Prediction of Cyber Security Attacks through Analysis of Windows Event ID Using Ants Colony Optimization & ANN - EFFPA

S D Jadhav
Research Scholar,
Shri Guru Gobind Singhji Institute of Engineering & Technology, Nanded, India.

Dr. B R Bombade
Associate Professor,
Shri Guru Gobind Singhji Institute of Engineering & Technology, Nanded, India.

Abstract *— Cyber security threat provides a new challenge & opportunities for researchers, system administrators and incident response experts and it is also a Cyber security threat to government & private organizations. This paper proposed a novel method of using Windows Event ID log analysis to identify and predict the cyber security attack. In this model we are proposing to use two stage analysis. At initial level we propose to first apply the Ant Colony Optimization (ACO) technique to optimize the Windows Event ID logs as per our experiment requirements. Once Windows Event ID logs are optimized, then at second stage, we propose to apply the Artificial Neural Network based Error Feed Forward Propagation Network (ANN-EFFPA) algorithm. The ANN-EFFPA is first trained by providing sample set of Windows Event ID logs. Such sample Windows Event ID logs are based on the previously occurred cyber security attack patterns. Different to different test conditions and datasets, helped the system to get trained effectively and helped proposed novel system to get evolved and matured. Our novel experiment has proven that, the result generated is 92% accurate for correct identification and prediction of cyber security threats through Windows Event ID log analysis.*

Keyword *— Cyber Security Attacks, Windows Event ID logs, Genetic Algorithm, Ant Colony Optimization, Artificial Neural Network, Error Feed Forward Propagation Analysis, Windows Event Identifiers, Log Analysis, Anomalies, Incident of Compromise.*

I. Introduction

With the internet growth, cyber security threat of also started to looming over the netizens. Initially everyone was not aware of the cyber security threats. At present, state and private industry are fighting day and night against the menace of cyber-crime across the globe. The menace of cyber-economic crime has crossed loss to industry worth of billion-dollars in various countries like USA, UK, China, Canada, Japan, Australia, India & other such countries [1, 2, 3 and 4]. Cyber attackers are always trying to find out the loopholes in the existing security systems and exploit them. For every such actions of attacker, the Windows operating system generates Windows Event ID logs. Windows Event ID log file provides crucial information about who accessed the system, when the system was accessed, what files were opened/altered/deleted, what changes were made to any file/system configuration, list and details of Transactions, errors, flags, day-date-time of all events, etc. valuable information is available in Windows Event ID log file and based on same one can trace back the cyber-attacks. But, at present various available log processing systems have issues such as, poor log optimization algorithms, poor process for identification of malicious events in given logs, and no proper facility to have real-time analysis of logs, etc. Therefore, it is necessary to have real time analysis of the Windows Event ID logs for prevention of cyber-crime.

Optimization is a process which is nothing but finding out the best set of parameters that are necessary to solve the problem statement, which also mean in other words to discard the least essential parameters, minimize the computation load and compilation time with increase in accuracy of result. The Genetic Algorithm (GA) based Ant Colony Optimization algorithm has been previously applied on various engineering problems [5, 6]. It has demonstrated its usefulness in complex optimization process. Few researchers have used the same on various computer science problems, right from data base issues to network route selection. In our proposed research method, we proposed to apply the Ant Colony Optimization algorithm over the Windows Event ID logs.

A. Introduction to Ant Colony Optimization (ACO)

It is widely known that when food source is located, then an ant moves back from food source location to its colony and during that process it secrets and lays down a special chemical called *Pheromone*, to trace back its path. Ant pheromones is made up of a variety of chemical compounds, including pyrazines, pyrrolidine-based alkaloids, and monoterpenes. There exist few major Pheromone categories, for example, when danger is close to the ants, they spray *Alarm Pheromones* that generates alarm against specific danger, *Escape Pheromones* are used to declare the call to escape when enemy invades their colony, when food is located by an ant, it heads back to its colony and while moving back the ant sprays *Food Trail Pheromones* and other such pheromones exist [7].

When any ant find outs a *Food Pheromone* trail, it triggers a social response among worker ants to reach the food. Once they reach and collect the food, then while moving back, it is necessary that worker ant's must find out the shortest path from food source to its colony. Hence worker ants spray their Pheromone liquid while heading back towards the colony and hopes to find out the shortest path apart from the previously laid down pheromone trail paths. Hence multiple paths are available for any ants to reach the food source. But, in order to save the time, energy and efforts, it is necessary to select only a single path which has minimum distance. Therefore, in order to select the shortest path from all the available paths, ant identifies such path whose *Food Pheromone* trail intensity is high. This is because, in this process, the surface temperature or heat, wind, rain, and human activities sometime destroys the Pheromone acid trail. Considering only atmospheric conditions as disruptor, one can say that only shortest path will have strong intensity of *Food Pheromones Trail.* This is because, greater the distance then lowers the chance that the Food Pheromones track will survive from heat, wind, dust, humidity, rain, etc. atmospheric conditions and vice-a-versa. This is well depicted in Figure 1(a), where ants intelligently discard all the longest paths where Food Pheromone trail is found to be weaker and selects the shortest path through its intelligence [8].

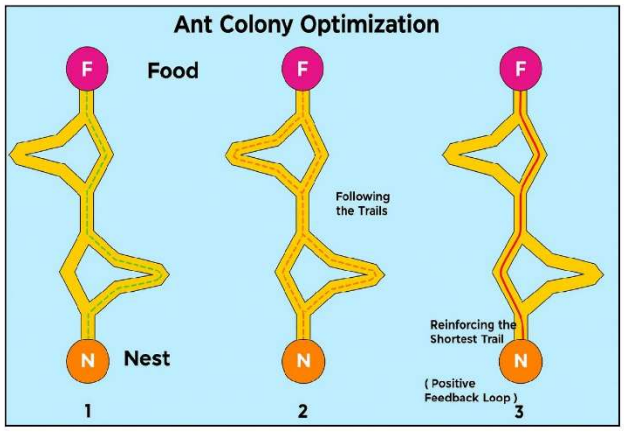

Figure 1(a)

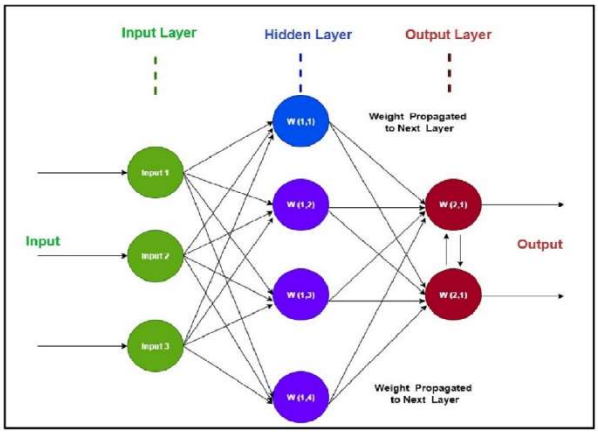

Figure 1(b)

Figure 1 (a): ACO Technique for finding the shortest path using Food Trail Pheromones.
1 (b): Formal Depiction of ANN Algorithm based EFFPA.

B. Introduction to ANN - Error Feed Forward Propagation Algorithm (ANN-EFFPA)

The Artificial Neural Network - Error Feed Forward Propagation Algorithm (ANN-EFFPA) consist of single layer or multiple hidden layers. Its formal depiction is shown in Figure 1(b). The EFFPA is a method that involves the computation of the error gradient at each layer of the network. The error gradient is then used to update the respective node of neural network. EFFPA propagation is widely used in AI and deep neural networks due to its efficiency, accuracy and faster machine learning approach.

C. Introduction to Windows Event Identifiers (ID)

There could be more than 500+ event log IDs in Windows Operation System and each Event ID is unique and has its own purpose and definition. There is no fix number of ID's as each year new ID's are added by Microsoft. Table 1 provides a glimpse of few of such Windows Event ID's. The Windows Event Id's are normally stored at location ***C:\Windows\System32\winevt\Logs.***

Table 1. Glimpse of few Windows Event ID's and their definition.

Windows Event ID	Details	Windows Event ID	Details
529	Logon Failure - Unknown user name or bad password	627	Change Password Attempt
531	Logon Failure - Account currently disabled	629	User Account Disabled
576	Special privileges assigned to new logon	630	User Account Deleted
611	Removing Trusted Domain	644	User Account Locked Out
612	Audit Policy Change	676	Authentication Ticket Request Failed
617	Kerberos Policy Changed	4625	Failed Logon attempts
621	System Security Access Granted	4740	User Account Lockout
622	System Security Access Removed	4723	An attempt was made to change an account's password
625	User Account Type Changed	1100	The event logging service has shut down

II. Materials and Methods

Proposed Novel Model (ACO and ANN-EFFPA)

A. Dataset Details: The standard dataset represents 58 consecutive days of de-identified event data collected from five sources from their internal computer network. It includes Windows based authentication events of individual desktop systems, Active Directory Domain Controllers, Process start and halt events, suspicious events, failed authentication events, etc. [9, 10 and 11]. The Dataset is of 12-GB size and contains 1,648,275,307 events, total 12,425 users, 17,684 computers and 62,974 processes and is available in chunks and in CSV, JSON, etc. formats. Similarly, we have created our own local lab dataset, which consist of 10 Million Microsoft Windows Events ID's collected from individual desktop systems of our research institute. Both the collectively dataset has required contents and are of size that are enough to conduct our proposed experimentation. Both the datasets need to be optimized first and then to be used for further experimentation.

B. Setting up the Overall Generic Process Flow: The overall generic process flow is divided into two main parts, Part A and Part B. This is well depicted in figure 2.

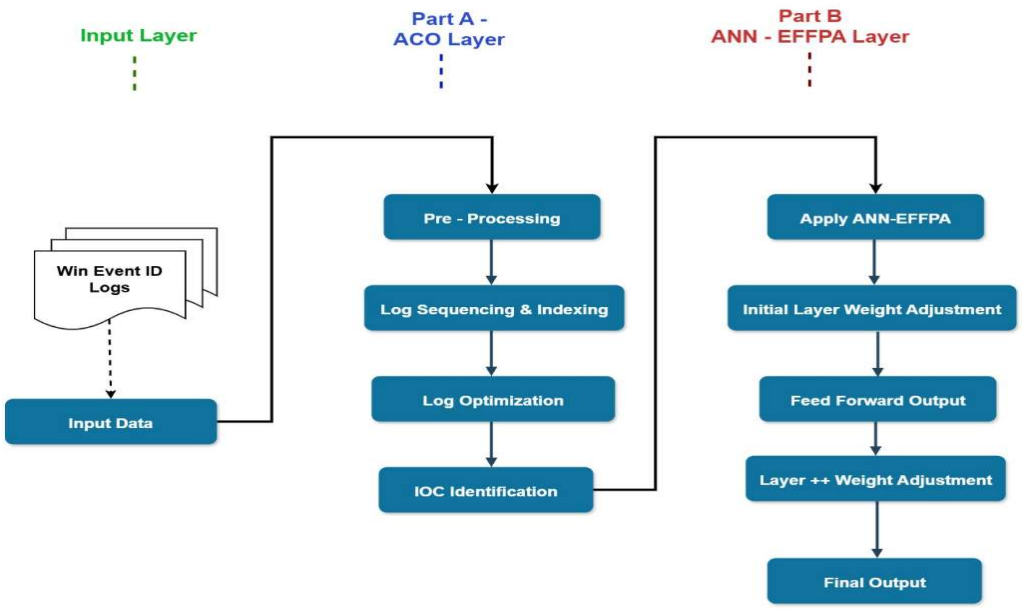

Fig. 2. Proposed Novel Method to Predict Cyber Security Attacks by analysis of Windows Event ID using ACO and ANN-EFFPA.

Part A – It is initial process and consist use of Genetic Algorithm based Ant Colony Optimization and its primary function is to do pre-processing of logs (cleansing – remove empty fields, duplicate entries, etc.), Sequencing and indexing of logs in ascending date-wise manner, optimize the logs as per given parameters. Post optimization, find out those Windows Event ID's which can lead to possible anomalies and Incident of Compromise (IOC). This process of finding out the possible Windows Event ID's is done by using the previous cyber security attack scenarios and sorting out those Windows Event ID which are noted as suspicious one. The Output of Part A later becomes input to Part B.

Part B – It is lateral process which consist of Artificial Neural Network (ANN) based Error Feed Forward Propagation Algorithm (EFFPA). The final output is forwarded for weight adjustment of immediate next layer and training the system accordingly. Improvements are done in order to minimize or manage the False Acceptance Rate (FAR) and False Rejection Rates (FRR).

Please note - If we compare behavior of ants (Pheromone based chemical trails and number of alternative paths it creates from such chemicals) with the various windows event ID's, then one can consider that various such event ID's are nothing but the various number of available paths. Whereas the shortest path out of various number of available paths will be those Windows Event ID's that helps to identify the anomalies and Incident of Compromises occurred on various computers, attacked by the attacker during cyber-attack.

III. Results and Discussions

For experimentation, we have used a server class machine of our local research lab and also used the online Google Co-lab facility provided by Google Inc. Figure 3 (a) provides the glimpse of very first Graphical User Interface (GUI) of proposed novel system. The GUI consist of 6 tabs and one display window. The six tabs facilitate user for; a) To upload the Windows Event ID logs, b) Application of proposed novel model over logs, c) Finding all anomalies and Incident of compromise (IOC) and d) Visualize the output generated using graphs e) Exit Tab to exit the program. If you observe figure 3 (a), then we can find out that user has successfully uploaded a Windows Event ID log file and its details (File name & total number of logs) are shown in its display window along with the message of successful uploading the log file.

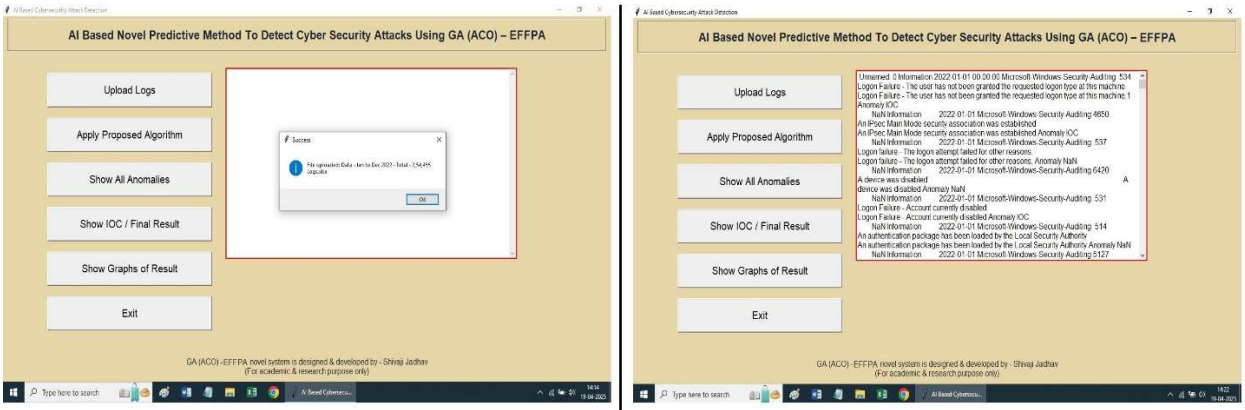

Figure 3(a) Figure 3(b)

Figure 3 (a): GUI Window showcasing pop-up message highlighting successful uploading of logs.
3 (b): GUI showcasing the list of all the Anomalies found in given log dataset.

Since the logs are uploaded successfully, now apply the proposed algorithm through pressing the second tab "Apply Proposed Algorithm". Post successful compilation of all Windows Event ID logs as per proposed novel method, the system highlights success message and then user need to press the third tab "Show All Anomalies". This will highlight all the anomalies in the provided dataset of Windows Event ID logs, as shown in figure 3 (b).

Kindly note that, not all listed anomalies are considered as a threat or potential future Incident of Compromise. This is because sometime anomalies can be suspicious at first but only after close observation, one can conclude that the respective anomaly is a real cyber security threat or not. The condition for an anomaly to become IOC varies per cyber-attack incident. Each time, we have to set a threshold under which an anomaly can be considered as IOC. This depends on the global threat vector parameters that evolves constantly and with respect to updates and upgrades in technology platforms including hardware, software and firm-wares. This is because cyber criminals always take advantages of this changing technology platforms and the loopholes exist in it. The moment when one loopholes is fixed by the manufacturer / developer, the next moment cybercriminal find outs and exploits another loophole. Therefore, based on same, the Ant colony optimization algorithm defines weather the respective anomaly is IOC or not. Once user click on the fourth tab "Show IOC/ Final Result" then system generates the final output. This is shown in figure 4. Please note, with each Windows Event ID log file processing, the generated results are internally feed forwarded to ANN-EFFPA model. The weight of respective node is adjusted based on the input provided. This helps to train the proposed model.

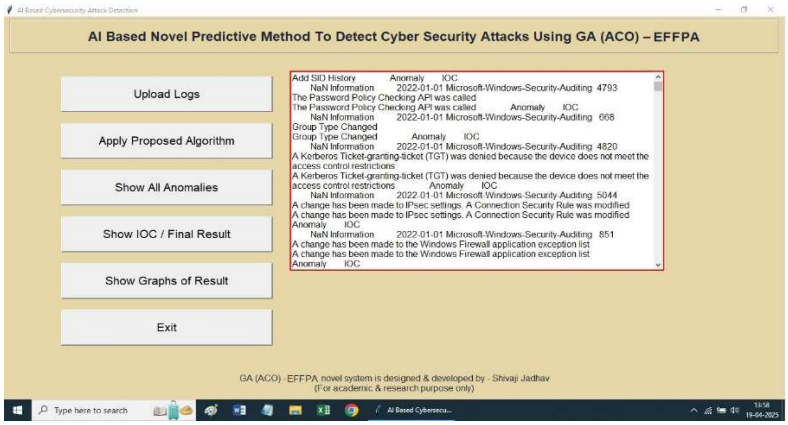

Figure 4. GUI showing list of all the Incident of Compromise (IOC's) found in given log dataset.

Further, by using the tab “Show Graphs of Result” will plot the graphs that gives the overall statistics of total number of Windows Event ID logs marked as anomaly and total number of Windows Event ID logs marked as Incident of compromise in month-wise manner. This again helped us to understand that not all anomalies will be treated as IOC. This is well depicted in figure 5(a) where month-wise statistics of various anomalies and IOC’s for two consecutive years 2022 and 2023 respectively are provided. Similarly, figure 5(b) provides details of month-wise statistics of various anomalies and IOC’s found for consecutive year 2024 respectively.

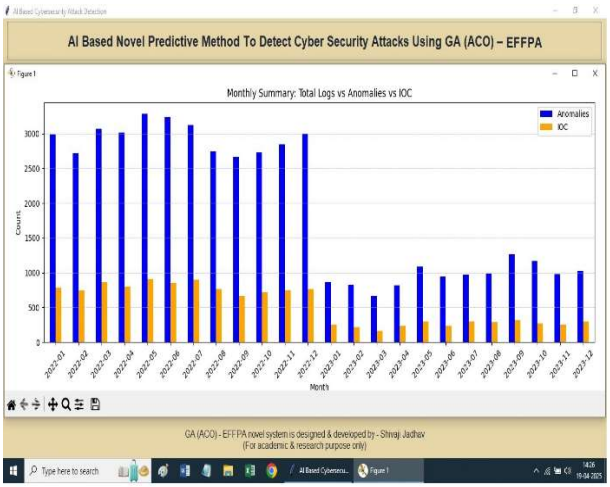

Figure 5(a)

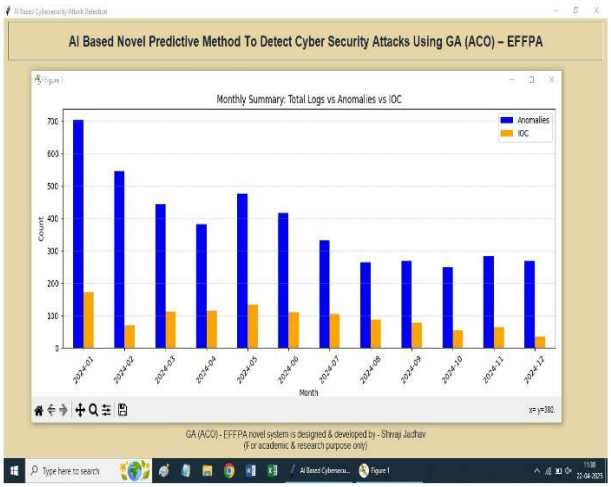

Figure 5(b)

Figure 5(a). Graphs with Anomalies and IOC’s found in given log dataset (for year 2022 & 2023).
5(b). Graphs with Anomalies and IOC’s found in given log dataset (for year 2024).

We have done numerous such experimentation with various standard dataset like Kaggle listed Los Alamos National Laboratory (LANL), our own local research lab generated log datasets and other such available datasets. Based on all figures and tables, it is observed that, the proposed novel method and implemented system correctly identifies and marks each such anomaly and IOC found in given log datasets. We can call this marking as Pheromone based marking of GA-ACO, and the correctness of such markings will help to adjust the weight of each node of the ANN-EFFPA model. This will be treated as a learning cycle for the ANN-EFFPA model.

Limitations - Sometime, the algorithm may identify and mark any simple log event as anomaly or IOC incorrectly, which is false identification of a simple log event. Such actions can be measured as False Acceptance Rate (FAR) and it is well represented in figure 6(a). It consists of 10 iterations for log compilation and with each Iteration Cycle (IC) the system gets trained and matured. The false acceptance rate at first iteration cycle is approx. 18.08% and with each successive iteration cycle the false acceptance rates falls up to 8.20% till the tenth iteration cycle. This proves that the algorithm is getting enough training through each iteration cycles to correct itself.

Similarly, the accuracy of the proposed system of GA-ACO & ANN-EFFPA gets slowly improved in stepwise manner. From 1st iteration cycle (IC) till up to the 10th iteration cycle the accuracy of the system has reached approximately up to approx. 92%. This is compared with the accuracy of only GA-ACO model execution with no ANN-EFFPA module. The comparison is well depicted in the graph shown in figure 6(b).

Please note that, the primary focus of our research was to compile system logs in our proposed novel GA (ACO) & ANN-EFFPA model. In this paper, we have presented our experimentation using the Windows Event ID’s logs generated by Windows Operating System. But, we have also tuned our system and tried this experiment with few other machine based logs that have different formats and parameters. This include but not limited to IDS, IPS, Firewall, and others.

Sr. No.	Number of Logs	False Acceptance Rate (FAR)
1	10,000	18.08%
2	1,000,00	18.02%
3	5,000,00	17.05%
4	10,000,00	16.00%
5	40,000,00	12.35%
6	80,000,00	11.88%
7	100,000,00	11.50%
8	10,00,00,000	10.40%
9	1,64,82,75,307	10.35%
10	3,50,40,55,208	8.20%

Figure 6(a)

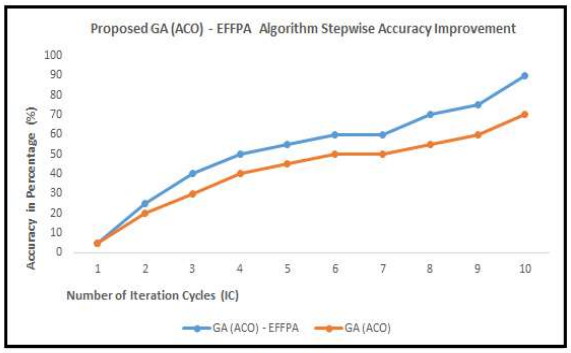

Figure 6(b)

Figure 6(a). Number of Logs and Proposed Systems False Acceptance Rate (FAR)
6(b). Graph showing step-wise Improvement in accuracy (92%) of proposed novel system

IV. Conclusion

Identification of potential cyber threat is a global challenge. Major Cyber security system providers are working day and night to secure digital assets. Organization spends billions of dollar on such cyber security tools across the globe but still cyber-attack occurs. This is because cyber-criminal always tries to find the loopholes in existing system and exploits it. Therefore, it is need of the time to work on such research problem statement that is threating to the world at present. In this paper we presented a novel method for identification and prediction of cyber security attack by using Windows Event ID logs. In proposed novel method we used the combination of Genetic Algorithm based Ant Colony Optimization technique and ANN based Error Feed Forward Propagation Algorithm to solve the problem statement. Our proposed novel system initially takes Windows Event ID logs as input and using GA-ACO it tries to discard all such logs which are not directly related to security incident events. The Ant colony optimization algorithm first narrow down the size of given log data by its unique optimization method. The Windows Event ID's that are marked as suspicious in previous cyber-attack scenarios helps the ant colony optimization model to tune up its optimization parameters.

Once logs are optimized, our system then marks them and provides it as input to ANN-EFFPA algorithm. The Error Feedforward algorithm adjust its node weight based on the identification of anomalies and IOC's. Further, it calculates and quantifies the error magnitude and uses it as like a learning curve. This helps the ANN-EFFPA algorithm to improve itself through each learning cycle. The proposed novel system was designed and developed in Python programming language and tested against various online available standard datasets and our own local research lab generated log datasets. It has proved through our experimentation that the proposed novel system is highly accurate and provides 92% accuracy for correct identification of anomalies and IOC in given Windows Event ID log dataset.

References

[1] *Vergara Cobos, Estefania and Cakir, Selcen. 2024. A Review of the Economic Costs of Cyber Incidents. Washington, DC: World Bank.*
[2] *Mihail Antonescu, Ramona Birău, Financial and Non-financial Implications of Cybercrimes in Emerging Countries, Procedia Economics and Finance, Volume 32, 2015, Pages 618-621, ISSN 2212-5671, https://doi.org/10.1016/S2212-5671(15)01440-9.*
[3] *Sunny A, A study on financial cyber-crimes, trends, patterns, and its effects in the economy. Allied Academics Addiction & Criminology - Addiction & Criminology (2024) Volume 7, Issue 1. DOI: 10.35841/aara-7.1.186.*
[4] *National Crime Record Bureau (NCRB) Annual Report on Crimes in India 2022. Volume I. Ministry of Home Affairs, Government of India.*
[5] *K. F. Man, K. S. Tang and S. Kwong, Genetic algorithms: concepts and applications in engineering design, in IEEE Transactions on Industrial Electronics, vol. 43, no. 5, pp. 519-534, Oct. 1996, doi: 10.1109/41.538609.*

[6] *B. Chandra Mohan et. all., A survey: Ant Colony Optimization based recent research and implementation on several engineering domain, Science Direct Elsevier Journal of Expert Systems with Applications, Volume 39, Issue 4, March 2012, Pages 4618-4627. https://doi.org/10.1016/j.eswa.2011.09.076.*
[7] *M. Dorigo, M. Birattari and T. Stutzle, Ant colony optimization, in IEEE Computational Intelligence Magazine, vol. 1, no. 4, pp. 28-39, Nov. 2006, doi: 10.1109/MCI.2006.329691.*
[8] *Anand Nayyar et. all., Ant Colony Optimization – Computational Swarm Intelligence technique, 3 rd Interational Conference on Computing for Sustiniable Global Development , 16-18 March 2016, pp 392-398.*
[9] *LANL Dataset availaible on https://csr.lanl.gov/data/2017/*
[10] *LANL Dataset availaible on https://csr.lanl.gov/data/cyber1/*
[11] *Kaggle Datas at https://www.kaggle.com/discussions/general/335189*