ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 5, Issue 4, April 2016

Internal Attack Identification and Inhibition System using Cosine Similarity for Abnormal User Detection

S. Jerina Catherine Joy[1], Dr.S.Singaravelan[2], Mrs. M.Piramu[3]

[1]PG scholar, PSR Engineering College, Sivakasi, Tamil Nadu, India

[2,3]AssistantProfessor, Computer Science and Engineering, P.S.R. Engineering College, Sivakasi, Tamil Nadu, India.

***Abstract*—Network security plays a vital role in the prevention of data transmission network from the unauthorized access and illegal activities performed by the malicious users. Various security mechanisms include encryption and decryption, password protection, firewalls, anti-malware applications, Intrusion Detection Systems etc. Even though, these mechanisms prevents the network from security attacks, they fail to identify the internal attackers. The internal attackers causes equivalent damage as the external attackers. As the internal attackers imitate the authorized users, it is tedious to identify those attackers. In order to address the abovementioned issues, this paper proposes an Internal Attack Identification and Inhibition System (IAIIS) by combining the data mining and forensic techniques. The System Call (SC) of the users at the kernel are analyzed in the proposed IDS.The Cosine Similarity (CS) is applied to compare the profiles of the users to find, whether the user is a normal user or an abnormal user. The experimental results evaluate the proposed system in terms of detection accuracy, decisive rate threshold, response time, abnormal user count, and false alarm rate.**

***Index Terms*— Intrusion detection system, Internal Attack Identification and Inhibition System, Cosine similarity, System call, Hellinger distance, Hash function, Sketch dataset.**

I. INTRODUCTION

The growth of internet and communication led to the wide use of networking and its resources. The geographically dispersed nodes are linked together using a specific topology to transmit the data from one location to another. The data transferred from a source to destination are highly prone to attack. The disclosure of sensitive data leads to the destruction of the system. The primary attacks that influence the data networks are phishing, Denial of Service (DoS), viruses, IP spoofing attacks, worms, etc. The data networks transmit data via routers, in which the security is provided. Here, network security plays a vital role in data security by protecting the network from unauthorized and illegal users. The security can be provided by the construction of the firewall or IDS. The anti-virus software, cryptographic schemes, authentication and identity management are the other types of security mechanisms that can be implemented for data protection in networks. These mechanisms concentrates on the prevention of external attacks, but the attacks caused by the internal intruders also cause serious damage to the system. The dictionary attack is the familiar attack caused by the internal users to hack the data of other users by attempting the login passwords or patterns.

Once they get access to their profile, they may attack the data either in the passive mode or in the active mode. In the passive mode, the attackers just reads the data and uses it for other purposes, whereas the in active mode, the data will be either modified or deleted from the profile. The active attacks are more dangerous, when compared with the passive attacks. The general characteristics of IDS are reliability, transparency, and durable monitoring, less false alarm notification and high performance. IDSs are categorized into two types as follows: misuse-based IDS and anomaly-based IDS. Only known attacks that are available in the database can be detected using the signature based IDS. The anomaly based IDS is capable of detecting anonymous attacks that take place within the network. Even though it can detect the unknown attacks, it bypasses the attack if it is an imitation of the normal traffic. Fig 1 shows the implementation of internal IDS in the network.

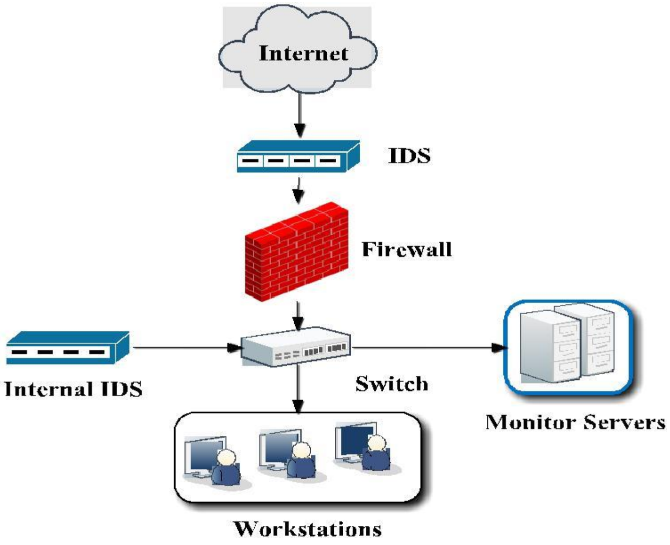


Fig 1. Implementation of Internal IDS in a network

The existing IDS undergoes the following challenges:

- They fail to detect the internal attacks caused by the internal intruders of the system.
- The System Calls (SCs) at the kernel of the Operating System (OS) are not considered for intrusion detection.
- Even though the systems detect the intrusions, they never isolated the attackers for attack prevention.
- Simultaneous detection of intrusions and virus is not possible.

In order to address these issues, an Internal Attack Identification and Inhibition System (IAIIS) is designed by integrating the data mining and forensic techniques. The user logs collected by monitoring the user activities at the kernel level. The System Call (SC) of the individual users are analyzed and compared using Cosine Similarity (CS) to decide whether it is a normal or abnormal activity. The main intent of this paper is to detect the internal intruders with increased accuracy and reduced response time.

The remaining sections of this paper are organized as follows: Section II reviewed the existing approaches of IDSs to detect the malware and attacks. Section III provides an overall description of theInternal Attack Identification and Inhibition System (IAIIS). Section IV shows the performance analysis of the proposed system. Section V presents the conclusion and future work of the paper.

II. Related Work

This section reviews the traditional approaches regarding the IDS for providing network security. *Chen, et al.* [1] suggested prototype based Autonomic Security management (ASM) strategy for estimation, detection and identification of security attacks. The strategy performed proper planning for providing protection to the networking system. Firstly, the system and network parameters were collected by the sensors.

The collected information were sent to the forecasters and the Intrusion detection Systems (IDS). A multi-objective controller was assigned for the selection of appropriate protection strategy based on the signatures of the attacks. The experimental study showed that there was an improved QoS and reduced overhead.*Lu, et al.* [2]proposed a behavioralsignature generation system, namely, DiffSig to restrict the attackers from snooping the flow of traffic information. The suggested system handled the dependencies in a more efficient way based on which the new malwares can be detected.*Chadli, et al* [3] designed a multi-agent based IDS architecture to improve the performance and security of the network. The proposed architecture is a combination of cluster-based and cooperation-based hierarchical model. A knowledge base, which contained the global ontology data, was used to improve security. When compared to the existing IDSs, the hierarchical architecture performed better in terms of performance and security.

Saeed, et al [4] proposed a dynamic model checking based multi-agent architecture to detect intrusions in an efficient manner. The suggested architecture had a group of cooperative collaborating agents placed in the hierarchy. The communication rules were specified for interaction between the nodes. Temporal logic rules were applied in the agents to combine the obtained results. Dynamic model checking algorithm was used to check whether the specific formulae was satisfied or not. When compared with the previous works, the proposed architecture provided promising results. *Houque, et al* [5] presented an IDS using Genetic Algorithm (GA) to detect various intrusions in the network. Darwininan's principle was used to optimize the population of candidate solutions. The set of chromosomes obtained in the pre-calculation step was used to find the attack type. When compared to the existing IDSs, the proposed IDS resulted in high detection rate.

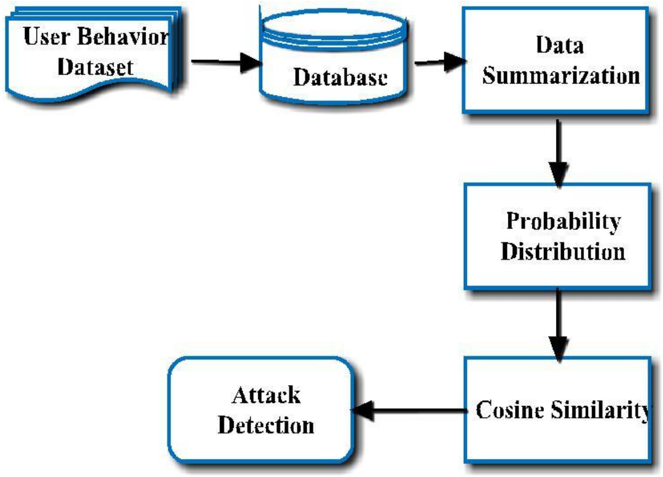
Koc, et al. [6] proposed a Hidden Naïve Bayes (HNB) multiclass classifier based network IDS to detect the network events such as normal and attack events. This model was a combination of PKI discretization and INTERACT feature selection methods. The detecting accuracy DOS attacks were improved through HNB classifier. When compared with SVM, the performance of HNB based IDS was superior in terms of accuracy, classification cost, and error rate. *Chen, et al.* [7] proposed Network Intrusion Detection and Countermeasure mechanism (NICE) to detect the vulnerabilities in multi-phase distributed environment. The OpenFlow network programming was leveraged for monitoring and controlling the distributed programs over the virtual switches. The attacks detection was improved and the consequences of the attacks were mitigated via the switches used in the proposed IDS. *Trivedi, et al.* [8] presented a Reputation Based Intrusion Detection System for Mobile adhoc networks (RISM) to strengthen the defense of attacks. The concepts such as redemption, monitoring system, path management and fading were applied in the proposed design. A routing protocol was integrated with an IDS to observe the reliability of nodes along with their trust values. When compared with DSR, the suggested RISM outperformed in terms of packet delivery ratio and routing overhead.

Karthick, et al. [9] proposed an adaptive network IDS by integrating HMM based traffic model and a probabilistic classifier to detect the unknown attacks and to minimize the misclassification rate. The anomalies in the potential traffic were prevented and the IP addresses for the potential attacks were eliminated. The results showed that the hybrid approach achieved good performance. *Hassan, et al.*[10] designed an IDS using Genetic Algorithm (GA) and fuzzy logic to identify the intrusion activities in computer networks. The main intent of the proposed IDS was to identify the attacks of the unauthorized users in the network traffic. The primary attacks such as probes, DoS, U2R and R2L were detected. The fuzzy confusion matrix was used to measure the fitness function of the chromosomes. Even though, it was cost effective, it generated more false alarm rates. *Toulouse, et al.* [11] proposed a consensus based network IDS to identify the malicious behaviors of the network using a naïve Bayes classifier. An iterative average consensus protocol was used to compute the probability values of the attacks. The results showed that the communication cost is low, whereas the convergence speed is high. *Panda, et al* [12] designed an IDS using hybrid intelligence approach by combining classifiers to achieve the best detection rate. Two-class classification strategy and ten-fold cross validation method was used for final classification. The meta learning strategies were used along with grading and END to enhance the performance of the proposed system.

Selvaraj, et al [13] proposed a novel idea using honey pot technique and packet data analysis for intrusion detection in both network and anomaly based systems. The system was trained with a sample of malware or deep analysis of packet inspection. Honey pot had a large amount of energy and a strong field of security, which made the proposed system better than the existing IDSs. *Uddin, et al* [14] proposed a signature-based multi-layer Intrusion Detection System (IDS) to detect the threats with high success rate. It automatically created multiple small databases in a dynamic fashion to update signatures of new malware at regular intervals. The challenges of signature-based systems were signature creation, storage, maintenance, updation, and avoiding traffic flooding. A complementary payload-based anomaly detection system was used to detect the newly emerging attacks. Multiple IDS were deployed in different layers. The multi-layer approach optimized the detection rate and also detected uncommon attacks. *Sharma, et al* [15] presented a signature-based IDS using an identity-based scheme to achieve network protection from worm-hole attacks. The suggested scheme resisted the process of certificate distribution between the nodes in the network. The computational overhead was decreased in the identity-based signature scheme due to the elimination of signature distribution.

III. Proposed Method

This section presents the overview of the proposed Internal Attack Identification and Inhibition System (IAIIS). The main objective of this paper is to detect the internal attackers with high accuracy and less response time. Fig 2 describes the overall flow of the IAIIS.



The proposed IAIIS consist of four stages as follows:

- Data loading
- Dataset sketching
- Cosine similarity calculation
- Attack detection and Prevention

A. Data Loading

In the data loading process, the raw data is preprocessed before loading the required data. Preprocessing includes elimination of the inconsistent and incorrect data to improve the results of the proposed system. The unwanted, noisy and repeated data are removed to decrease the amount of data to be loaded. The dataset is made superior through clustering and aggregation techniques of data mining. In the proposed IDS, the user patterns are given as input in the form of user log files. The SC patterns of the user are mined from the log files using the pattern mining concepts. The dataset is loaded with user ID, process ID and system calls of the user. The user data are gathered in Comma Separated Value (CSV) but it is crucial to use the CSV formatted data for further processing. The dataset requires text format for easy processing and hence the CSV format is converted to text file format in the preprocessing stage. This conversion removes the unwanted data present in the CSV files.

B. Dataset Sketching

The reduced dataset that contains the SC patterns, user ID and process ID is summarized by applying hash functions. For each and every attribute available in the dataset, the probability of the attacks are computed using sketch data distribution. The internal attacks are identified by the analysis of similarity scores for the SC patterns and the user activities. The size of the dataset is condensed to attain a summarized sketch dataset. The data aggregation techniques are integrated along with the hash function for dataset reduction.

C. Cosine Similarity Distance

Cosine similarity is used to measure the degree of resemblance between the user profiles and the actual dataset.The degree of similarity between any two traffic patterns are measured by the cosine similarity. According to the degree of resemblance the normal and abnormal users are classified in the proposed system. The cosine similarity can be represented by the following formula.

$$, = \frac{A \times B}{A \times B} \qquad (1)$$

Where,

A - Patterns of the user profile
B - Patterns in the actual dataset

The probability values that attain low degree of similarity is considered the abnormal user profiles, whereas the profiles with high cosine similarity is the normal user profile. The actual dataset contains the traffic patterns obtained, while the network is accessed by a normal user. So, degree of cosine similarity is high for normal behaviors, as there is a match in the actual dataset.

Steps of the proposed IAIIS

Step 1: The raw dataset containing user profiles and the SC patterns is preprocessed.

Step 2: In the preprocessing stage, dataset in CSV format is converted to text file format.

Step 3: The dataset is summarized to get the sketch dataset.

Step 4: For each and every entry in the sketch dataset, the probability values are calculated.

Step 5: A table is created to store the calculated probability values.

Step 5: The Cosine similarity was measured between the probability values using the formula in the equation 1.

Step 6: The normal and abnormal users are distinguished based on the degree of resemblance.

Step 7: The entries that obtain the low degree of resemblance is considered as abnormal users.

Step 8: The SC monitor forward the abnormal user profiles to the SC filter for segregation.

Step 9: The attacks are inhibited using the isolation process of the SC filter.

D. *Attack Detection and prevention*

Attack is a harmful damage caused by the malicious users or software that try to compromise the security mechanisms of the network. The intruders may steal hardware, software or sensitive data, run the Trojan or other malwares in the system to corrupt or modify the data and try to degrade the network performance by the over utilization of the resources. So, it is essential to detect and prevent the attacks to protect the network from malware disasters. Attacks are classified as external and internal attacks, in which most of the security mechanisms concentrate on external attacks. The intruders in the external attack snoop the network to collect the sensitive information. The external attacks are further categorized as remote and local external attacks. War dialing, DoS, brute force attacks are the few examples of remote external attacks. These attacks targets the servers located at the remote locations that offer services to the users of the organization. The DoS attack dumps the server with a large number of requests to prevent the services offered by the server. The war dialing is similar to DoS attack that dumps the private exchange by making unwanted calls to keep the exchange always busy. The shared access to some of the resources in an organization leads to local external attacks. The implementation of firewall around the network mitigates the external attacks.

The internal attacks are caused due to the insiders of an organization. Foot printing, privilege escalation, port scanning, and backdoor installation are some of the examples of internal attacks. Foot printing creates the traces of the network to identify the ports that are not secured and open. It also collects the configuration information of the network, which is the basis for other activities. Once foot prints are gathered, the steps are taken to bypass the network via the open ports. In privilege escalation, the attacker aims to access the privileges of an administrator through the low level accounts. Backdoors are installed in the network to enter the network, whenever the attacker needs access. The internal attacks can be addressed using an IDS. This paper focuses on the internal attacks of the network.In the proposed system, the attacks are identified using the Cosine similarity. During the execution of shell commands, the SC patterns generated at the kernel are maintained in the log files of the individual users. To predict whether the user is a normal user or an abnormal user, SC patterns are compared with the user behavior. The internal intruders are segregated by the SC filters on the reception of a notification from the SC monitor. If a user tries the authentication code of another user, it can be detected using the usage patterns of the users.

IV. Performance Analysis

This section presents the performance analysis of the IAIIS. The proposed system is compared with the IIDDS that uses Hellinger Distance (HD) for attack detection. The performance of the proposed system is evaluated in terms of

- Detection accuracy
- Decisive rate threshold
- Response time
- Abnormal user count
- False alarm rate

A. *Detection Accuracy*

Detection accuracy is defined as the ratio of the number of triggered alerts to the total number of alerts in the trace.

$$= \frac{}{} \qquad (2)$$

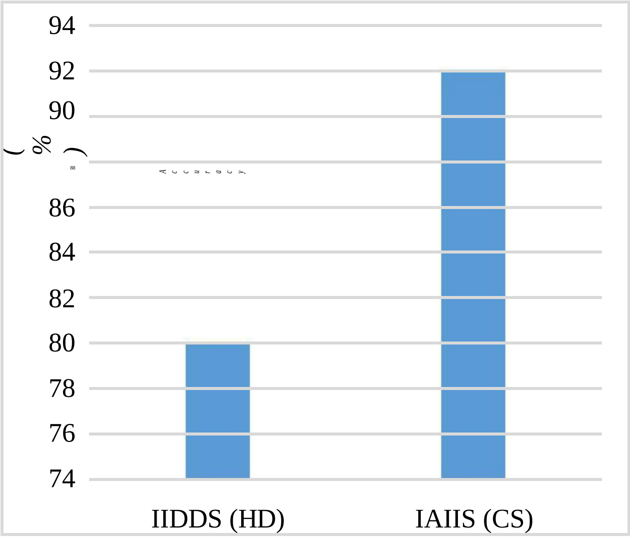


Fig 3. Detection accuracy of IIDDS (HD) and IAIIS (CS)

The comparison between detection accuracy of the existing and the proposed system is shown in the graph plotted in Fig 3. The detection accuracy of IIDDS using HD is 80% and the detection accuracy of IAIIS using CS is 92%. When compared to IIDDS, there is a 12% increase in the detection accuracy of IAIIS. Thus, the proposed IDS achieved higher detection accuracy, when compared to the existing system.

B. *Decisive Rate Threshold*

Decisive rate threshold can be defined as the similarity score between the current user profile and the profile of other users. The decisive rate thresholds of the HD based IIDDS and the CS based IAIIS is compared in Fig 4. The decisive rate threshold of the IIDDS is 0.025, whereas the decisive rate threshold of the IIDDS is 0.038. The decisive rate threshold of the proposed system is increased due to the use of cosine similarity for pattern recognition. When compared to the HD based IIDDS, the decisive rate threshold of the proposed IAIIS is high.

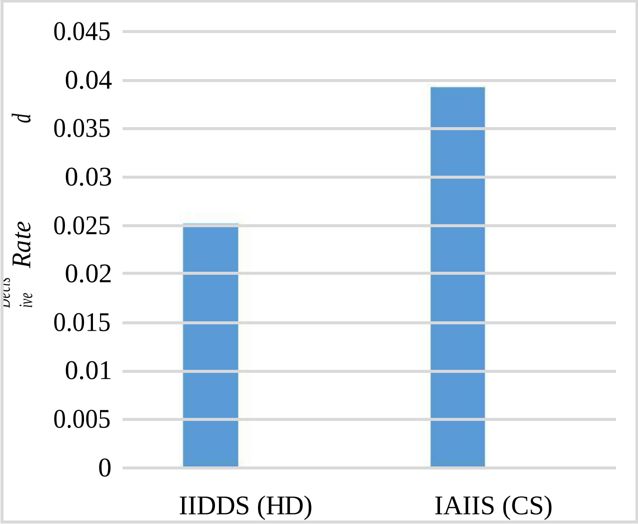


Fig 4. Decisive rate threshold of IIDDS (HD) and IAIIS (CS)

C. *Response Time*

Response time can be defined as the time taken to produce the trigger, when an attack is detected. Fig 5 illustrates the analysis of the response time of both the IIDDS and IAIIS. In HD based IIDDS, the response time is provided in 25 ms, whereas, the CS based IAIIS responds in 18 ms. Hence, the proposed IAIIS resulted in lesser response time than the HD based IIDDS.

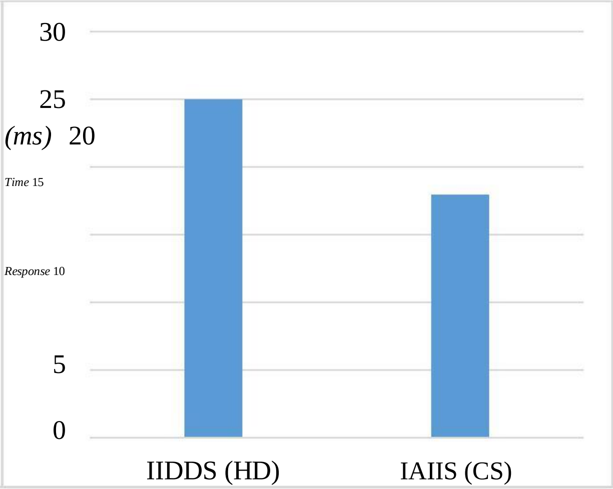


Fig 5. Response time of IIDDS (HD) and IAIIS (CS)

D. *Abnormal User Count*

Abnormal user count is the detection rate of the number of attackers or malicious users in the system. The graph in fig 6 depicts the comparison of abnormal user count in both the existing and proposed system. The abnormal user count in IIDDS is 50 and the abnormal user count of IAIIS is 25. Therefore, the abnormal user count of the proposed system is reduced by 50% on comparison with the existing system. The SC filter and the cosine similarity reduces the abnormal users in the proposed IAIIS.

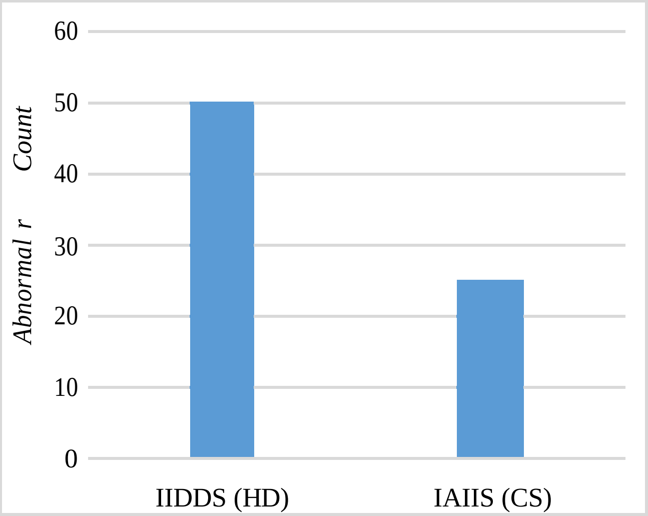


Fig 6. Abnormal user count in IIDDS (HD) and IAIIS (CS)

E. *False Alarm Rate*

False alarm rate is defined as the ratio of the number of wrongly triggered alerts to the number of total alerts.

$$= \frac{\quad}{\quad} \qquad (3)$$

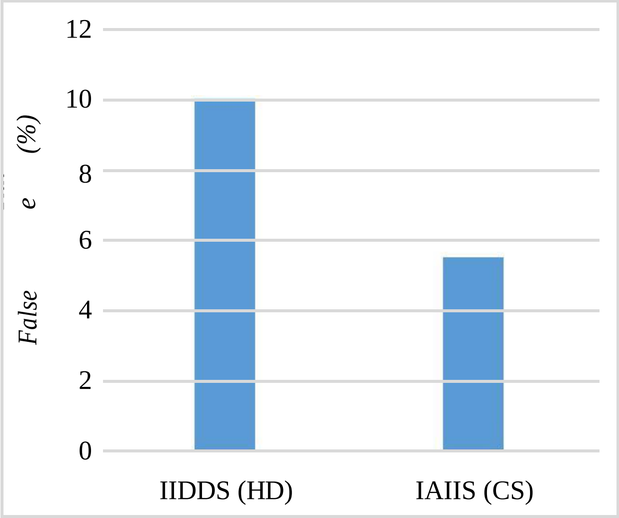


Fig 7. False alarm rate generated by IIDDS (HD) and IAIIS (CS)

Fig 7 describes the graph plotted against the false alarm rates of the proposed and existing systems. The IIDDS produced a false alarm rate of 5.5%, but the IAIIS generated the false alarm of 10%. When compared to the existing system, the false alarm rate is reduced in the proposed system. Hence, it is clearly understood that the proposed HD based IIDDS outperformed the existing CS based IAIIS in terms of detection accuracy, decisive rate threshold, response time, abnormal user count, and false alarm rate.

V. Conclusion and Future Work

This paper proposed an Internal Attacker Identification and Inhibition System (IAIIS) to defend the internal intruders from the network. The user logs are gathered from the kernel level of the OS. The logs contained user ID, process ID and the SC patterns is preprocessed and converted in to the sketch dataset. The cosine similarity is applied to improve the detection accuracy of the proposed IDS. The internal attackers are isolated using the SC filters according to SC pattern of individual users. The performance results of the proposed system is compared with the HD based IIDDS in terms of detection accuracy, decisive rate threshold, response time, abnormal user count and false alarm rate. When compared with IIDDS, the accuracy and the decisive rate threshold is high. The abnormal users are reduced in the proposed system by the use of SC filters. The system responds with small time duration and minimal false alarm rate. Hence, the overall results showed that the proposed IAIIS outperformed the HD based IIDDS.

REFERENCES

[1] Q. Chen, S. Abdelwahed, and A. Erradi, "A model-based approach to self-protection in computing system," in *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference*, 2013, p. 16.

[2] H. Lu, B. Zhao, X. Wang, and J. Su, "DiffSig: resource differentiation based malware behavioral concise signature generation," in *Information and Communication Technology*, ed: Springer, 2013, pp. 271-284.

[3] S. Chadli, M. Saber, M. Emharraf, and A. Ziyyat, "A new model of IDS architecture based on multi-agent systems for MANET," in *Complex Systems (WCCS), 2014 Second World Conference on*, 2014, pp. 252-258.

[4] I. A. Saeed and A. Selamat, "Multi-agent Architecture with Dynamic Model Checking for Malware Detection," *LabuanSchool of Informatics Science*, vol. 15, p. 47, 2013.

[5] M. S. Hoque, M. Mukit, M. Bikas, and A. Naser, "An implementation of intrusion detection system using genetic algorithm," *arXiv preprint arXiv:1204.1336*, 2012.

[6] L. Koc, T. A. Mazzuchi, and S. Sarkani, "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier," *Expert Systems with Applications*, vol. 39, pp. 13492-13500, 2012.

[7] C.J. Chung, P. Khatkar, T. Xing, J. Lee, and D. Huang, "NICE: Network intrusion detection and countermeasure selection in virtual network systems," *Dependable and Secure Computing, IEEE Transactions on*, vol. 10, pp. 198-211, 2013.

[8] A. K. Trivedi, R. Kapoor, R. Arora, S. Sanyal, and S. Sanyal, "RISM--Reputation Based Intrusion Detection System for Mobile Ad hoc Networks," *arXiv preprint arXiv:1307.7833*, 2013.

[9] R. R. Karthick, V. P. Hattiwale, and B. Ravindran, "Adaptive network intrusion detection system using a hybrid approach," in *Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on*, 2012, pp. 1-7.

[10] M. M. M. Hassan, "Network intrusion detection system using genetic algorithm and fuzzy logic," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, 2013.

[11] M. Toulouse, B. Q. Minh, and P. Curtis, "A consensus based network intrusion detection system," in *IT Convergence and Security (ICITCS), 2015 5th International Conference on*, 2015, pp. 1-6.

[12] M. Panda, A. Abraham, and M. R. Patra, "A hybrid intelligent approach for network intrusion detection," *Procedia Engineering*, vol. 30, pp. 1-9, 2012.

[13] R. Selvaraj, V. M. Kuthadi, and T. Marwala, "Honey Pot: A Major Technique for Intrusion Detection," in *Proceedings of the Second International Conference on Computer and Communication Technologies*, 2016, pp. 73-82.

[14] M. Uddin, A. A. Rahman, N. Uddin, J. Memon, R. A. Alsaqour, and S. Kazi, "Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents," *IJ Network Security*, vol. 15, pp. 97-105, 2013.

[15] D. Sharma, V. Kumar, and R. Kumar, "Prevention of Wormhole Attack Using Identity Based Signature Scheme in MANET," in *Computational Intelligence in Data Mining—Volume 2*, ed: Springer, 2016, pp. 475-485.

Author's Information:

S.Jerina Catherine Joy received the BE degree in Francis Xavier Engineering College, Anna university,Tamilnadu,India and PG scholar in PSR Engineering college, Sivakasi , Tamilnadu, India.

S.Singaravelan received the PhD, degree in Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli, in 2016. He received the M.E., degree in Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli, in the

year 2007. He received the B.E., degree in Electronics and Communication Engineering from Francis Xavier Engineering College, Tirunelveli, 2004.Currently he is a Assistant Professor in Computer Science and Engineering Department of P.S.R. Engineering College, Sivakasi, Tamil Nadu, India. He has published more than 10 International Journals including Scopus Index in the areas of Digital Image Processing, Content Based Image Retrieval, Object Recognition, Pattern Regonition.He is a Life member in ISTE.

M. Piramu received the BE degree in A.K.C.E. Krishnan Kovil-M.K.University, Tamil Nadu, India and Master's degree in MepcoSchlenk Engineering College, Sivakasi, Tamil Nadu, India. Currently she is a Professor in Computer Science and Engineering Department of P.S.R. Engineering College, Sivakasi, Tamil Nadu, India.