cyber Security And Intrusion Detection System Using Data Mining Techniques

Aasifa Arabi[1], Qurat Ashraf[2], Yasir Saleem Khan[3], Ambreena Muneer[4], Suhail Javid[5]
[1, 2, 3, 4, 5] Dept of Computer Science Engineering
[1, 2, 3, 4, 5] IqbalInstitute of Technology and Management

Abstract- *Cyber security is of key importance to personal computer users, military, and other organisations. Security has became a major concern and the history of security allows a better understanding of the emergence of security technology. The field of network security is vast and in an evolutionary stage. The range of study includes a brief history back to internet's beginnings and the development which is happening now in network security. In order to understand the research being performed today, background knowledge of the importance of security, types of attacks taking place in the networks. This paper explains the literature study on network security in various domains.*

Keywords- Cyber Security, Cloud Computing, Sensor Networks, Computer to Computer Networks, IOT

I. INTRODUCTION

The world is becoming more interconnected with the advent of the Internet and new networking technology. There is a large amount of personal, commercial, military, and government information on networking infrastructures worldwide. Network security is becoming of great importance because of intellectual property that can be easily acquired through the internet.

Network security starts with authorization, commonly with a username and a password. Network security consists of the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access, modification in system, misuse, or denial of a computer network and network-accessible resources. Basically network security involves the authorization of access to data in a network, which is controlled by the network admin. It has become more important to personal computer users, and organizations. If this authorized, a firewall forces to access policies such as what services are allowed to be accessed for network users. So that to prevent unauthorized access to system, this component may fail to check potentially harmful content such as computer worms or Trojans being transmitted over the network. Anti-virus software or an intrusion detection system (IDS) helps detect the malware. Today anomaly may also monitor the network like wire shark traffic and may be logged for audit purposes and for later on high-level analysis in system. Communication between two hosts using a network may be uses encryption to maintain privacypolicy.

II. NEED OF NETWORKSECURITY

System and network technology is a key technology for a wide variety of applications. Security is crucial to networks and applications. Although, network security is a critical requirement in emerging networks, there is a significant lack of security methods that can be easily implemented.

There exists a "communication gap" between the developers of security technology and developers of networks. Network design is a well-developed process that is based on the Open Systems Interface (OSI) model. The OSI model has several advantages when designing networks. It offers modularity, flexibility, ease-of-use, and standardization of protocols. The protocols of different layers can be easily combined to create stacks which allow modular development. The implementation of individual layers can be changed later without making other adjustments, allowing flexibility in development. In contrast to network design, secure network design is not a well- developed process. There isn't a methodology to manage the complexity of security requirements. Secure network design does not contain the same advantages as network design.

When considering network security, it must be emphasized that the whole network is secure. Network security does not only concern the security in the computers at each end of the communication chain. When transmitting data the communication channel should not be vulnerable to attack. A possible hacker could target the communication channel, obtain the data, decrypt it and re-insert a false message.

Securing the network is just as important as securing the computers and encrypting the message.

When developing a secure network, the following need to be considered:

1. *Access*: authorized users are provided the means to communicate to and from a particular network.
2. *Confidentiality*: Information in the network remains private.
3. *Authentication*: Ensure the users of the network are who they say they are.
4. *Integrity*: Ensure the message has not been modified in transit.
5. *Non-repudiation*: Ensure the user does not refute that he used the network.

An effective network security plan is developed with the understanding of security issues, potential attackers, needed level of security, and factors that make a network vulnerable to attack

III. ATTACK ON NETWORKS

This section describes the basic class of attacks which can be a cause for slow network performance, uncontrolled traffic, viruses etc. Attacks to network from malicious nodes. Attacks can be categories in two: "Passive" when a network intruder intercepts data traveling through the network, and "Active" in which an intruder initiates commands to disrupt the network's normal operation.

A. Active Attacks

Some active attacks are spoofing attack, Wormhole attack, Modification, Denial of services, Sinkhole, and Sybil attack.

1. *Spoofing:* When a malicious node miss-present his identity, so that the sender change the topology.
2. *Modification:* When malicious node performs some modification in the routing route, so that sender sends the message through the long route. This attack cause communication delay occurred between sender and receiver.
3. *Wormhole:* This attack is also called the tunneling attack. In this attack an attacker receives a packet at one point and tunnels it to another malicious node in the network. So that a beginner assumes that he found the shortest path in the network.
4. *Fabrication:* A malicious node generates the false routing message. This means it generate the incorrect information about the route between devices.
5. *Denial of services:* In denial of services attack, malicious node sending the message to the node and consume the bandwidth of the network. The main aim of the malicious node is to be busy the network node. If a message from unauthenticated node will come, then receiver will not receive that message because he is busy and beginner has to wait for the receiver response.
6. *Sinkhole:* Sinkhole is a service attack that prevents the base station from obtaining complete and correct information. In this attack, a node tries to attract the data to it from his all neighbouring node. Selective modification, forwarding or dropping of data can be done by using this attack.
7. *Sybil:* This attack related to the multiple copies of malicious nodes. The Sybil attack can be happen due to malicious node shares its secret key with other malicious nodes. In this way the number of malicious node is increased in the network and the probability of the attack is also increases. If we used the multipath routing, then the possibility of selecting a path malicious node will be increased in the network.

B. Passive Attacks

The names of some passive attacks are traffic analysis, Eavesdropping, and Monitoring.

Traffic analysis: In the traffic analysis attack, an attacker tries to sense the communication path between the sender and receiver. An attacker can found the amount of data which is travel from the route of sender and receiver. There is no modification in data by the traffic analysis.

1. *Eavesdropping:* This is a passive attack, which occurred in the mobile ad-hoc network. The main aim of this attack is to find out some secret or confidential information from communication. This secrete information may be privet or public key of sender or receiver or any secrete data.
2. *Monitoring:* In this attack in which attacker can read the confidential data, but he cannot edit the data or cannot modify the data

IV. LITERATUREREVIEW

Shi-Jinn Horng *et al.,* in [1] designed a new flow for intrusion detection system using Support Vector Machine (SVM) technique. The famous KDD Cup 1999 dataset was used to evaluate the proposed system. Compared with other

intrusion detection systems that are based on the same dataset, this system exhibited better performance in the detection of DoS and Probe attacks, and the best performance in overall accuracy.

Mohammad Wazid in [2] has used hybrid anomaly detection technique with the k-means clustering. WSN are simulated using Optimized Network Engineering Tool (OPNET) simulator and the resultant dataset consists of traffic data with end to end delay data which has been clustered using WEKA 3.6. In this experiment, it has been observed that two types of anomalies namely misdirection and black hole attacks were activated in the network.

Shun-Sheng Wang *et al.,* [3][4] have designed an integrated intrusion detection system using intrusion dataset from UCI repository .The dataset trained well using Back Propagation Neural Network (BPNN) and the output is used as an important parameter in Adaptive Resonance Theory (ART) model to cluster the data. Finally the outputs received from both techniques are compared and the ART model provided the best accuracy rate and overall performance.

Mohit Malik *et al.,* [5] applied the rule based technique for detecting the security attack in WSN. They identified ten important security attack types developed a fuzzy rule based system for calculating the impact of security attacks on the wireless sensor network.

Reda M. Elbasiony *et al.,* [6] proposed a hybrid detection framework using K-means clustering algorithm to detect novel intrusions by clustering the network connections. In this hybrid framework, the anomaly part was improved by replacing the k-means algorithm with the weighted k-means algorithm.

LeventKoc *et al.,* [7] proposed a new technique Hybrid Naïve Bayes (HNB) and excelled in a superior performance in terms of accuracy, error rate and misclassification cost. In early stages the traditional Naïve Bayes model are used but the result produced by HNB is better than traditional Naïve Bayes. The results they have produced indicate that this model significantly improves the accuracy for detecting the denial-of-services (DoS) attacks.

Wenying Fenga *et al.,* [8] introduced a new way of combining algorithm for the better result in detecting intrusions and classified the network activities into normal or abnormal by reducing the misclassification rate. It combined Support Vector Machine method and the Clustering based on Self-Organized Ant Colony Network to take the advantages by avoiding their weaknesses. This Experiments show that CSVAC (Combining Support Vectors with Ant Colony) outperforms better the SVM or CSOACN in terms of both classification rate and run-time efficiency.

Megha Bandgar *et al.,* [9] described a novel approach using Hidden Markov Models (HMM) to detect Internet attacks and described an intrusion detection system for detecting a signature based attack. They have performed single and multiple HMM model for source separation both on IP and port information of source and destination.

Dat Tran *et al.,* [10] proposed Fuzzy Gaussian mixture modeling method for network anomaly detection. It was a mixture of Gaussian distributions used to represent the network data in multi-dimensional feature space. Using fuzzy C-means estimation, Gaussian parameters were estimated and the whole work is carried out with the KDD Cup data set. The proposed method produced here is more effective than the vector quantization method.

Vahid Golmah in [11] developed a hybrid technique using C5.0 and SVM algorithm to evaluate the performance of the hybrid technique with DARPA dataset. The motivation behind this hybrid approach was to improve the accuracy of the intrusion detection system when compared to using individual SVM and C5.0. Due to the mixture of SVM and C5.0, it took less execution time.

Punam Mulak in [12] has used hybrid technique by combining Boundary cutting algorithm and clustering algorithm. The motivation for using this hybrid approach is to improve the accuracy of the intrusion detection system and to provide better result than other clustering.

Venkata Suneetha Takkellapati in [13] proposed a new system with Information Gain (IG) and Triangle Area based KNN algorithm is for selecting more discriminative features. Then the Greedy k-means clustering algorithm was combined with SVM classifier to detect Network attacks. This system achieved a accuracy detection rate and less error rate .All these experiments were conducted in KDD CUP 1999 training dataset.

Vaishali Kosamkar in [14] developed technique of combining C4.5 Decision Tree and Support Vector Machine (SVM) algorithm in order to achieve high accuracy and diminish the false alarm rate. For feature selection stage, the Correlation- Based Feature Selection (CFS) algorithm was used for better accuracy result.

Harmeet Kaurl in [15] designed a model to reduce the delay in the network and to produce an end to end data in good

speed. A simulated WSN using SPEED protocol was used. It was concentrating on two different performance parameters throughput and energy consumption for analysis. BCO (Bee Colony Optimization) algorithm was used to give better results with high throughput and low energy consumption.

H. Oh, I. Doh and K. Chae in [16], the authors proposed a real-time intrusion detection system based on the Self- Organizing Map (SOM); an unsupervised learning technique that is appropriate for anomaly detection in wireless sensor networks. The proposed system was tested using KDD'99 Intrusion Detection Evaluation dataset. The system groups similar connections together based on correlations between features. A connection may be classified as normal or attack. Attacks are classified again based on the type of attack. It took the system 0.5 seconds to decide whether a given input represents a normal behavior or anattack.

N. Ye and X. Li in [17], A data mining algorithm called Clustering and Classification Algorithm Supervised (CCA-S) was developed for intrusion detection in computer networks. The algorithm is used to learn signature patterns of both normal behaviors and attacks. Compared to anomaly detection techniques, the signature recognition techniques always produce true alarms, but not being the capability to detect unknown attacks. The algorithm's scalability and incremental learning were improved performance the decision tree algorithms.

G. Singh, F. Masseglia, C. Fiot, A. Marascu and P. Poncelet in [18], the authors addressed the main drawback of detecting intrusions by means of anomaly (outliers) detection. In their work, they added a new feature to the unknown behaviors before they are considered as attacks, and they claim that the proposed system guarantees a very low ratio of false alarms, making unsupervised clustering for intrusion detection more effective, realistic andfeasible.

K. Faraoun and A. Boukelif in [19], a genetic programming approach for multi-category pattern classification applied to network intrusion detection, proposed to reduce the input patterns dimension towards a better inter-classes discrimination, and achieved through non-linear transformations on the original datasets.

W. Lee, S. Stolfo, P. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop and J. Zhang in [20], a real time data mining based intrusion detection like accuracy, efficiency and usability in intrusion detection in real time environments. It used the artificial anomalies, multiple model and adaptive learning algorithms to address the above issues respectively.

K. Ioannis, T. Dimitriou and F. C. Freiling in [21], a light weight intrusion detection scheme was proposed to identify or detect the effect of attack in WSN by utilizing the concept of collaborative communication methodology. They also formulated the general rules for the WSNtoo.

D. Farid, J. Darmont, N. Harbi, N. Hoa and M. Rahman in [22], the authors addressed the complexity of the intrusion detection datasets, as most of them are complex and contain large number of attributes. Some of these attributes may be redundant or do not have significant contribution for intrusion detection. The aim of this work was to specify effective attributes from the training dataset to build a classifier using data mining algorithms. Experimental results on KDD'99 intrusion detection dataset show that the proposed approach achieves high classification rates and reduces false positives in such environment with limited computational resources.

J. Zhang and M. Zulkernine in [23], the authors focused on the high rate of false positive in intrusion detection associated with an intent of achieving a high rate of false positives in intrusion detection, a modified random forest algorithm was developed, and tested using WEKA tool, testing was conducted on KDD CUP 99 dataset for the above said claim.

M. Tavallaee, E. Bagheri, W. Lu and A. Ghorbani in [24], to overcome the short coming of KDD CUP 99 dataset, a new dataset called NSL-KDD [24] was proposed and presented a detailed statistical analysis model to evaluate the intrusion detectionsystems.

Campose *et al.*, [25] proposed a Database Centric Architecture for Intrusion Detection (DAID) system in Oracle 10g to address the challenges in designing and implementing data mining based intrusion detection systems. DAID offered numerous advantages in terms of scheduling capabilities, alert infrastructure, data analysis tools, security, scalability, andreliability.

K. Prothives and S. Srinoy in [26], an intrusion detection system based on Adaptive Resonance Theory (ART) and Rough Set Theory [38] to detect the known attacks and also new unknown attacks by creating new clusters using ART and RT.

H. Güneş Kayacık, A. Nur Zincir-Heywood and M. I. Heywood in [27], a feature relevance analysis [27] was conducted on KDD CUP 99 to enlist the effects of features in detecting the intrusion in systems.

Amini *et al.*, in [28] introduced an intrusion detection approach based on Adaptive Resonance Theory (ART) and Principal Component Analysis (PCA). The PCA is used for feature selection to reduce the computational complexity and training time of ART. Experimental results show that modifications proposed in this approach improved the speed and accuracy of detection

J. Xiao and H. Song in [29], an intrusion detection system called Unsupervised Neural Net based Intrusion Detector (UNNID) was introduced to provide the facilities for training, testing, and tuning of unsupervised Adaptive Resonance Theory (ART) with neural networks used for intrusion detection.

E. Skoudis in [30], to mention a few of the attacks Smurf attacks, also known as directed broadcast attacks, and are popular form of DoS packet floods. Smurf attacks rely on directed broadcast to create a flood of traffic for a victim. The attacker sends a ping packet to the broadcast address for some network on the Internet that will accept and respond to directed broadcast messages, known as the Smurf amplifier. The attacker uses a spoofed source address of the victim. If there are 30 hosts connected to the Smurf amplifier, the attacker can cause 30 packets to be sent to the victim by sending a single packet to the Smurf amplifier.

K. Labib and V. Rao Vemuri in [31], Neptune attacks can make memory resources too full for a victim by sending a TCP packet requesting to initiate a TCP session. This packet is part of a three-way handshake that is needed to establish a TCP connection between two hosts. The SYN flag on this packet is set to indicate that a new connection is to be established. This packet includes a spoofed source address, such that the victim is not able to finish the handshake but had allocated an amount of system memory for this connection. After sending many of these packets, the victim eventually runs out of memory resources. IPsweep and Portsweep, as their names suggest, sweep through IP addresses and port numbers for a victim network and host respectively looking for open ports that could potentially be used later in an attack.

T. Eldos, M. Khubeb Siddiqui and A. Kanan in [32], author presented a contribution to the network intrusion detection process using Adaptive Resonance Theory (ART1), a type of Artificial Neural Networks (ANN) with binary input unsupervised training. they presented the feature selection using data mining techniques, towards two dimensional dataset reduction that is efficient for the initial and on-going training, and reduce the dataset both vertically and horizontally, numbers of vectors and number of features.

Table I Research Direction In Network Security

S. No.	Security Threats	Security Requirement	Research Direction
1	Denial-of-service (DoS)	Availability	Intrusion detection
2	Unauthenticated or unauthorised access	Key establishment and trust setup	Random key distribution
3	Node capture and compromised node	Resilience to node compromise	Inconsistency detection and node revocation
4	Routing attacks	Secure routing	Secure routing protocols
5	Message modification	Integrity and authenticity	Keyed secure hash function
6	Message disclosure	Confidentiality and privacy	Link/network layer encryption
7	Intrusion and high-level security attacks	Secure group management, intrusion detection,	Intrusion and high-level security attacks

V. CONCLUSION

In this paper, we have evaluated many researchers approach for network security in WSN, IoT, Cloud Computing, WBAN, and Big Data. This article suggests a research area in the domain of security threats for WSN, WBAN, Cloud computing, IoT. In future smart home conditions, there will be multi-modal sensor explications that include the advantages reported. Table I depicts the research direction in network security.

REFERENCES

[1] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui- Lin Lai, Citra Dwi Perkasa, “A novel intrusion detection system based on hierarchical clustering and support vector machines”, *Elsevier Computer Network,* pp.306–313,2010.

[2] Mohammad Wazid, “Hybrid Anomaly Detection using K-Means Clustering in Wireless Sensor Networks”, Center for Security, Theory and Algorithmic Research, pp. 1-17,2014.

[3] Y.-J. Shen and M.-S. Wang, “Broadcast scheduling in wireless sensor networks using fuzzy hopfield neural network,” *Expert Systems with Applications*, Vol. 34, No. 2, pp. 900-907,2008

[4] Y. Wang, M. Martonosi, and L.-S. Peh, “Predicting link quality using supervised learning in wireless sensor networks,” *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 11, No. 3, pp. 71–83,2007

[5] Mohit Malik, Namarta kapoor, Esh naryan, Aman Preet Singh, “Rule Based Technique detecting Security attack for Wireless Sensor network using fuzzy logic”, *International Journal of Advanced Research in Computer Engineering & Technology,* Vol. 1, No. 4,, ISSN: 2278–1323, June2012.

[6] Reda M. Elbasiony, Elsayed A.Sallam, Tarek E. Eltobely,Mahmoud M. Fahmy, “A hybrid network intrusion detection framework based on random forests and weighted k-means” *Ain Shams Engineering Journal*, vol 4,pp.753–762,2013.

[7] Levent Koc, Thomas A. Mazzuchi, Shahram Sarkani, “A network intrusion detection system based on a Hidden

Naïve Bayes multiclass classifier", *Elsevier,* pp.13492–13500,2012.

[8] Wenying Fenga, Qinglei Zhangc, Gongzhu Hud, Jimmy Xiangji Huange, "Mining network data for intrusion detection through combining SVMs with ant colony networks", *Elsevier,* pp. 127-140, 2013.

[9] Megha Bandgar, Komal dhurve, Sneha Jadhav,Vicky Kayastha,Prof. T.J Parvat, "Intrusion Detection System using Hidden Markov Model (HMM)", *IOSR Journal of Computer Engineering (IOSRJCE)* e- ISSN: 2278-0661, p- ISSN: 2278- 8727Vol. 10, No. 3, pp. 66-70, Mar. - Apr.2013.

[10] Dat Tran, Wanli Ma, and Dharmendra Sharma, "Network Anomaly Detection using Fuzzy Gaussian Mixture Models", *International Journal of Future Generation Communication and Networking,* pp.37- 42,2012.

[11] Vahid Golmah, "An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM"**,** *International Journal of Database Theory and Application* Vol.7, No.2, pp. 59-70,2014.

[12] Punam Mulak, Nitin R. Talhar, "Novel Intrusion Detection System Using Hybrid Approach", *International Journal of Advanced Research in Computer Science and Software Engineering,* Vol. 4, No. 11, ISSN: 2277 128X, November2014.

[13] Venkata Suneetha Takkellapati1, G.V.S.N.R.V Prasad, "Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine", *International Journal of Engineering Trends and Technology,* Vol. 3, No.2012.

[14] Vaishali Kosamkar, Sangita S Chaudhari, "Improved Intrusion Detection System using C4.5Decision Tree and Support Vector Machine", *International Journal of Computer Science and Information Technologies,* Vol. 5, No. 2, pp. 1463- 1467,2014.

[15] Harmeet Kaur, Ravneet Kaur, "Crossbreed Routing Protocol for SPEED Terminology in Wireless Sensor Networks", *International Journal of Advance Research in Computer Science and management Studies,* Vol. 2, No. 7, ISSN: 2321-7782, July2014..

[16] H. Oh, I. Doh and K. Chae, "Attack classification based on data mining technique and its application for reliable medical sensor communication", *International Journal of Computer Science and Applications,* Vol. 6, No. 3, pp. 20-32,2009.

[17] N. Ye and X. Li, "A Scalable Clustering Technique for Intrusion Signature Recognition", *Proceedings of 2001 IEEE Workshop on Information Assurance and Security,*2001.

[18] G. Singh, F. Masseglia, C. Fiot, A. Marascu and P. Poncelet, "Data Mining for Intrusion Detection: from Outliers to True Intrusions", The 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'09), Thailand,2009.

[19] K. Faraoun and A. Boukelif, "Genetic Programming Approach for Multi-Category Pattern Classification Applied to Network Intrusions Detection", *The International Arab Journal of Information Technology,* Vol. 4, No. 3,2007.

[20] W. Lee, S. Stolfo, P. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop and J. Zhang, "Real Time Data Mining-based Intrusion Detection", *Proceedings of DISCEX II,* June2001.

[21] K. Ioannis, T. Dimitriou and F. C. Freiling, "Towards Intrusion Detection in Wireless Sensor Networks", *13th European Wireless Conference,* Paris, April2007.

[22] D. Farid, J. Darmont, N. Harbi, N. Hoa and M. Rahman, "Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification", *International Conference on Computer Systems Engineering (ICCSE 09),* Bangkok, Thailand, December2009.

[23] J. Zhang and M. Zulkernine, "Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection", *Symposium on Network Security and Information Assuranc-Proc. of the IEEE International Conference on Communications (ICC),* Istanbul, Turkey, June,2006.

[24] M. Tavallaee, E. Bagheri, W. Lu and A. Ghorbani, "A Detailed Analysis of the KDD'99 CUP Data Set", *The 2nd IEEE Symposium on Computational Intelligence Conference for Security and Defense Applications (CISDA),*2009.

[25] M. Campos and B. Milenova, "Creation and Deployment of Data Mining-Based Intrusion Detection Systems in Oracle Database10g"