EasyChair Preprint

№ 15109

AI-Driven APT Detection Framework: Early Threat Identification Using ML

Hidayat Ur Rehman, Zunera Jalil and Safa Fahim

September 27, 2024

AI-Driven APT Detection Framework: Early Threat Identification Using ML

Hidayat ur Rehman, Zunera Jalil, Safa Fahim
Department of Cyber Security, Air University, Islamabad, Pakistan
Hidayatkhan013@gmail.com, zunera.jalil@au.edu.pk, safafahim52@gmail.com

***Abstract*—An increasing number of hacktivists, state-sponsored hackers, cybercriminals, cyber terrorists, cyber spies, and cyberwarfare warfighters are attacking the systems. A balance between real-time cyberattack detection, cyber threat intelligence, and, most importantly, cyber early warning capability is needed for a successful cyber security strategy. Cyber threats are tough and complex to describe since it is challenging to pinpoint the origin of the attack. The motivation driving them, or even to forecast how the attack will play out in real-time. The challenge of drawing boundaries between national or international, public or private objectives makes it more difficult to identify cyber threats. The fight to counteract cyber threats is dynamic and ever more difficult because they are worldwide in scope and entail quick technological advancements. In this study, we focused on the cyber-kill chain, proposed a universal/generic cyber-kill chain model, and analyzed various Advanced Persistent Threat(APT) cyber-kill chain steps/concepts. We focused on the detection of Advanced Persistent Threats by using different machine-learning models like XGBoost, Random Forest, Decision tree, Adatboost, and K-Nearest Neigbor in real-time and achieved an accuracy of 99.95% by using the model of XGBoost.**

I. Introduction

Cyberattacks have occurred, and they have undergone significant evolution over time, and the transition from worms and viruses to malware and botnets. A new kind of attack known as "Advanced Persistent Threat (APT)" has emerged in recent years. Originally used to refer to cyberattacks against military institutions, the word "APT" has now expanded and is no longer exclusive to the military[1]. Nowadays, Smartphones are used by every individual from all around the world. The widespread proliferation of these electronic devices on the internet is the primary cause of the upsurge in cybercrimes. The issue of cyber security has been tricky and hard for the past few years. In comparison, cooperative security solutions, and cyberthreats have quickly advanced. Furthermore, access to knowledge availability has become vital in today's competitive environment, due to it has become an essential commodity. There is now more competition and complexity in the interaction between the system's security and attacks of cyber. The sophistication, gravity, and scope of these cyberattacks are increasing. In the past, unsuspecting hackers would employ software to steal money or steal identities from individuals. After the attack, the effects of these attacks were instantly apparent. But now with the tremendous boost of IT infrastructure, the rules have suddenly altered. The new usage models such as virtualization, cloud computing, and enhanced mobility, have prompted the traditional barriers among enterprises and security to break down or dissolve. This has rendered the environment surroundings more attractive to hackers. The most notable element of the risk landscape in this situation is the emergence of long-running, highly focused worldwide espionage sabotage campaigns by secret agencies. The different countries and extremist groups massively finance these secret agencies in order to launch schemes and campaigns of attack against specific targets. Attacks, this sophisticated and destructive are known as Advanced Persistent Threats (APTs).

As we have discussed in in-depth detail APT detection and its various cause factors. We all know how hazardous it is and to determine the solution for resolving the issue is aided by identifying the characteristics of it, such as its intricate nature, stealth, and persistence. Most APT attacks consist of multiple steps, each of which gives the additional resources, assistance, and expertise to infiltrate the organization. It is crucial to be aware that APT attackers generally do not give up or refuse to back unless they execute their goals; they may receive backing from other nations and organizations that have the capacity with resources to carry out their game plan. The incident response crew can track down the APT more Swiffer and effectively by linking these signs and killing the phases of the chain. Several studies have been done on it and addressed the specific aspects of structure framework, but they didn't cover the overall picture of a real attack.

In this research, we cover the whole aspects to meet the research objectives and then implement them to achieve the goal of the research. In the following, we have discussed some major research motivations in some points.

- Growing sophistication of Advanced Persistent Threats (APTs)
- Limitations of traditional APT detection approaches
- Need for early identification and proactive mitigation of APTs.
- Leveraging the power of AI and machine learning for accurate and efficient detection

This research will offer a thorough framework for modeling APT Attacks using a Cyber Kill Chain(CKC). Along with analyzing and outlining attacks against computer systems, this approach additionally assists in figuring out the characteristics of the adversary. The Kill-Chain attack model's output from the attack threat model can be used to determine the attack's phase. The incident response team can respond appropriately to these APTs by using the framework to determine the

attacker's goals, intent, and techniques. The framework can also map the many connections, linkages, and processes. To achieve the preceding goals, the following has to be fulfilled:

- To investigate and analyze APTs and Cyber Kill Chains (CKCs) that involve a wide range of various subjects and topics, including APTs, Cyber Adversaries, Cyber Indicators and the Indicator Life Cycle, Kill Chain, and Security Information.
- To Create and build a new framework.

For further description, we have divided the research into the following sections. In section I, we have introduced our topic in descriptive form with the research goals and objectives. In section II describe some background statistics of the APT attacks in the past. In Section III provides a literature review on the detection of anomalies. The description of the dataset, with a detailed description of models, is described in section IV and provides the experimentation environment and comparison between the models. And finally, section IV explains the conclusion and future work.

II. Statistics of Attack

The global cybersecurity landscape has faced rising threats in the past few years. Cybercriminals have taken advantage of the pandemic of the misaligned network as firms moved to remote settings. These attacks have spiked by 358% in 2020 over 2019[2]. The number of cyber-attacks has been expanded by 125% on a worldwide basis from 2020 to 2023. The threats accomplished by these attacks to people and business organizations continued. 493.33 million attacks were identified all around the world in 2022. The most common attack from these is phishing which is approximately 3.4 billion. In 2022, the average cost of a data breach worldwide was $4.35 million. The average price of security breaches brought on by lost or compromised credentials was $4.50 million. With an average cost of $10.10 million for data breaches in 2022, the healthcare sector has been the most expensive for breaches for 12 years running [2], [3]. In 2023, attacks have been an unprecedented year. Global attack frequency continuously increased day by day. 95.41% increase year to year and 11.2% expansion over Q2 as you can be seen in the figure [1].

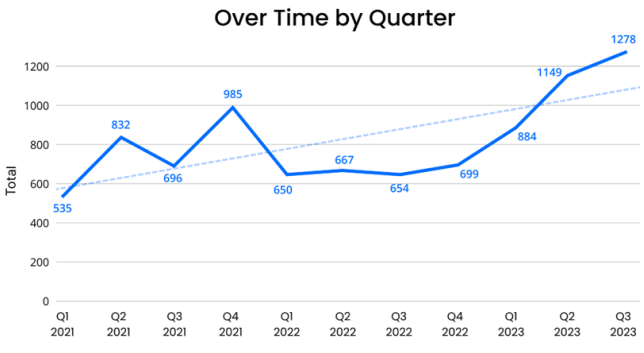

Fig. 1. Quarter on Quarter(QoQ) Statistics 2021 to 2023[4]

III. Literature Review

Malware attacks have progressively risen to the level of Worldwide industrial cyber espionage over the last decade. The literature review approach involved investigating the most significant cyberspace attacks in the past. Below we have explained the whole literature in the mannered form for a better understanding of attacks.

The first hierarchy based on a Cyber Kill chain (CKC) depending on banking Trojan features was proposed by Kiwi et al. [5] in 2017. This hierarchy can be used for learning more about risk mitigation and identification techniques. However, this structure must be extended to include all additional families. In this research, the researcher examined the real-world dataset gathered by the major banking organizations in the UK of 127 banking Trojans 2014 from December to 2016 January. Traditional methods of monitoring and mitigating advanced persistent threat (APT) attacks have shortcomings. These methods are often ineffective because they are based on signatures or rules that can be easily evaded by attackers. Machine learning detection methods are proposed as a solution to these shortcomings. The authors propose a novel method to detect APT attacks in IoT contexts, using prior knowledge input (PKI) and unsupervised learning. The PKI and a Progressive PKI model can group features and simplify the training process, resulting in better performance with fewer features. The tests demonstrate that the PKI and Progressive PKI models enhance the APT detection precision significantly under an IoT dataset. They have used the SCVIC-APT-2021 dataset for the implementation of their model. PKI is a machine learning model that uses unsupervised clustering to obtain prior knowledge about the data. This prior knowledge is then incorporated into a supervised model to improve the model's performance. The authors of the study reported that PKI achieved a best macro average F1 score of 81.37%, which is 10.47% higher than the baseline results. This suggests that PKI is a promising approach for improving the performance of machine learning models [6].

A trained machine learning model can monitor network traffic in real-time with high accuracy and raise an early alert before data exfiltration. However, conventional machine learning methods are not adequate to efficiently detect APT attacks due to the lack of data on APT attacks. Here are some additional details about the shortcomings of traditional methods and the challenges of using machine learning to detect APT attacks: Traditional methods of monitoring and mitigating APT attacks are often ineffective because they are based on signatures or rules that can be easily evaded by attackers. For example, an attacker can simply change the signature of their malware to avoid detection. Machine learning methods can be more effective at detecting APT attacks because they can learn to identify patterns in data that are indicative of an attack. However, machine learning methods require a large amount of data to train, and there is limited data available on APT attacks, which makes it difficult to train effective machine learning models. Despite the challenges, machine learning is a promising approach for detecting APT attacks. As more data on APT attacks becomes available, machine-learning models

will become more effective at detecting these attacks [7].
In [8], Hasan et al. aimed to establish an effective identification model for advanced persistent threat (APT) attacks to prevent and reduce their impact. Machine learning has the potential to detect and predict cyber security threats, including APT. This study used several boosting-based machine learning methods to predict various types of APTs that are consistent in the cyber security domain. Additionally, Explainable Artificial Intelligence (XAI) was used to provide actionable insights to domain stakeholders and practitioners in this domain. The results, particularly XG Boost with a weighted F1 score of 0.97 and Shapley Additive explanations (SHAP)-based explanation, demonstrate that boosting methods and machine learning models paired with XAI are indeed promising in addressing cybersecurity-related dataset problems. This can be extrapolated to new avenues of challenging research by effectively deploying boosting-based XAI models. Knowledge-based models can provide neural networks with a set of rules or knowledge that can be used to make predictions. This can help to reduce the training time of neural networks, especially for large and complex networks Knowledge-based models can also help to improve the performance of neural networks by providing them with a better understanding of the problem that they are trying to solve. Overall, knowledge-based models can be a valuable tool for improving performance and reducing the complexity of neural networks [9].
The SCVIC-APT-2021 dataset was created in a laboratory setting and covers five stages of advanced persistent threat (APT): initial compromise, pivoting, lateral movement, reconnaissance, and data exfiltration. The F1 score was used to evaluate performance, as it takes false positives and negatives into account. The highest macro average F1 score for the SCVIC-APT-2021 dataset is 81.37%, which is 10.47% higher than the previous best result [10]. The approach in [11] is a promising new direction for APT attack detection. While the limitation of this approach is that doesn't include other network traffic components, such as HTTP, TLS, and Flow. We believe that this will further improve the accuracy of our approach. APT attacks are a major challenge for information security systems. The proposed approach can detect APT domains and IPs with high accuracy. The proposed approach is based on the analysis and evaluation of network traffic components using machine learning. The proposed approach can be used to improve the security of information systems.
Unsupervised machine learning techniques can effectively identify advanced cyberattacks that target network infrastructures. It highlights that these methods remain successful even when encountering new attack patterns that were not seen during training and validation. Re-training is unnecessary in such cases. To enhance this approach, it is suggested to develop methods for adapting the detection threshold. Based on the obtained results and observations, future work will focus on automating threshold selection and exploring practical applications. Additionally, the research will involve testing simulation environments and creating extensive datasets to evaluate the detection of advanced persistent threat (APT) attacks [12].
In [13], the researcher proposed the framework which is integreted on the big data by the machine learning models i;e Deep neural network system, Random Forest and Support Vector Machine.The data undergo on the specific process using Resilient Distributed Datasets and one hot encoding methodology. Spiltting dataset into train and test in the spark platform. From the experimentation, the researcher achieve the initial phase accuarcy of 95% to 98%.

IV. Proposed Framework

The aim and main focus of this research is to establish a multistage framework solution that will minimize and limit the destruction of cyber-attacks. With the aim of this, we sketch and create the model according to its requirement for the achievement of our proposal aim. We split our proposed approach into five different phases which we are written below:

- Data Selection
- Data Analysis
- Preprocessing and Feature Extraction
- Implemented Methodologies
- Performance Criteria of Evaluation

In our proposed research, we have created real-time APT Detection systems which are comprised of two major parts. We have utilized different steps for the first one which is for training our model and it is used for detection purposes. The other one is for the deployment of our model in which we have also utilized different steps for training such as packet capturing, data filtrations, and data feature extraction, and then applied the machine learning approach which predicts the log files for the objective of assessment. Figure [2] shows the whole framework that we have used in our research.

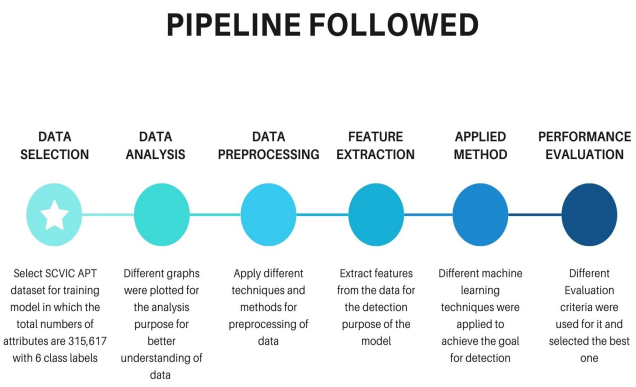

Fig. 2. Steps following for the implementation

A. Dataset Description

Based on the research analysis performed in [10], the dataset SCVIC-APT-2021 [14] is one of the most recent benchmark datasets for the identification of advanced persistent Threats (APT) in traffic through the network. The dataset comprises 84 attributes with 315,617 rows of the dataset. Six class labels serve as the target labels, in accordance with the description of

the dataset. These include data exfiltration, initial compromise, lateral movement, regular traffic, reconnaissance, and pivoting. They formed the foundation for their choice of common attack techniques by adhering to the worldwide knowledge base of adversary tactics and approaches.

B. Data Filtration

Selecting a certain amount of portion of your data set and utilizing it for viewing or analysis is referred to as data filtration. The procedure of filtering can frequently be (though but not always) transient; just a little part of the entire dataset is put into use in the computation. It is fundamental for determining important data, disposing of unnecessary data, and enhancing the quality of information data.
Using filtration can help with:

- Examine the findings for the specific time frame.
- Analyze the outcome for certain areas of interest.
- Exclude inaccurate or “bad” observations from an investigation.
- Develop and validate statistical models.

To choose the situation you want to include in your evaluation, first need to establish a filtering rule or logic. Data ”subsetting” or ”drill-down” are other terms for filtering. he use of filtering and provides an example of a filtered data set. In order to recognize and stop potentially dangerous data or traffic, filtering is an essential part of network and data security. This preserves a network’s integrity and aids in the prevention of cyberattacks.

C. Data Preprocessing

Data preprocessing is important for enhancing the quality of data to improve the results of the system. For applying machine learning and deep learning approaches, the analysis of data is significantly important. For this component, we have investigated the samples of missing data. In the Preprocessing step, we performed integration and aggregation of the testing and training data sets so that the number of counts of seven classes is below.

D. Feature Extraction

Feature extraction plays a pivotal role in classification and detection, and it helps in the facilitation of input to differentiate the distinctive features for an effective process of classification. Feature extraction employs identifying the patterns and boosts the accuracy of the classification significantly. In addition, it is essentially used for reduction in dimensionality input to eradicate extraneous and unwanted information, which makes possible faster and more effective classification in APT detection. Therefore, feature extraction methods acquire considerably in a few years to accommodate the classification demand of APT. It should be highlighted that feature extraction is an essential module that looks and carries out. It makes the utilization of feature engineering [15] in which we must look first to identify the APT attack’s features in the pertinent data collection in terms of their kind. According to features of SCVIC-APT-2021[16], the datatypes object, float64, and int64.

1) Feature Extraction Techniques: We have extracted features by utilizing two techniques i.e., Pearson correlation and the second one is Principal Component Analysis (PCA). As we know, because of the simplicity and capacity to help determine the level of a certain degree of correlation between the input and output variables, Pearson correlation has been extensively used for variable selection. In addition, variables with significant variations impacting the output variable were determined using Principal component analysis. On the other hand, non-linear modeling approaches employ linear forms of variable selection.
In [17], Principal Component Analysis(PCA) and Pearson correlation(PCC) are linear techniques, therefore despite the fact that they are frequently employed effectively in input parameter identification, their applicability in nonlinear is dubious. We have fetched all the columns and found the correlation. In Figure [3], we can see the Pearson correlation heatmap after applying Principal Component Analysis by eliminating columns. The elimination criteria were that we took only those that features have an 80% correlation or less or we can say that we took the relative features whose mask is less than 0.8. We have dropped all the remaining features from the feature sets and the remaining 46 in which the top feature is shown in Figure [4] which we have considered in our study.

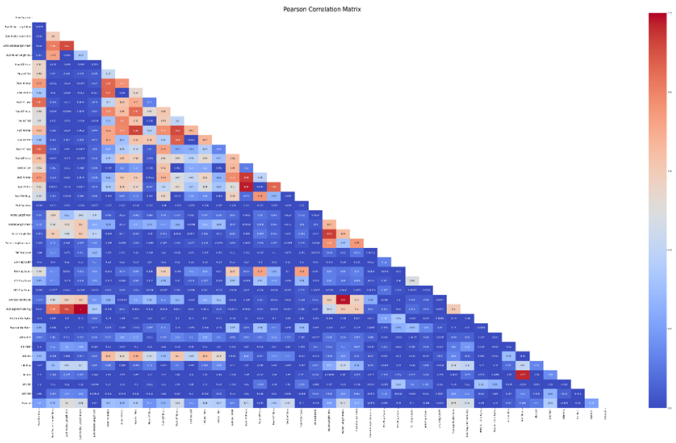

Fig. 3. Pearson Correlation Matrix After Applying Principal Component Analysis

2) Feature Score: The significance of each feature (variable) in relation to the model’s prediction is indicated in the feature score. In a nutshell, it establishes the level or degree of the value that a particular variable has for an existing model or prediction. In general, we use a numeric number known as the feature score that represents the relevance of each aspect; the higher the score, the more important the feature is. The feature score has a lot of benefits such as the relationship between independent variables (features) and dependent variables (targets) can be ascertained. Variable significant scores would allow us to identify and eliminate elements that aren’t relevant. The model may run more quickly or even perform better if the number of useless variables is decreased. Furthermore, a popular tool for ML model interpretability is feature importance. It is feasible to deduce

why the ML model produces specific predictions from the scores and how characteristics might be changed to alter the model's predictions.

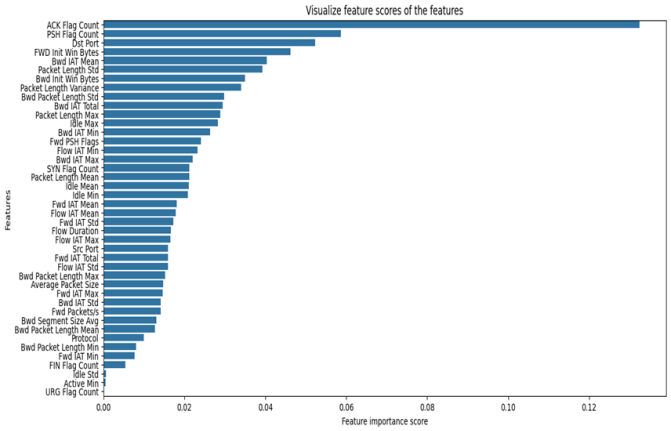

Fig. 4. Attributes Feature Score Visualization

E. Classification using different ML models

For the classification of APT, we have utilized different models of machine learning like XgBoost, Random Forest, Decision Tree, etc.

1) XgBoost: XGBoost was built mainly with gradient-boosted trees for speed and performance. It is a method of machine boosting or applying boosting to machines that was initially carried out by Tianqi Chen [18] and then by other developers. This tool is a part of the DMLC, or Distributed Machine Learning Community.

2) Random Forest: The random forest classifier is a well-known technique used in machine learning. It belongs to a group of methods known as ensemble methods, which combine multiple models to make predictions. The idea behind the random forest is to construct a set of deciding trees, forming what we refer to as a "forest". Each tree is trained independently to ensure accurate classification.

3) Decision Tree: A decision tree generates a tree-like representation of the patterns present between records of data that have been collected in the past. The new record has been categorized by the classification module, which additionally acts as a predictor for the appropriate type of value[19]. The evaluation of decision tree analysis recommends a number of techniques, including the process of pruning, stop rule, and classification standard. The many decision trees are created based on how they are combined. In order to swiftly and reliably create decision trees, a variety of algorithms are being explored; as a result, newly developed algorithms are being announced.

4) AdaBoost: The AdaBoost technique is ideally suited for situations in the real world. In simple words, it is related to the real scenarios that exist in the world. Since it requires the basic classifier's accurate identification rate to be marginally higher than a random guess, as compared to knowing in advance the lower bound of weak learning's predictive accuracy [20].

5) KNN: An adaptation of the instance-based learner algorithm, which makes use of training examples, or instances, is k-nearest neighbors. Learning entails adding new instances to the repertoire. The objective is to categorize an unidentified instance according to its similarity index—which is frequently a distance—about the training instances. The new sample is then assigned a predicted class label. The number of neighbor instances that must be compared is indicated by the variation of k in k-NN, and the class label is determined by the majority class elements' vote.

6) LSTM: The LSTM network was created to recognize long-term dependencies and relationships between data elements within a sequence. LSTMs are designed to eliminate some of the difficulties of conventional RNNs, like a gradient problem that gets weaker with every step of learning, making it impossible to use them in cases where we need to remember some information for a while. Using memory cells and gating mechanisms, this issue is solved by the unique architecture.

7) ANN: An approach of supervised machine learning technique that simulates how the brain processes information. It is composed of connected neurons of the network arranged into input, output, and additional hidden layers. By employing a sigmoid function to modify the triggering weights fed to the neurons and then adjusting through backpropagation, the process can be accomplished by learning by example [21], [22].

8) AutoEncoder: Autoencoders are a technique used in unsupervised learning. The basic idea of an autoencoder is to reconstruct the input data at the output level. They consist of two main components: encoder and decoder. The encoder takes the input data and maps it to a low-dimensional representation called a latent space. The decoder then takes this representation and attempts to reconstruct the original input data.

F. Evaluation Parameters

We have used four evaluation performance matrices i.e.; precision, recall, accuracy, and F-1 score.

1) Precision: Precision is known as the ratio of correctly identified attack connection records to total records located. Equation illustrates the precision.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

2) Recall: It determines the percentage of attack connection records that are correctly classified in relation to the total amount of attack connection records.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

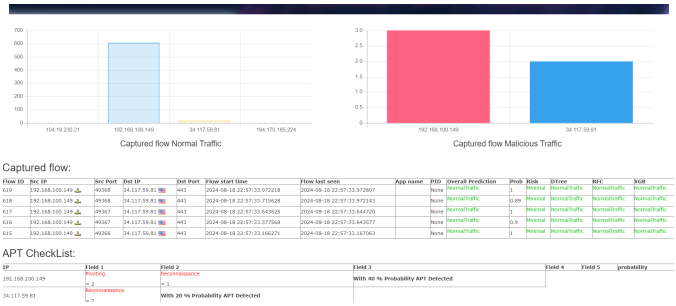

Fig. 5. APT Model Prototype Deployment UI

3) F-1 Score: It is the accuracy and recall measure's harmonic mean. It works very well with highly imbalanced datasets.

$$F-1score = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{3}$$

4) Accuarcy: It determines a ratio between all of the dataset's correctly identified connection records. It is advantageous when each class has equal importance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

V. Experimental Setup

We have divided our experimentation into two parts.

- Model Training
- Model Development

A. Model Training:

Several independent processes come together to create the overall structure of the framework.The overall architecture is shown in Figure [5] for the system. The selection of data and monitoring of it is the initial step for this approach. In this process, the dataset is examined carefully. Additionally, the dataset undergoes data preprocessing which includes activities like cleaning data, visualization of data, and feature engineering. Feature extraction is the part of feature engineering in which we extract the features of the dataset by applying specific steps. Then after that, these extracted features are divided into training and testing sets of data with a ratio of 80-20. The learning algorithm utilized the training data, and the final model was created. By using the various metrics assessment, the best model was employed to evaluate the final model in comparison to the testing set.

B. Model Development:

In the development of our model, we have utilized the Npcap module for packet capturing on Windows in which the Wireshark library was utilized to capture network packets. The whole description of Npcap is as follows. Data Filtration and Feature engineering was applied to it to predict the logfile on

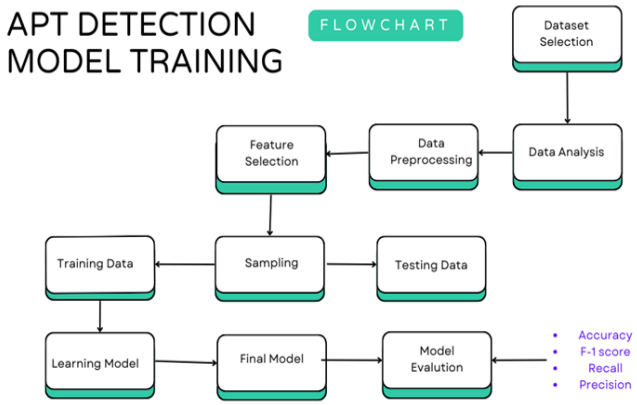

Fig. 6. APT Training Model Flowchart

runtime and a machine learning model was deployed for this prediction.

Our proposed methodology is executed on Windows 10 with 3 processor core i7 10th generation with 8 GB RAM With Nvidia GPU. We have utilized the Flask framework for the development of Python, HTML, and JavaScript. We employed Google Colab for the training of machine learning and deep learning models. Python was used for the whole procedure to train the model. We can see in Figure [6] our UI APT System.

As you can see from the graph the normal traffic flow is captured. On the x-axis, we have shown the different IPs from which the traffic comes, and, on the y-axis, it shows the traffic flow or the number of packets of the normal traffic. The IP address attributes are the base that grows on the x-axis while the number of packets increases on the y-axis. As a new IP address is added to the stack, the x-axis keeps the record of the newly added IP address. Similarly, when a new packet of the same IP address is observed it increases the count on the y-axis shown in Figure [7]. The malicious traffic records are kept

Fig. 7. Normal Traffic Flow

on the right side of the UI. The below graph is for malicious flow capture. On the X-axis it is shown that the different Ips from the different packets come from. As far as the whole record of packets can be kept in the middle of the UI using a tabular structure, which indicates that whenever a new packet

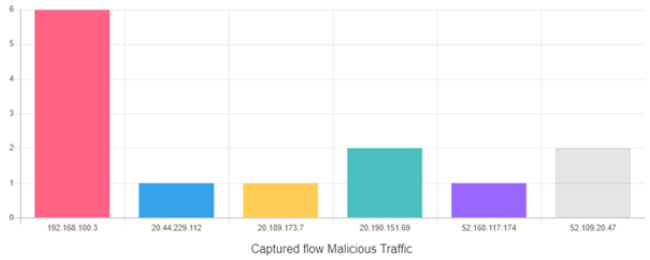

Fig. 8. Malicious Traffic Flow

is captured by the IDS it is shown in the table. First, the IDS captured the packet pass it to the Machine learning model and then some attributes of the packet and the Machine learning model decision are shown on the table. There are totally three models deployed that use the highest count technique to show the overall decision to the user. For example, if one model says that the current packet is showing the behaviors of APT, we treat it as a false positive, but if two models predict that a certain packet showing the APT cycle, we treat it a legitimate decision and the malicious packet graph will update itself in red color which will warn the user that a certain IP address is performing the malicious activity. Figure [9] shows the source port and source IP with the destination port and IP. The table has a column of overall prediction which is the most important. It shows that the prediction is either malicious or normal which is dependent on the probability of the APT cycle.

Captured flow:

Flow ID	Src IP	Src Port	Dst IP	Dst Port	Flow start time	Flow last seen	App name	PID	Overall Prediction	Prob	Risk	DTree	RFC	XGB
54	192.168.100.149	65174	34.117.59.81	443	2024-08-18 22:53:34.485858	2024-08-18 22:53:34.740564		None	NormalTraffic	0.89	Minimal	NormalTraffic	NormalTraffic	NormalTraffic
53	192.168.100.149	65173	34.117.59.81	443	2024-08-18 22:53:34.100243	2024-08-18 22:53:34.101159		None	NormalTraffic	1	Minimal	NormalTraffic	NormalTraffic	NormalTraffic
52	192.168.100.149	65173	34.117.59.81	443	2024-08-18 22:53:33.640715	2024-08-18 22:53:34.100168		None	NormalTraffic	0.72	Low	NormalTraffic	NormalTraffic	NormalTraffic
51	192.168.100.149	65172	34.117.59.81	443	2024-08-18 22:53:33.648432	2024-08-18 22:53:33.649367		None	NormalTraffic	1	Minimal	NormalTraffic	NormalTraffic	NormalTraffic
50	192.168.100.149	65172	34.117.59.81	443	2024-08-18 22:53:33.383110	2024-08-18 22:53:33.648314		None	NormalTraffic	0.72	Low	NormalTraffic	NormalTraffic	NormalTraffic

Fig. 9. Capture flow Prediction Table

C. Evaluation Results with Comparative Analysis in Tabular Form

Below we can see the models in which accuracy before and after augmentation is mentioned in tabular form in table [1]. As we can see the highest accuracy achieved before and after Augmentation is XgBoost than after Random Forest and so on.

TABLE I
APT CLASSIFICATION MODEL PERFORMANCE

Sr. No	Models	Accuracy Training Before Augmentation	Accuracy Training After Augmentation
1	XGBoost	99.93%	99.95%
2	RF	99.90%	99.89%
3	DT	99.82%	99.82%
4	AdaBoost	99.07%	97.83%
5	KNN	98.41%	97.02%
6	LSTM	98%	97%
7	ANN	98.41%	96.98%
8	Auto Encoder	98.34%	96.89%

D. Evaluation Results Analysis in Form of Graph

We have shown the depiction of the proposed experimentation of our research and the techniques that we have applied. We evaluate different Machine Learning models, including DT, RF, etc., and compare their results in the graphical form below before and After Augmentation. In the x-axis, models of machine learning are shown whereas in the y-axis, the percentage of performances matrix can be visualized

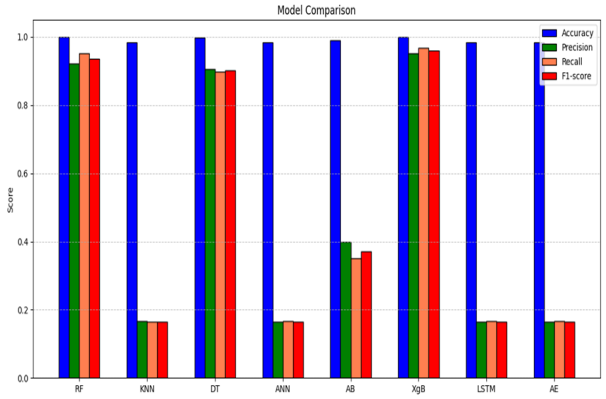

Fig. 10. Compartive Analysis graph before Augmentation

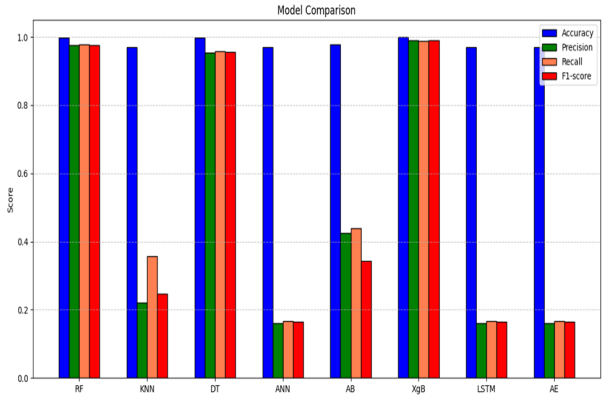

Fig. 11. Comparative Analysis graph after Augmentation

VI. CONCLUSION AND FUTURE WORK

Information security plays a crucial role in ensuring the development of dependable and secure networks by allowing information technology to grow rapidly. In the research, we have utilized a novel approach for APT identification using an APT dataset with explainable machine learning techniques. We have staggeringly examined the dataset and investigated extensively to process it effectively and efficiently. The proposed system was tested on the dataset SCVIC-APT-2021 and compared precision, accuracy, recall, and F1-score by other algorithms that are utilized in the research. At first, we analyzed data and preprocessed it to improve the quality of data to improve and achieve the best results. After that, feature extraction was applied using Principal Component Analysis and Pearson Corelation and find the feature score.

Seven machine learning models i.e., XGBoost, Random Forest, KNN, ANN, AdaBoost, Autoencoder, Decision Tree, and one deep learning model LSTM have been used in the research. Data augmentation technique is utilized for the comparison of the results before and after augmentation. The analysis of the performance shows that XGBoost performs very well compared to other techniques. The detection rate of XGboost was 99.93% before augmentation and 99.95% after augmentation. Random Forest is the second highest in performance i.ie.99.90% before augmentation and 99.89% after augmentation. The Autoencoder gives less comparatively to the other models i.e. 98.34% before augmentation and 96.8% after augmentation. Our research offers reasonable intuition to research objectives

A. Future Work

For future work, we propose several enhancements based on the findings of this study: Extend the research on algorithms to identify additional types alongside APT. Develop a self-learning model that can be deployed on the network for both malicious prediction and ongoing training, addressing the limited number of malicious classes. Implement our model on a blockchain to safeguard against adversarial AI attacks.

REFERENCES

[1] M. Ussath, D. Jaeger, F. Cheng, C. Meinel, Advanced persistent threats: Behind the scenes, in: 2016 Annual Conference on Information Science and Systems (CISS), IEEE, 2016, pp. 181–186.

[2] the-latest-cyber-crime-statistics.
URL https://aag-it.com/the-latest-cyber-crime-statistics/

[3] cybersecurity-statistics.
URL https://www.techopedia.com/cybersecurity-statistics

[4] Key Ransomware Statistics for Q3 2023 — corvusinsurance.com,

[5] D. Kiwia, A. Dehghantanha, K.-K. R. Choo, J. Slaughter, A cyber kill chain based taxonomy of banking trojans for evolutionary computational intelligence, Journal of computational science 27 (2018) 394–409.

[6] Y. Shen, M. Simsek, B. Kantarci, H. T. Mouftah, M. Bagheri, P. Djukic, Prior knowledge based advanced persistent threats detection for iot in a realistic benchmark, in: GLOBECOM 2022-2022 IEEE Global Communications Conference, IEEE, 2022, pp. 3551–3556.

[7] F. N. Khan, K. Zhong, W. H. Al-Arashi, C. Yu, C. Lu, A. P. T. Lau, Modulation format identification in coherent receivers using deep machine learning, IEEE Photonics Technology Letters 28 (17) (2016) 1886–1889.

[8] M. M. Hasan, M. U. Islam, J. Uddin, Advanced persistent threat identification with boosting and explainable ai, SN Computer Science 4 (3) (2023) 271.

[9] P. Watson, K. Gupta, R. Mahajan, Development of knowledge-based artificial neural network models for microwave components, in: 1998 IEEE MTT-S International Microwave Symposium Digest (Cat. No. 98CH36192), Vol. 1, IEEE, 1998, pp. 9–12.

[10] J. Liu, Y. Shen, M. Simsek, B. Kantarci, H. T. Mouftah, M. Bagheri, P. Djukic, A new realistic benchmark for advanced persistent threats in network traffic, IEEE Networking Letters 4 (3) (2022) 162–166.

[11] C. Do Xuan, Detecting apt attacks based on network traffic using machine learning, Journal of Web Engineering (2021) 171–190.

[12] H. Neuschmied, M. Winter, B. Stojanović, K. Hofer-Schmitz, J. Božić, U. Kleb, Apt-attack detection based on multi-stage autoencoders, Applied Sciences 12 (13) (2022) 6816.

[13] M. Basi, T. Shang, J. Liu, Y. Jiang, W. Wang, Apt detection based on rsdn framework, in: 2024 IEEE 14th International Conference on Electronics Information and Emergency Communication (ICEIEC), IEEE, 2024, pp. 1–6.

[14] J. Liu, Y. Shen, M. Simsek, B. Kantarci, H. T. Mouftah, M. Bagheri, P. Djukic, Scvic-apt-2021 (2022). doi:10.21227/g2z5-ep97.
URL https://dx.doi.org/10.21227/g2z5-ep97

[15] S. N. S. A. Sham, K. K. Ishak, N. A. M. Razali, N. M. Noor, N. A. Hasbullah, Iot attack detection using machine learning and deep learning in smart home, JOIV: International Journal on Informatics Visualization 8 (1) (2024) 510–519.

[16] J. Liu, Y. Shen, M. Simsek, B. Kantarci, H. Mouftah, M. Bagheri, P. Djukic, A new realistic benchmark for advanced persistent threats in network traffic, IEEE Networking Letters 4 (2022) 1–1. doi:10.1109/LNET.2022.3185553.

[17] R. May, G. Dandy, H. Maier, Review of input variable selection methods for artificial neural networks, Artificial neural networks-methodological advances and biomedical applications 10 (1) (2011) 19–45.

[18] S. S. Dhaliwal, A.-A. Nahid, R. Abbas, Effective intrusion detection system using xgboost, Information 9 (7) (2018) 149.

[19] Decision Tree — mastersindatascience.org, https://www.mastersindatascience.org/learning/machine-learning-algorithms/decision-tree/, [Accessed 19-08-2024].

[20] J. Wang, X. Xiong, N. Zhou, Z. Li, W. Wang, Early warning method for transmission line galloping based on svm and adaboost bi-level classifiers, IET Generation, Transmission & Distribution 10 (14) (2016) 3499–3507.

[21] C. M. Bishop, Neural networks and their applications, Review of scientific instruments 65 (6) (1994) 1803–1832.

[22] J. A. Anderson, An introduction to neural networks, MIT press, 1995.