A Unified Behavioral Attack DNA Framework for Global Cyber Threat Detection Using Multi-Dataset Learning

Deepta Chakravarty[1], Mayukh Mondal[2], Disha Chaudhury[3],Dr. Lakshmi Dhevi B[4]

Department of Networking and Communications SRM Institute of Science and Technology Kattankulathur, Chennai, Tamil Nadu

DOI: https://dx.doi.org/10.51244/IJRSI.2026.1305000195

Received: 26 May 2026; Accepted: 01 June 2026; Published: 09 June 2026

ABSTRACT

Cyber threats are becoming increasingly complex, and laws around data sovereignty are leading to increasingly complex regulations; this has rendered global collaboration in cybersecurity not only necessary but challenging. Traditional intrusion detection systems (IDSs) are often based on centralized architectures or signatures, which are inadequate against these attacks, which can change and evolve over the course of an attack. This paper introduces a single system of detecting the global cyber threats in the shape of the behavioral representation known as — Attack DNA. The features are combined in a systematical fashion to generate a common feature space for heterogeneous datasets of intrusion detection (CIC-IDS2017, NSL-KDD, UNSW-NB15 and TON_IoT); packet count, byte count, flow rate, etc. are used to produce a standardized description of network behavior in a variety of environments. It compares the sequence and feature-based models with the Long Short-Term Memory (LSTM) networks and Random Forest classifiers; the feature-based models are shown to be better in the case of non-temporal aggregated data. Also, a geographical analysis is conducted to assess differences in the pattern of attacks by different geographical sources, illustrating the usability of the framework in cybersecurity situations on a global scale. Results highlight the significance of standardized behavior for scalable, interpretable and cross-dataset intrusions.

Keywords—Cyber Threat Detection, Attack DNA, Behavioral Analysis, Intrusion Detection Systems, Multi-Dataset Learning, Random Forest, LSTM, Feature Engineering, Network Security.

INTRODUCTION

An increased number of digitalized systems and capabilities to communicate with the systems have rendered the modern computing environment as a highly complex and integrated environment. From cloud computing and IoTs to large enterprise networks, the use of digitalized systems is commonplace in the most basic aspects of day-to-day life. The proliferation of connectivity has also brought about significant advances in technology, but this has given attackers an increasingly wide area to attack and made cybersecurity a critical topic for any company or government in the world [18].

These attacks are more than a single attack; they represent a very advanced attack process which relies on weaknesses of various layers of an organization's systems. The threats of today are more stealthy, adaptable, and long-term; and they can evade normal protection methods like firewalls and intrusion detection systems (IDS). The attack process consists of stages including reconnaissance, lateral movement, privilege escalation and data theft, and these will be difficult to detect using traditional approaches [16]. Cybersecurity measures that exist at present rely mainly on signature-based and centralized analytics of data. Signature based can detect threats that have already been seen, but are unable to detect zero day threats or changing threat patterns.

Moreover, centralized approaches depend heavily on aggregation of large volumes of raw data, thus making them hard to scale up with great latency and problems of privacy [19]. Those constraints highlight the need for new more scalable techniques. Data sovereignty and privacy laws have also added to the difficulties faced by cyber security in recent years. Less sharing of critical data between organizations, across borders and beyond

organizational boundaries, thereby reducing the usefulness of any threat intelligence systems based on collaboration. The contradiction is that while it is important to collaborate internationally in order to identify such distributed cyber threats, the raw data sharing is constrained.

Consequently, organizations tend to run on limited visibility, limiting their chances of identifying coordinated attacks [11], [12]. In order to overcome these issues, a new framework is required which will allow detecting the global threats without using raw data sharing. Here, one way to address the issue is to shift the emphasis in the raw log analysis to behavioral intelligence. Behavioral methods are used to examine system activity patterns, like the duration of connections, packet transmission, and dynamic flow of data instead of the individual events. These patterns give a more advanced abstraction of system behavior and can effectively detect anomalies and maintain the privacy of data [16]. In this paper, we introduce the idea of Attack DNA, a single behavioral definition of a network operation, derived from structured properties of the duration, number of packets, number of bytes and flow rate.

The proposed approach is to map the various and dataset-specific features into a common feature space, enabling the integration of different intrusion detection datasets such as CIC-IDS2017, NSL-KDD, UNSW-NB15, and TON-IoT. Such standardization makes it easier to learn across datasets and enhances generalization to various network settings [15]. The proposed system involves the comparison of the sequence-based machine learning approach with feature-based machine learning approach, i.e., Long Short-Term Memory (LSTM) networks and Random Forest classifiers respectively. The LSTM is used to learn temporal dependency of sequential data, while the feature-based model, e.g., the Random Forest is used to learn discriminative patterns from structured inputs.

Moreover, a geographical analysis is added in order to analyze the variation of the attacks' trend according to the geographical attack sources, and so it provides hints about the behavior of cyber threat globally and helps to the distributed analysis [11, 12]. Overall, this paper demonstrates that the behavioral feature standardization results in scalable and interpretable and cross-dataset intrusion detection. The proposed solution aims to minimize the difference between different data sets and provide a common analysis tool to guide to more resilient and universally applicable cyber security solutions.

LITERATURE SURVEY

Today's cybersecurity landscape has evolved to become intelligent, adaptive, and privacy respecting in order to address the complexity of cyber threats. The existing work can be revisited in the light of the data heterogeneity, behavioral modeling, and feasible deployment constraints all of which coincide with the goals of the proposed Attack DNA framework.

Challenges in Multi-Dataset and Heterogeneous Intrusion Detection

One of the significant weaknesses that are affecting the current research in intrusion detection is the inability to interoperate with different datasets and environments. To solve this issue, Wang et al. [15] proposed a multi-domain attention-based Bidirectional LSTM model that demonstrates good generalization. The model, however, requires a lot of data for labeled data and high computing power and is not scalable to real world application. This reinforces the need for a unified feature space that can help to ease the task of cross-dataset learning.

Similarly, Zhang et al. [18] emphasized the importance and the role of deep learning techniques in cyber-physical systems, but noted that most models have been trained on a single dataset, reducing their adaptation capabilities. Imrana et al. [7] also found the sequence-based models as effective models in small data sets, and ineffective in heterogeneous data sources.

CNN-BiLSTM based methods and autoencoder-based methods are combined with hybrid methods such as the methods proposed by Said et al. [14] and Bi et al. [3] respectively to improve feature extraction. But these models are more complex and are not common to all datasets. All these restraints highlight the absence of a commonly

agreed upon behavioral representation, and the current work, the Attack DNA framework, aims to map the different datasets to this common feature space.

Behavioral Modeling versus Temporal Dependency in Attack Detection

The recent studies focus on modelling behavioral patterns of cyber-attacks using sequence models and probabilistic models. In the field of sequential data, Yang et al. [17] proposed an intrusion detection system based on BiLSTM with high detection rate and large training data size. Kang et al. [8] introduced ActDetector, a multi-stage attack modeling using sequence embeddings and LSTM networks, that is capable of effectively modeling the attack progression.

Cheimonidis and Rantos [4] also explored more probabilistic modeling on the basis of Markov chains and Bayesian networks to predict the evolution of the attack. They can be used to model attack sequence but are very time sensitive and their usage is limited to data with predefined structure and cannot be applied with aggregated data or non-sequential data.

Xuan [16] mentioned the correlations between the domain features and the IP features to detect APT, and the authors proved that the capability to detect APTs using behavioral features is better than static signatures. But this has special techniques of feature engineering and cannot be extended to other sets of data.

From this research, it seems that there is a clear contradiction: sequentially, Temporal Modelling is a sensible approach, but not every time when analyzing the Heterogeneity and Aggregation of the Dataset. For this, the Attack DNA technique was developed which focuses on behaviors which are constant across datasets, namely duration, number of packets, number of bytes, flow rate, etc..

Privacy, Collaboration, and Real-World Threat Intelligence Collection

Learning, distributed, based on deception techniques, has been adopted as a solution to address the requirements of intrusion detection in a privacy-preserving and collaborative manner. To allow model training without exchanging raw data, Raza et al. [11] proposed a system based on federated learning. Although it works, this method causes overhead in communications and systems.

Revathi and Kingslin [12] extended the idea of using blockchain with federated learning, together with the use of cryptographic algorithms, to ensure the security and inalterability of collaboration. They are, however, complicated computing techniques which are difficult to scale up to large deployments. The multi-party secure computation proposed by Liu et al. [9] is not efficient for real-time intrusion detection system.

On the other hand, Alohali et al. [1] utilized blockchain technology together with Explainable AI techniques to improve transparency in cyber security systems, though issues related to scalability arise. These pieces highlight the importance of privacy as well as the tradeoff between security and effective system functioning.

At the same time efforts have been equally carried out to record real life attacker actions through deception techniques such as honeypots etc. Rabah et al. [13] suggested an ensemble learning model that combines honeypot information in detecting real-time DDoS attacks with a high degree of accuracy, consuming a lot of resources. Naik et al. [10] suggested a dynamic honeypot system which is adaptively learnable but not very scalable.

Franco et al. [6] provided a survey of honeypot architectures and discussed the use of honeypots to collect behavioral data, while Ali et al. [2] demonstrated the use of a honeypot for real world applications for detecting SSH attacks. Also, Fan et al. [5] demonstrated the role of interpretability in the detection systems based on an RNN.

All these studies indicate that systems that preserve privacy and use deception play a positive role in threat intelligence, but are not typically coupled with single and scalable detection systems. The suggested Attack DNA framework fills this gap with allowing privacy-preserving behavioral abstraction and global collaboration without sharing of raw data.

PROPOSED METHODOLOGY

This section provides a proposed framework for the detection of global cyber threats and how it can be designed and implemented using a single behavioral representation, Attack DNA. The methodology focuses on the process of mapping the intrusion detection data from different sources into a common feature space, and evaluating the performance of different machine learning paradigms that address the question of how well they can operate in this kind of environment. The limitations of the current systems to deal with cross-dataset variability and the absence of a shared abstraction of behavior, as mentioned in previous literature [7], [18], are what prompt the proposed approach.

System Overview

The system proposed takes a behavior-based and data-centric perspective of intrusion detection. As opposed to the classical methods based on analysis of raw data and inclusion of certain features of the data set, the proposed model is oriented towards abstraction from the data and representation of the behavior of the network using a generalized version of the data, thus moving on to the new line of research in cyber security. The ability to transition from raw data to behavioral models will create greater flexibility and resilience to new threats [16].

The entire structure involved in the process are data set collection, features extraction, preprocessing, train and evaluate the model. In the first step, the features extraction from the raw data of different origins is considered. As the base for representation of Attack DNA, we use mapping of extracted attributes into a single space of features. For this reason, this type of architecture of the IDS is helpful to resolve the issue that existing IDS is facing, which is the issue of differences between data sets [15].

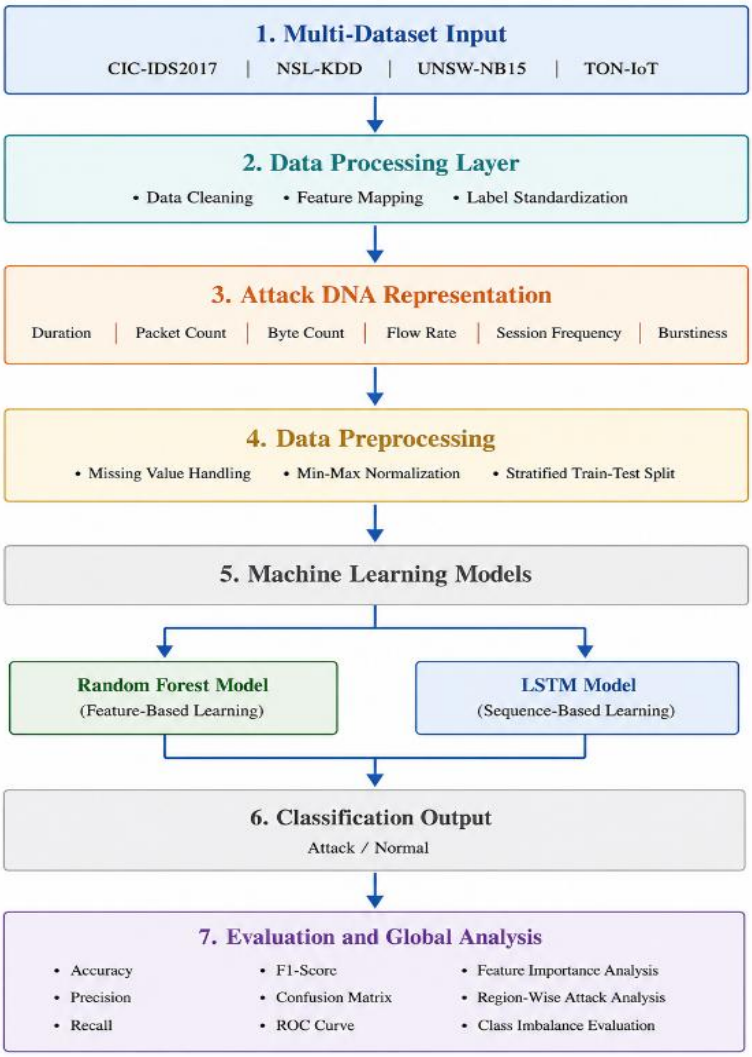

Fig. 3.1 Proposed Architecture of the Attack DNA Framework for Global Cyber Threat Detection

Figure 3.1 illustrates the structure of the Attack DNA proposal. The multiple inputs are taken as data and converted to behavioral forms, followed by classification and analysis on a global level. There have been two learning methods employed for this purpose. The first one is the sequence-based LSTM learning model [8], [17] and the second one is the feature based Random Forest classification model for learning [11, [12]. In Attack DNA, an additional analysis layer has been embedded to help to detect geographical differences of the cyber-attacks, which guarantees privacy-based global analysis [11], [12] that is region-aware.

Dataset Collection and Integration

The datasets that are planned to be used for detection of intrusion benchmark are CIC-IDS2017, NSL-KDD, UNSW-NB15, and TON-IoT (in general, these sets of data can be called as representative in terms of the network environment and type of attacks). The results of the previous studies showed that standard models trained from one data set were not domain adaptable [15].

However, the datasets have varying structures of their features, and thus it is impossible to merge them in any form. This difficulty is addressed by an official procedure for integration, and the features that could be useful for analysis because they can explain the behavior of the network are identified and subsequently mapped into a common pool of standardized features. This technique allows for the same processing of multiple data sets, and does not rely on the characteristics of each data set.

Each set of data would be labelled with a geographical location, e.g. US, EU, India, Global and across-region comparisons would be available to create an atmosphere of a global cyber threat. It allows distributed threat modelling without any data transfer/flow, which satisfies the privacy-driven practices in the field of cybersecurity [11, 12]. The result is a single data set that ties together data from several sources into a cohesive system.

Attack DNA Feature Engineering

This paper's novel inputs are the techniques for extracting network behavior into a few generalized attributes, called Attack DNA. What makes this approach different from the classic approaches where high dimensional attributes specific to the dataset are taken into account is that it is based on the use of low dimensional attributes that can be generalized across different datasets [16].

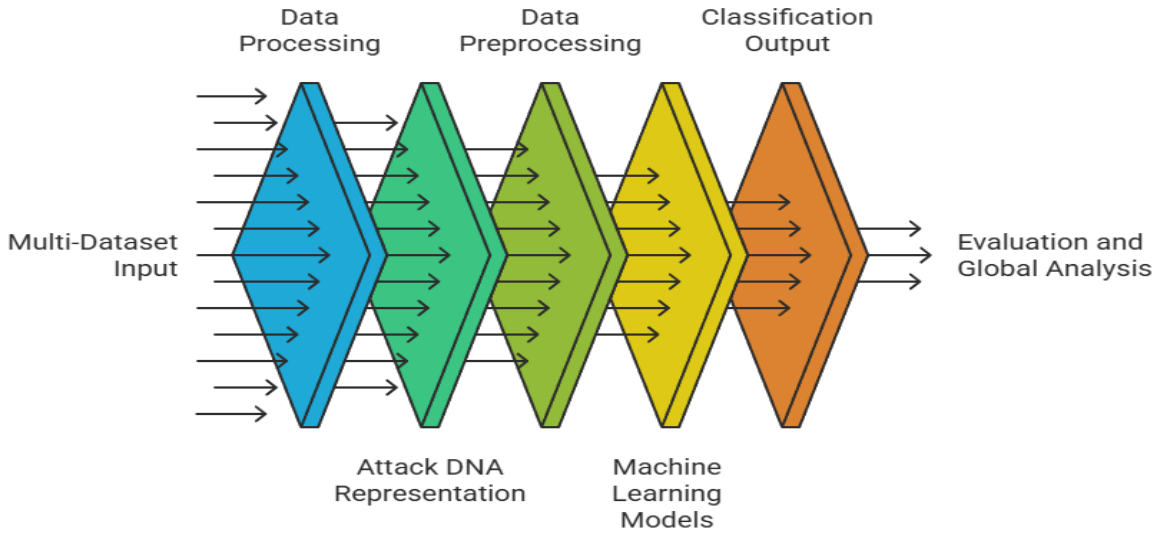

Fig. 3.2 Conceptual Flow of Attack DNA-Based Cyber Threat Detection Pipeline

Figure 3.2 depicts the conceptual transformation process, according to which the input raw multi-dataset data are gradually transformed into the Attack DNA representations and then used for classification and analysis purposes. The approach avoids any use of dataset-specific structures, which increases the interpretability of the

method through mapping diverse types of data into a common feature space and therefore no dependency on dataset-specific structures. All features have behavioral significance, allowing learning of general trends and not only trends in a particular dataset.

Table 3.1 Importance of Features in Attack DNA

Feature	Description
Duration	Connection duration
Packet Count	Total packets exchanged
Byte Count	Total bytes transmitted
Flow Rate	Traffic rate
Session Frequency	Frequency of communication sessions
Burstiness	Traffic fluctuation intensity

The significance of every feature in Attack DNA framework is given in *Table 3.1*. These selected characteristics are: Duration, Packet count, Byte count, Flow rate, Session Frequency and Burstiness. The duration measures the length of time when network traffic is active, packet count and the byte count measure the amount of traffic that occurs, the flow rate measures intensity of network traffic, the session frequency measures the repetition patterns of network traffic, and the burstiness measures sudden fluctuations in network traffic. These features combine to give a more complex behavioral abstraction that can capture both steady state and anomalous network activities over heterogeneous data sets [15].

Data Preprocessing and Model Design

The data is then preprocessed, after feature extraction, to give a homogeneous and machine learning friendly data. This involves the process of dealing with missing and invalid values, elimination of inconsistency and the handling of undefined or infinite values. The feature scales are standardized using Min-Max normalization, a widely used normalization technique in intrusion detection systems [18] to convert all features into the same scale, thus improving the stability and convergence of the model.

Also, the class labels are converted to binary labels (attack and normal) and the region labels are stored to be used afterwards to analyze. These resulting data can be viewed as a clean and structured representation of behavior and can be trained on to overcome the issues with heterogeneous data integration highlighted in the previous studies [15].

To evaluate the efficiency of the proposed Attack DNA representation, both feature-based and sequence-based learning methods are presented, so that the comparison of the modeling approaches can be done on common behavioral data.

Random Forest Model

The more general classifier is the Random Forest as it works well for structured and tabular data. Being an ensemble method of learning, it builds a number of decision trees with random sample of the features and combines their output to enhance classification.

In this framework, the best technique is to use Random Forest, because it is nonlinear, it does not assume time dependence and it is noise-tolerant. Moreover, it offers measures of feature importance, thus making behavioral attributes interpretable. These characteristics render it suitable to behavior-based intrusion detection as has been witnessed in previous research [16].

LSTM Model

A Long Short-Term Memory (LSTM) network is used to determine the suitability of sequence-based learning to intrusion detection. LSTM models are built to capture the time sequence of a series of data points, and so far, have been used to identify cyber threats [8,17].

In this paper, the dataset is restructured into a sequential one for passing in the LSTM. The integrated data is not always a continuous time series, however, as it is the aggregate of many data sources. Given this consideration the LSTM model is not used as the primary detection mechanism, but as a comparative baseline. This is similar to the previous finding that the sequence-based models don't perform well when used in non-temporal or heterogeneous data settings [4, 7].

The LSTM consisted of one LSTM layer of 64 hidden cells followed by a dropout layer (dropout_rate = 0.2) and two dense layers. The model was trained for 5 epochs with 32 batch size and Adam optimizer, and binary cross-entropy loss. The training was done using a 20% validation split.

Due to the sequential nature of processing and parameter updates for recurrent networks, LSTM networks are prone to higher computational overhead for modeling temporal dependencies. The Attack DNA dataset is not temporal and is largely aggregated, which reduces the usefulness of temporal modelling. Therefore, the extra computational effort of LSTM did not correspond to the performance increase when compared to the RF classifier.

Global Region-Based Analysis

To expand the framework further to include more than the traditional intrusion detection, a region-based analysis is added to give the world a view of cyber threats. A region label is provided with each data sample that can be used to compare data to geographically distributed environments.

The distribution of classes, frequency of attacks and the behavioral differences of regions are analyzed through statistical analysis. This helps in determining threat pattern in region and assists in cyber threat analysis on the global scale. Such views, which are distributed and region sensitive, have emerged as a major focus of current cyber security models [11, 12].

This proposed framework will enable non-shared data to be used to identify a global threat on a global scale, using behavioral abstraction and region-based insight.

EXPERIMENTAL RESULTS AND EVALUATION

This segment will assess the proposed Attack DNA framework by applying various intrusion detection datasets to the statistical analysis, visualization and model-based analysis. The suitability, scalability and flexibility of these techniques will be evaluated across various contexts. The assessment will be based on the effectiveness of classification, interpretability and generalization of the learned knowledge to other homogeneous data sources.

Dataset Distribution Analysis

Before using the model, the structure of the merged data set is examined to obtain knowledge on the distribution of classes and regions.

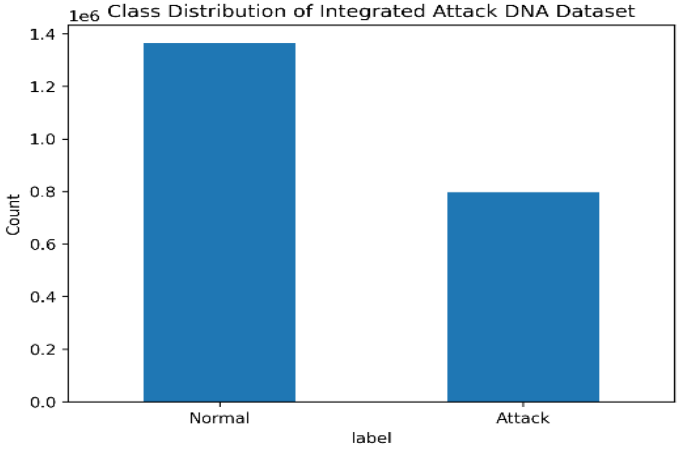

Fig. 4.1 Class Distribution of Integrated Dataset

Figure 4.1 shows the distribution of the normal and attack sample. The data set is highly imbalanced, having many attacks. This imbalance represents the actual circumstances of intrusions, which always have more attacks than normal traffic [17]. Such a balance has to be achieved by using complex models that can deal with imbalanced data.

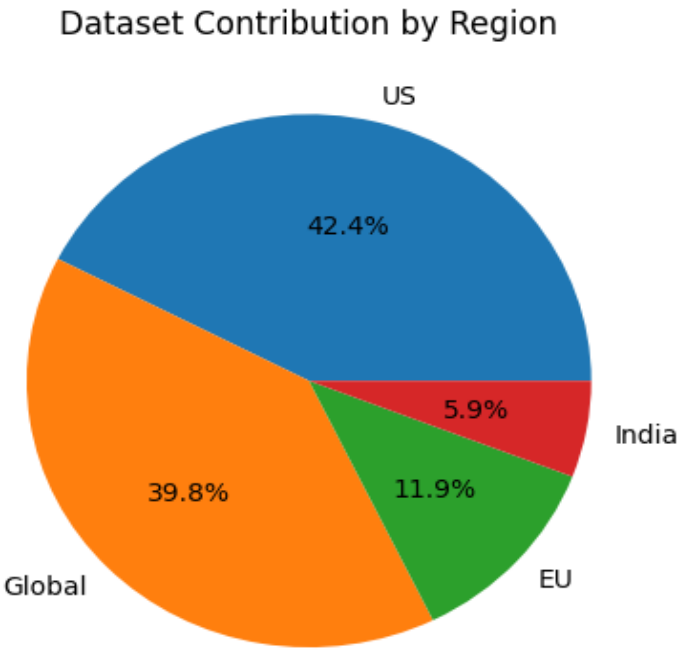

Fig. 4.2 Region-wise Dataset Distribution

Figure 4.2 displays the distribution of the data over the geographical region; and it can be observed that geographical contributions come from many sources. This helps in achieving heterogeneity in the traffic flow pattern as well as increases the generalization capability of the model in different scenarios [15]. The important thing is to be aware of not only the idea of imbalance within classes, but also regional diversity that directly affects the training of the model and its evaluation.

Class Imbalance Handling

As seen in *Figure 4.1*, the integrated dataset is imbalanced, that is, there are a lot of attack samples compared to normal traffic samples. In intrusion detection, one of the common problems is class imbalance, as the machine learning model can be biased towards majority class and give false estimation of model performance if accuracy is only taken into account.

A class imbalance train-test split strategy was used to ensure that the class distributions of both train and test sets were the same, which helps to alleviate the effects of class imbalance when evaluating the model. Additionally, the performance of the models was evaluated by various evaluation measures such as precision, recall and F1-score as well as accuracy. However, a special focus was given to the recall of the attack class, which is very important for cyber security applications, where if the malicious actions are not detected, serious security risks may occur.

The results acquired showed that RF model achieved high precision and recall even though the distribution of classes was not balanced with its high accuracy in differentiating between normal and abnormal traffic patterns. Using multiple evaluation metrics gave a more robust evaluation of the performance of the model than accuracy alone, and allowed for a security-oriented view of the effectiveness of classification.

The proposed extension of the framework could be extended to explore more sophisticated class imbalance handling techniques like cost-sensitive learning, adaptive class weighting, synthetic oversampling and intelligent sampling strategies to further improve the detection performance in highly imbalanced intrusion detection settings.

Region-wise Attack Analysis

A study of attack distribution, done region-wise, to get a picture of the global distribution of attacks.

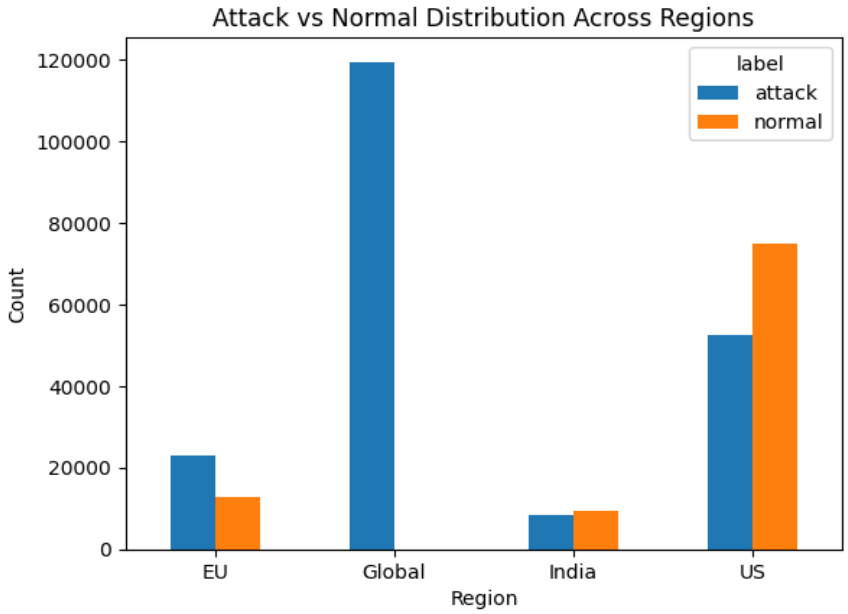

Fig. 4.3 Region-wise Attack Distribution

There are regions (*Figure 4.3*) to compare the attack and normal samples and the intensity of the attack is very different. These differences can be attributed to the differences of the data set, attack conditions generated in the simulation and the network dynamics that were used.

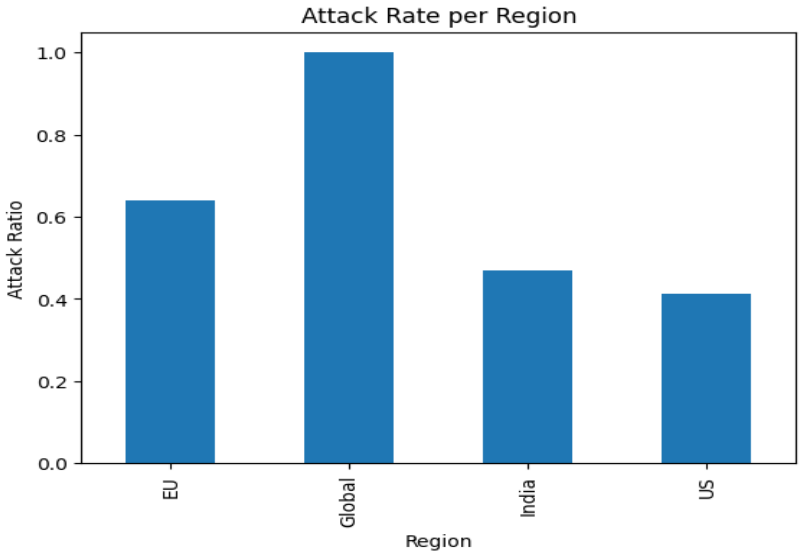

Fig. 4.4 Region-wise Attack Rate Analysis

The rate of attack (attack samples/total samples) is shown in *Figure 4.4*, and is a normalized comparison of the regions. This highlights the reality that areas are more likely to be at a higher relative threat level although they have less data. This result indicates that cyber threat distribution is not even and the need for region-specific

analysis to be included in an IDS system. All these viewpoints are even more relevant in a cybersecurity distributed architecture [11, 12].

Model Accuracy of Random Forest vs. LSTM

Random Forest model is considered as the main Attack DNA representation classifier.

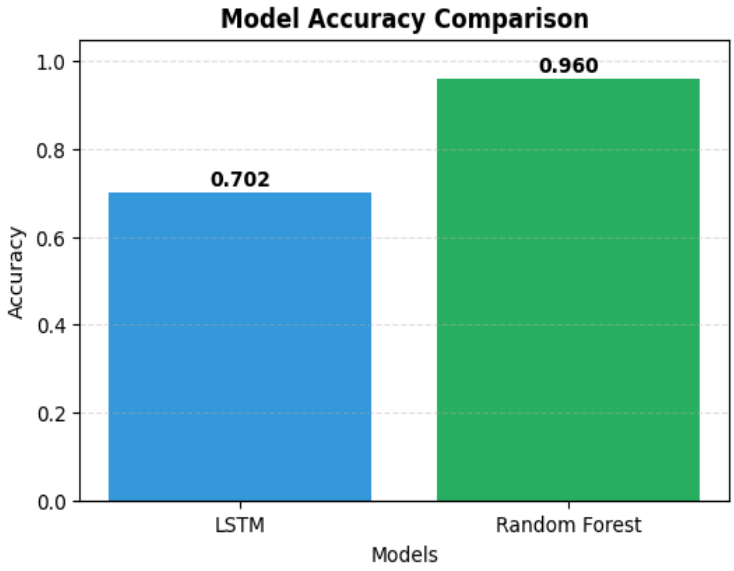

Fig. 4.5 Model Accuracy Comparison

Comparison of the model's correctness is done in *Figure 4.5*, where it is evident that the Random Forest is more accurate in this scenario compared to LSTM, as shown in section E.

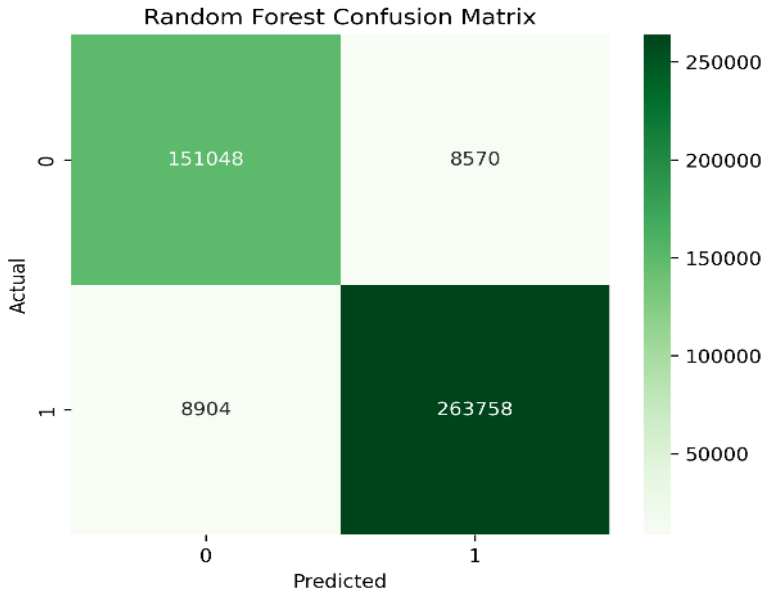

Fig. 4.6 Confusion Matrix for Random Forest Model

Figure 4.6 shows the confusion matrix, with a high true positive and true negative rate, and low false rates. This implies that the model can identify the normal and malicious traffic appropriately.

Other performance metrics that show that the model is robust include accuracy, recall and F1-score, where especially the instances of attack were identified. The attack class recall is particularly important because it decreases the chances of undetected threats [16]. Overall, the performance of Random Forest is high and stable

(95.96%) and the algorithm is suitable to be used in feature-based intrusion detection in aggregated and multi-datasets.

Model Performance (LSTM)

The LSTM model is tested to check the sequence-based learning on integrated data.

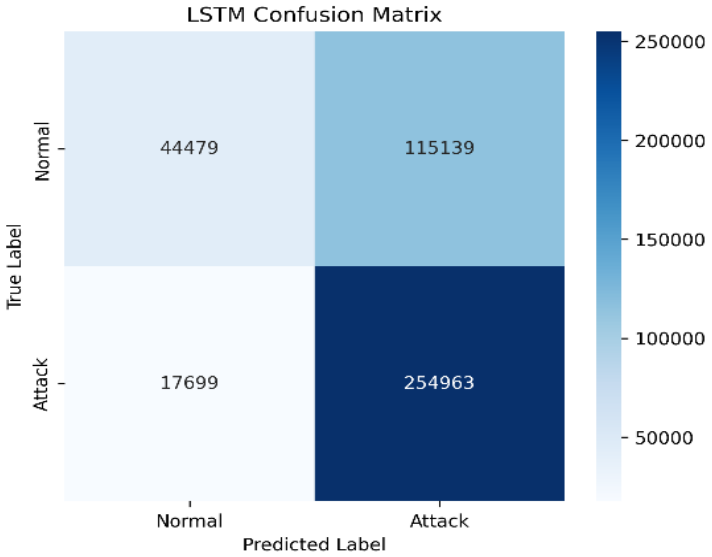

Fig. 4.7 Confusion Matrix for LSTM Model

As per the confusion matrix (*Figure 4.7*), the misclassification rates were observed to be more than in the case of Random Forest, particularly while differentiating between attack and normal samples. The LSTM model gave an accuracy of about 70.19% which was significantly poor as compared to the Random Forest Classifier. Although the model kept its rate of recall at a high level, the precision of the attack detection was degraded and the number of false alarms from normal traffic was increased.

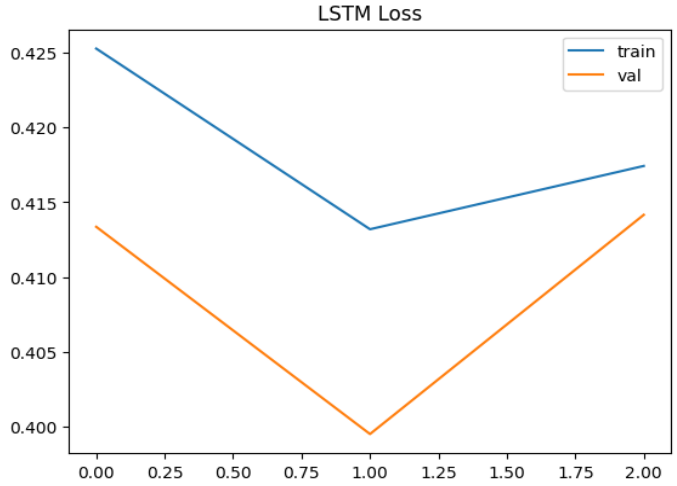

Fig. 4.8 Training and Validation Loss of LSTM Model

Figure 4.8 illustrates training loss and validation loss, where the losses are approaching each other, but there are some minor variations between epochs.

The relatively low performance could be attributed to the lack of a natural time dimension in the data as it is an aggregate sample from several sources. The LSTM models have been designed to take advantage of the sequential dependencies [8], [17] and will not be as effective if the sequential dependency is not present [4], [7].

Despite these drawbacks, the LSTM model is valuable for providing some insights into the scope of applicability of sequence-based techniques in heterogeneous environments.

Model Comparison

Random Forest and LSTM are evaluated and compared to confirm that there are differences in effectiveness of the model.

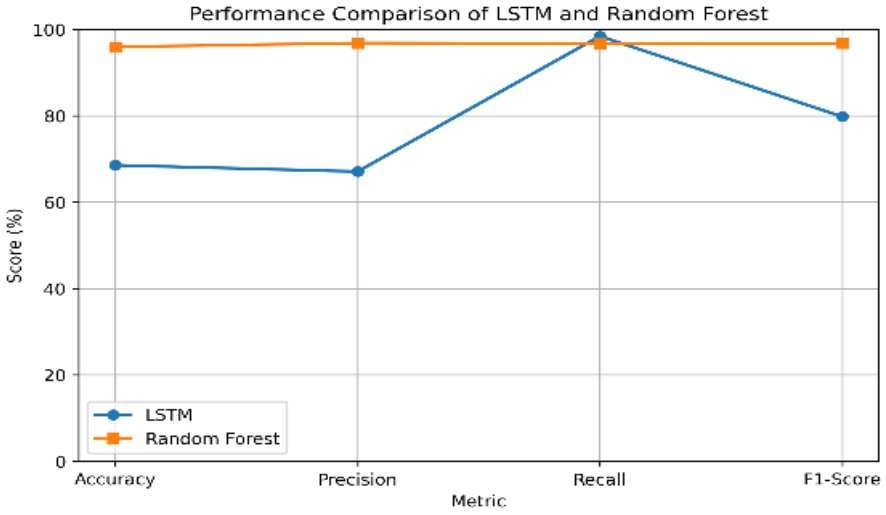

Fig. 4.9 Model Performance Comparison

As *Figure 4.9* indicates, the Random Forest has much better performance and is stable in comparison with LSTM. The performance difference shows feature-based learning is more appropriate to datasets with weak time dependencies.

The key takeaway of this analogy is that sequence-based models work better on time-series data and feature-based models are more flexible for aggregated and heterogeneous data. This is consistent with previous studies by different researchers on the shortcomings of temporal modeling [4],[7].

ROC Curve Analysis

The model performance at various classification thresholds is measured by Receiver Operating Characteristic (ROC) curves.

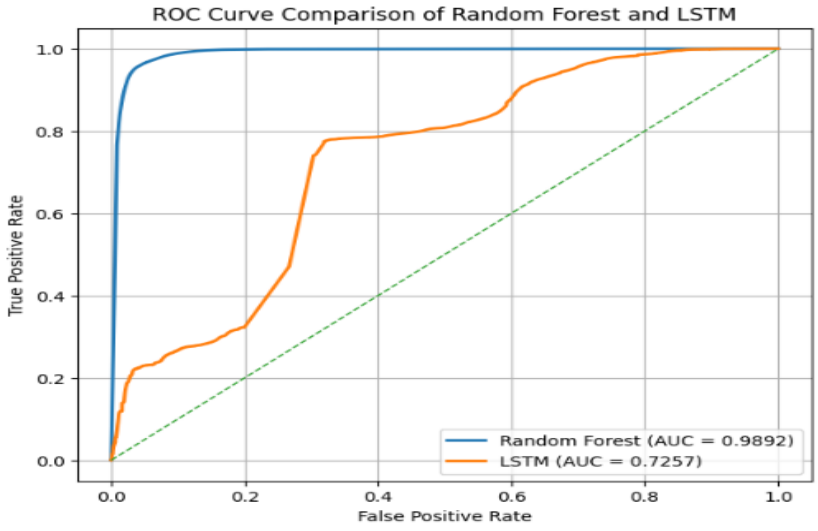

Fig. 4.10 ROC Curve Comparison of Models

As *Figure 4.10* indicates, Random Forest has a larger Area Under Curve (AUC) which means that it can better discriminate between attack and normal classes. The curve has a good trade-off between the true positive rate and false positive rate. The curve of the LSTM model is less prominent, meaning that it is less confident in terms of classification. An feature analysis of ROC provides an assessment independent of threshold, which further validates the strength and validity of the Random Forest model in this study.

Feature Importance Analysis

The Random Forest model is used to conduct feature importance analysis to increase its interpretability.

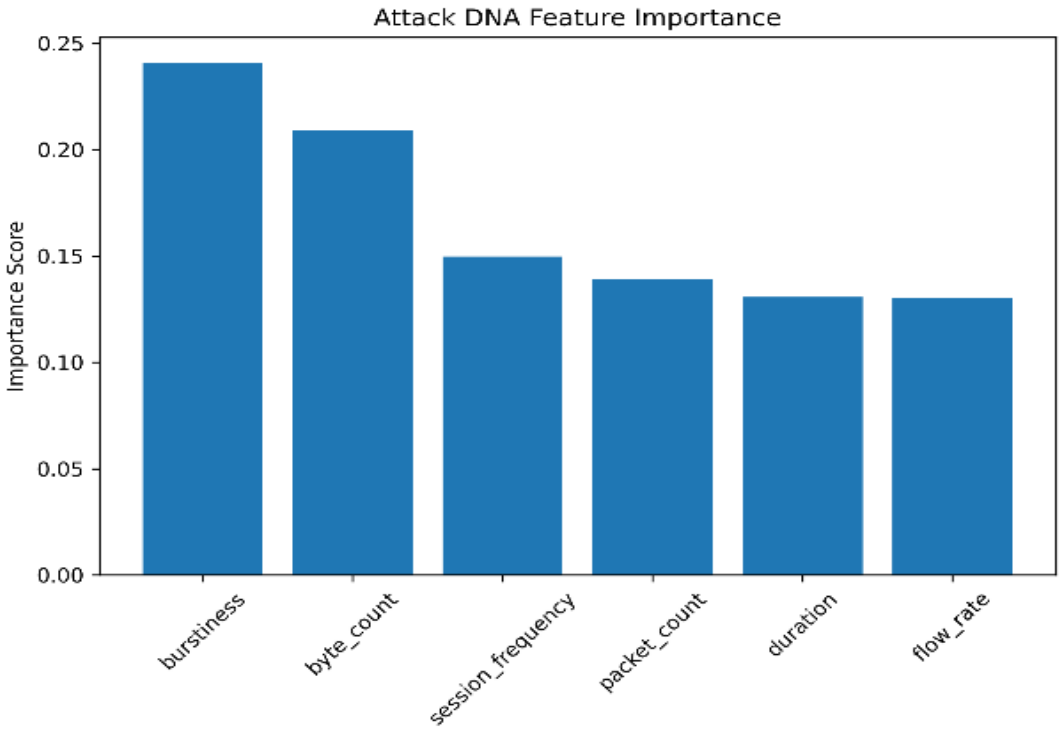

Fig. 4.11 Feature Importance Analysis

According to *Figure 4.11*, the most relevant characteristics, shown are the number of packets and the number of bytes, indicating that the amount and strength of the traffic might be important indicators of malicious traffic. Flow rate and length of time are also employed to help classify and record the action of flow and time aspects.

The suggested approach key strength is the recognition and deciphering meaningful features that provide transparency in decision making for the model. This improves trust and usability in real world cybersecurity applications [16]. Overall, the results indicate that meaningful behavioral patterns can be predicted well with the Attack DNA framework, which can be accurately, interpretatively, and scalable leveraged to detect cyber threats in diverse situations.

CONCLUSION

In this paper, Attack DNA, a behavioral abstraction framework for cyber threat detection over a heterogeneous set of intrusion detection datasets, is presented. The proposed approach included converting the data from four IDSs (NSL-KDD, UNSW-NB15, TON-IoT, CIC-IDS) into a common feature space, which comprised of duration, packet count, byte count, flow rate, session frequency, and burstiness. It has been shown in experimental results that RF outperforms LSTM, with an accuracy of 95.96%, whereas LSTM shows an accuracy of 70.19%, indicating that the former is better than the latter in the proposed behavioral representation, as it is a learning model based on features and not on sequences. The results confirm the effectiveness, scalability and interpretability of Attack DNA as a practical tool for cybersecurity analytics in a heterogeneous environment.

FUTURE WORK AND REAL-TIME DEPLOYMENT POSSIBILITIES

Research can further extend the Attack DNA framework to include more behavioral attributes that can represent more complex attack behavior. Possible features include protocol entropy, behaviour of the TCP flags, payload statistics, source-destination reputation indicators and advanced temporal features. Further advancements in the incorporation of explainable artificial intelligence techniques and boosting the use of ensemble learning methods could add to the detection accuracy and model interpretability.

The proposed framework also offers opportunities to deploy in real time. Future versions could also have the ability to combine with Security Information and Event Management (SIEM) systems like Splunk and Wazuh, facilitating ongoing risk tracking, automated alert generation, and incident evaluation. These streaming packet processing and online feature extraction mechanisms might enable near real-time intrusion detection in dynamic network environments.

Furthermore, it would be prudent to validate it in a cloud-native intrusion detection system, in a Security Operations Center (SOC) and on the edge. Transformer-based architectures, as well as other sophisticated deep learning models, could also be explored on a streaming network data source with increased temporal relationships, expanding Attack DNA to future generations of cybersecurity monitoring systems.

REFERENCES

1. M. A. Alohali, M. Aljebreen, N. Ahmad, S. Alahmari, S. S. Albouq, A. Alqazzaz, H. Alkhiri, and Y. Said, “Privacy Preserving Blockchain Integrated Explainable Artificial Intelligence with Two Tier Optimization for Cyber Threat Detection,” 2025.
2. W. Ali, A. Sajid, T. A. Ghodke, R. Malik, N. Malik, and K. Kaushik, “Honeypot Comparison of Attack Detection and Mitigation of SSH Attack,” 2024.
3. J. Bi, Z. Guan, H. Yuan, and J. Zhang, “Improved Network Intrusion Classification with Attention-Assisted Bidirectional LSTM and Optimized Sparse Contractive Autoencoders,” 2024.
4. P. Cheimonidis and K. Rantos, “A Proactive and Time-Sensitive Cyber Risk Assessment Model Integrating Markov Chains and Bayesian Networks,” 2025.
5. M. Fan, Z. Si, X. Xie, Y. Liu, and T. Liu, “Text Backdoor Detection Using an Interpretable RNN Abstract Model,” 2021.
6. J. Franco, A. Arış, B. Canberk, and S. Uluagac, “A Survey of Honeypots and Honeynets for IoT, IIoT, and CPS,” 2021.
7. Y. Imrana, Y. Xiang, L. Ali, and Z. Abdul-Rauf, “A Bidirectional LSTM Deep Learning Approach for Intrusion Detection,” 2021.
8. J. Kang, H. Yang, Y. Zhang, Y. Dai, M. Zhan, and W. Wang, “ActDetector: A Sequence-Based Framework for Network Attack Activity Detection,” 2022.
9. Q. Liu, Y. Zhan, and B. Wang, “Secure and Efficient Cloud-Based Multi-Party Private Set Intersection with Union Protocol,” 2025.
10. N. Naik, C. Shang, P. Jenkins, and Q. Shen, “D-FRI-Honeypot: A Secure Sting Operation Using Dynamic Fuzzy Rule Interpolation,” 2021.
11. M. Raza, M. J. Saeed, M. B. Riaz, and M. A. Sattar, “Federated Learning for Privacy-Preserving Intrusion Detection in Software-Defined Networks,” 2024.
12. P. Revathi and S. Kingslin, “A Review of Privacy-Preserving Intrusion Detection for Healthcare Edge-IoT,” 2026.
13. M. A. O. Rabah, H. Drid, Y. Medjadba, and M. Rahouti, “Detection and Mitigation of DDoS Attacks Using Ensemble Learning and Honeypots in SDN-UAV Architecture,” 2024.
14. R. Ben Said, Z. Sabir, and I. Askerzade, “CNN-BiLSTM: A Hybrid Deep Learning Approach for Network Intrusion Detection,” 2023.
15. X. Wang, J. Liu, and C. Zhang, “Network Intrusion Detection Based on Multi-Domain Data and Ensemble Bidirectional LSTM,” 2023.
16. C. D. Xuan, “Detecting APT Attacks Based on Network Traffic Using Machine Learning,” 2021.

17. Y. Yang, S. Tu, R. H. Ali, H. Alasmary, M. Waqas, and M. N. Amjad, “Intrusion Detection Based on Bidirectional Long Short-Term Memory with Attention Mechanism,” 2023.
18. J. Zhang, L. Pan, Q.-L. Han, C. Chen, and S. Wen, “Deep Learning Based Attack Detection for Cyber-Physical System Cybersecurity: A Survey,” 2022.