ZT-NIDS: Zero Trust, Network Intrusion Detection System

Abeer Z. Alalmaie[1][a], Priyadarsi Nanda[1][b] and Xiangjian He[2][c]
[1]*School of Electrical and Data Engineering, University of Technology Sydney, Sydney, Australia*
[2]*School of Computer Science, University of Nottingham, Ningbo, China*

Keywords: Zero Trust, Network Intrusion Detection, Network Security, CNN-BiLSTM, Attention, Cybersecurity.

Abstract: Zero Trust security can tackle various cyberthreats. Current trends in security monitoring must shift to a “never trust, always verify” approach, as data security is threatened when cloud-based third parties access network traces. Network Intrusion Detection System (NIDS) can be exploited to detect anomalous behaviour. Convolution Neural Network (CNN), Bi-directional Long Short Term Memory (BiLSTM) based classifiers and Auto-Encoder (AE) feature extractors have presented promising results in NIDS. AE feature extractor can compress the important information and train the unsupervised model. CNNs detect local spatial relationships, while BiLSTMs can exploit temporal interactions. Furthermore, Attention modules can capture content-based global interactions and can be applied on CNNs to attend to the significant contextual information. In this paper, we utilized the advantages of all AE, CNN and BiLSTM structures using a multi-head Self Attention mechanism to integrate CNN features for feeding into BiLSTM classifier. We use the bottleneck features of a pre-trained AE for an Attention-based CNN-BiLSTM classifier. Our experiments using 10, 6 and 2 categories NID system on UNSW-NB15 dataset showed that the proposed method outperforms state-of-the-art methods and achieved accuracy of 91.72%, 89.79% and 93.01%, respectively. Plus, we introduced a balanced data sampler for training 10 categories of NIDS.

1 INTRODUCTION

Today, activities of many private sectors and government organizations depend on technology. Thus, due to the surge in number of attacks, the cyber resilience of crucial infrastructures is vital. Cybersecurity is the establishment and compilation of methods to defend cyberspace from incidents. Hence, it is considered as a crucial element in private and government industries. Undoubtedly, technology and cybersecurity are continuously evolving for better user experience. In the same vein, attackers are also advancing their techniques to exploit the critical infrastructures for unauthorized access. Therefore, it is important to prevent, detect, and respond to cyberattacks quickly, so the damage to the critical infrastructures can be reduced (Dabbaghzadeh, et al., 2021). So, many organizations have initiated to improve cybersecurity awareness (Newhouse, et al., 2017).

Traditionally, organizations focused on perimeter defence and give authorized access to network traffic on the internal network. Hence, unauthorized lateral movement within the environment has been one of the biggest challenges for organizations networks. Perimeter firewalls are less useful for detecting and blocking attacks from inside of the network and cannot protect subjects outside of the enterprise perimeter. Network perimeters security alone can no longer be effective for providing enterprise security and organizations should incessantly focus on improving cybersecurity management systems. There is a persistent increment in types of cyber-attacks on the ICT structure with the subsequent major propensities (Sanandaji and Falahati, 2021).

Maturity models are used as consultants and reference structures for information system management in various types of companies. A Cybersecurity Maturity Model (CMM) is a standard to determine the maturity of a protection system and guidance on how to attain the higher level. Therefore,

[a] https://orcid.org/0000-0001-7733-8163
[b] https://orcid.org/0000-0002-5748-155X
[c] https://orcid.org/0000-0001-8962-540X

Alalmaie, A., Nanda, P. and He, X.
ZT-NIDS: Zero Trust, Network Intrusion Detection System.
DOI: 10.5220/0012080000003555
In *Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT 2023)*, pages 99-110
ISBN: 978-989-758-666-8; ISSN: 2184-7711

some cybersecurity maturity mechanisms have been suggested to reduce the impact of cyberattacks, considering country level protection. Moreover, regional, and international organizations have conducted other studies, but they aim on score and ranking countries corresponding to their national-level cybersecurity top systems. CMM aims to find methods for developing the company's protection and introduce a level-based advancement approach.

Zero Trust satisfies these significant characteristics by treating all users, devices, data, and service requests similarly. It shifts from the traditional security policy of all assets in an organization being open and accessible to requiring continuous authentication and authorization for any asset to be accessible. Most existing corporate networks are flat. The weakness of the traditional hub-and-spoke network model lies in its architecture. Crossing the chasm from trust to distrust via a firewall is inherently risky. Instead, Zero Trust no longer distinguishes between “inside” and “outside” the network perimeter. A Zero Trust Architecture (ZTA) addresses this trend by focusing on protecting resources, not network perimeters, as the network location is no longer viewed as the prime component to the security posture of the resource.

In other words, Zero Trust is a set of cybersecurity principles used to create a strategy that focuses on moving network defences from wide, static network perimeters to concentrating more narrowly on users, systems, and individual or small groups of resources.

In this respect, we are eliminating the concept of trust within the network and believe there are no more trusted interfaces, users, packets, and applications. Across the industries, security professionals are shifting the security diameter to zero security trust state of mind and quick adopting the approach of Zero Trust security network model. It is notable that Zero Trust is more than just concept. It is a robust security model that follows 7 security important principles: Data, Devices, Workload, Automation and Orchestration, Visibility and Analytics, Users, and Network (Alalmaie, et al., 2022).

The rest of the paper is organized as follow. In section 2, some of the main related works are studied. In section 3, the Zero Trust concept is explained briefly. The chosen approach and the proposed method are presented in sections 4 and 5, respectively. In section 6, the used evaluation methods are mentioned. Next, experimental results are reported in section 7. Eventually, in section 8, a conclusion and some future works are given.

2 RELATED WORKS

The Systems Security Engineering Capability Maturity Model (SSE-CMM) was designed on the CMM paradigm to evaluate assurance of security engineering procedures and capability for customers as a standard mechanism. THE SSE-CMM prevents organizational security by having various features (Nahari and Krutz, 2003).

The Community Cyber Security Maturity Model (CCSMM) was proposed to tackle various problems related to information sharing, metrics, training, testing, technology, random threats, structured risks, and well-structured risks. The CCSMM is designed based on comprehensive working experience with communities/states emerging and implementing cybersecurity practices, and it has five maturity levels as security awareness, process development, information availability, tactics development, and full security operational capability (White, 2011).

In (Eusgeld, et al., 2011), researchers examined the weaknesses associated with the combination of industrial control systems and the underlying critical infrastructures. Plus, an integrated model was suggested to address the selection of the suitable method, and it investigate interdependencies between the critical infrastructures.

In (Mettler and Blondiau, 2012), authors proposed a maturity model to measure the skill levels of federal important infrastructure safety attempts, taking the maturity measures based on the obtained core causes into consideration. The model did analyse of data concerning to nationwide cybersecurity developments through base level technique to obtain the fundamental reasons of the vulnerability of vital infrastructures to cyber risks. More, it confirms the maturity standards by proposing the effects to specialists according to a Delphi survey.

In (Karabacak, et al., 2016), authors have focused on the advancement of a wide-ranging maturity mechanism to apply in the Hospital Information System (HIS) environment. Thus, they discussed on different problems, like most important influencing factors by information system managers, which are associated to the maturity phases and maturity-affecting element assessment in the perspective of the maturity periods. Plus, they suggested HIS Maturity Model (HISMM). They prepared a questionnaire to understand the critical infrastructure requirements, using the design science research methodology. Further, they conducted a survey, considering different categories and designed an initial maturity model. Moreover, they implemented a qualitative assessment technique for interviews to attempt and analyse the existence or the unavailability of one or many attributes in the text. The final maturity model stage was designed based on data analysis, strategy, people, systems and its infrastructure, electronic medical record, and information security.

In (Nussbaum and Park, 2018), they noticed that the difficulties in traditional IT agreement are generally obvious in cybersecurity, as they have exclusive cybersecurity products and services. Thus, they suggested a prototype, explaining a set of measures and restrictions that have influence on the decision-making process of local government towards the cybersecurity agreement. Further, they discussed different conditions for government contract and challenges of cybersecurity contracting.

In (Saleem, et al., 2019), authors suggested a multidimensional holistic model to confirm protection through all aspects of the information resilience, by innovative technologies, intellectual procedures, and constant evaluations. This layered defence model was integrated to assess the security strengths of critical infrastructures and suggest the best procedures for cybersecurity analysts towards a robust method before combining third-party products.

In (Renteria, et al., 2019), authors came up with an enabler-founded digital administration maturity framework, considering three criteria as digital government, multidimensional, and availability. Hence, this model is comprised of 7 aspects: leadership, regulatory regime, and data to provide more precise and specific suggestions, strategy, governance, technology, and organization.

In (Gourisetti, et al., 2019), authors designed the cybersecurity weakness improvement framework through empirical model (CyFEr), using multi-scenario criteria-based vulnerability analysis and multitiered constraint-based optimization. CyFEr was implemented on the NIST cybersecurity structure, checking with a practical cyberattack. It also provides potential solutions to focus on the user requirements to reach an essential cybersecurity maturity, finding the top-ranking solution based on the scalar values.

In (Niazi, et al., 2020), authors suggested a Requirement Engineering Security Maturity Model (RESMM) to support software-based industries, suggesting specific requirements in a better way for secure software advancement. They studied to know the security needs and improved Sommerville's requirements engineering procedures. Further, they also did a questionnaire study, considering identified requirements into account. Eventually, they considered related works, improved Sommerville's procedures and feedback from the security experts.

In (Bazi, et al., 2017), authors suggested a cloud migration maturity model to ensure dynamism in cloud migration procedures. It helps managers for a complete migration overview to present a strategic plan, achieving an effective management.

As weaknesses, maturity mechanisms only offer a least compliance framework instead of a required cybersecurity standard that work in an emerging cyber environment. Thus, the mechanism should be practiced not only by the management, but by security specialists to evaluate the complete protection of the group/structure and considering measure to reinforce limitations of any explicit features of the organisation as recognised by the valuation.

3 ZERO TRUST

The proliferation of cloud computing, mobile device use, and the Internet of Things has dissolved traditional network boundaries. Hardened network perimeters alone are no longer effective for providing enterprise security in a world of increasingly sophisticated threats. Zero Trust concept compiles strict identity-based verification for every user and device trying to access resources on a private network as well as cloud, regardless of whether they are sitting within or outside of the network perimeter. No single specific technology is associated with Zero Trust; it is a holistic approach to network security that incorporates different principles and technologies. ZTA is designed to protect digital environments by leveraging network segmentation, preventing lateral movement, providing Layer 7 threat prevention, and simplifying granular user-access control.

Zero Trust is a cybersecurity paradigm focused on resource protection and the premise that trust is never granted implicitly but must be continually evaluated. Zero Trust implies not to trust any entity inside or outside of network perimeter at any time. It is focused on eliminating trust from within an organization, dictates that no implicit assumptions should be made about the credibility of users, devices, applications, or data accessing or being accessed on an organization's network. It provides the visibility and IT controls needed to secure, manage, and monitor every device, user, application, and network belonging to or being used by the organization and its employees to access data. Zero Trust scrutinizes any incoming or outgoing traffic. The difference between this and other security models is that even internal traffic, meaning traffic that does not cross the perimeter of the organization, must be treated as a potential danger as well.

Zero Trust provides an occasion for scalable protection structure throughout numerous distinct associations. In (Kindervag, 2010) stated that keeping trust on the cloud and networks is too critical job and then, suggested that it is better to eliminate the idea of trust. Further, author proposed a Zero Trust mechanism to improve the protection technologies for forthcoming changeability. By looking at failures inside organizations to stop cyberattacks, especially lateral movements of threats inside their networks, they realized the traditional security model operated on the outdated assumption that everything inside an organization's network could be trusted. Instead, Zero Trust inverts model, directing IT teams according to the guiding principle of "never trust, always verify" and redefining the perimeter to include users and data inside the network. In this broken trust model, it is assumed that a user's identity is not compromised, and all users act responsibly and can be trusted. In the Zero Trust model trust is a vulnerability. So, network users (threat actors and malicious insiders) are free to move laterally and access/exfiltrate whatever data they are not limited to.

In ZTA, none of the pillars are reliable, no question what kind of unit it is, even though it is part of the network. Zero Trust absolutely redefines the method to resource separation – a fundamental theory where resources must be remained safe, are categorized all together, and separated carefully or kept unconnected from illegal contact at any type of areas. These mechanisms also present the occasion to micro-segment systems, allowing groups to change their demands with not reforming their whole network. In micro-segmentation, networks are carved into small granular nodes all the way down to a single machine or application. Security protocols and service delivery models are designed for each unique segment. The free flow of data that was once one of the cornerstones of the Internet needs to be confined to protect networks from penetration, customers from privacy violations, and organizations from attacks on infrastructure and operations.

ZTA is an end-to-end approach to network/data security that encompasses identity, credentials, access management, operations, endpoints, hosting environments, and the interconnecting infrastructure. The focus should be on restricting resource access to those with a "need to know." Traditionally, agencies have focused on perimeter defence, and authorized users are given broad access to resources. Hence, unauthorized lateral movement within a network is one of the biggest challenges for organizations. The Trusted Internet Connections (TIC) and agency perimeter firewalls provide strong Internet gateways. This helps block attackers from the Internet, but the TICs and perimeter firewalls are less useful for detecting and blocking attacks from inside the network. Thus, the Zero Trust separation access framework can be thought the next generation mechanism of a firewall, expanding micro subdivision of the networks to achieve flexibility, scalability, and virtualization easiness.

4 THE CHOSEN APPROACH

Our model is proposed against a semi-honest adversary who will be conducting the task of IDS but might have various incentives to disclose the identity or the exact value of some attributes of some records. Our assumptions are: Firstly, the adversary is semi-honest. Secondly, the adversary can observe the anonymized version of the data. The adversary also knows the underlying anonymization algorithm. Thirdly, the adversary knows the original version of a subset of the records. He/she has collected this information through running some semantic attacks. Accordingly, we assume that the adversary knows the original prefix values of α% of the prefixes. Finally, the adversary objective is to find the original version of the most possible number of the anonymized records. We call these estimations of the adversaries of the original version of the anonymized records "matches". The larger the number of true matches is, the higher the adversary advantage will become, and the higher privacy leakage incurs.

4.1 IDS in Third-Party Setting

Undeniably, conducting reliable IDS greatly depends on the accuracy of the received dataset compared to its original version. Specifically, most of detection algorithms rely on learning the benign from the high-dimensional and large-size datasets. So, only defence mechanisms with minimal modification to the format/entries in the dataset can lead to trustworthy intrusion detection. Beside the existing homomorphic encryption and computation over aggregated data techniques, Multi-View (MV) approach (Mohammady, et al., 2018) has been introduced as an effective method to maximally benefit the best of both worlds. This approach outputs a prefix preserving version of IP address attributes, and an accurate copy of values in other attributes. Such a minimally modified version of the network traces can be used to accurately conduct analyses. Moreover, is ensured through hiding that real view among a set of indistinguishable fake views.

4.2 Defence: The MV Approach

We first need to formally introduce the prefix preserving anonymization PP (-, *K*) which is basically a cryptographic mapping function like CryptoPAn (Mohammady, et al., 2018) that relies on a secret key *K*. The most important property of this function is that it can preserve the prefixes of numeric-value attributes. Thus, if two real addresses share first *X* bits, e.g., 150.10.10.1, 150.10.20.1 share 14 bits in their prefixes, are mapped to another two anonymized addresses, e.g., 97.61.5.252, 97.61.5.252, which share first *X* bits. As we showed in previous sections, PP is vulnerable to different classes of semantic attack, and the MV approach presented in the following was designed to secure its output. The schema of the MV approach is presented in Figure 1.

There are 7 steps to this approach, the first five are initiated on the data owner side and the final three steps involve the data analyst side. We note that this approach assumes that the confidential attribute is the IP address, and if IPs are kept secure, the adversary cannot infer any sensitive information.

4.2.1 Implementing Effective Zero Trust Model at the Data Owner Side

Step 1: Data owner generates two 256 bits Cryptographic keys (*K0* & *K1*), and the original data is anonymized using PP (-, *K0*).
Step 2: The anonymized trace is partitioned.
Step 3: Each partition is anonymized but repeated for a different number of times at different partitions. Therefore, seed trace is not prefix-preserved.
Step 4: The seed trace and some supplementary parameters are outsourced to data analyst. The pseudo vector and the seed view generation are designed such that after “r” number of times view generation, the real view (which is identical to the view in step one will be retrieved.

4.2.2 Accuracy Guarantee at the Data Analyst side

Step 5: Analyst generates N views based on seed view and supplementary parameters.
Step 6: Analyst analyses all N views and generates corresponding reports.
Step 7: Data owner retrieves report corresponding to the “real view”, using a private information retrieval (PIR) protocol in a way that analyst cannot identify which view was retrieved.

Clearly, the quality of the views generated in the MV approach is the main factor of the adds-on confidentiality the network trace will receive. Specifically, if all other (fake) generated views are too far or too close from the original trace (prefix-wise and in the presence of some adversary knowledge), the MV approach may end up compromising a high level of privacy. In particular, the adversary can discard many of the (far) fake views by looking at the prefix relations in IP addresses and compare them to his/her adversary knowledge to identify any inconsistency. Conversely, a design with fake views generated too close to the real view incurs drastic privacy leakage. In the latter case, the fake views are not fake. Therefore, in (Mohammady, et al., 2018), authors proposed a metric called “the indistinguishability” to reflect the distance of each view from the real view. Based on this formalization, they suggest two schemes for their partitioning algorithms, i.e., IP-based, and distinct IP based partitioning: each with their own partition sizes. They conclude that a distinct IP based partitioning with a customized pseudo vector can significantly (50 times less) reduce the privacy violation. We will detail on these methodologies in our next report.

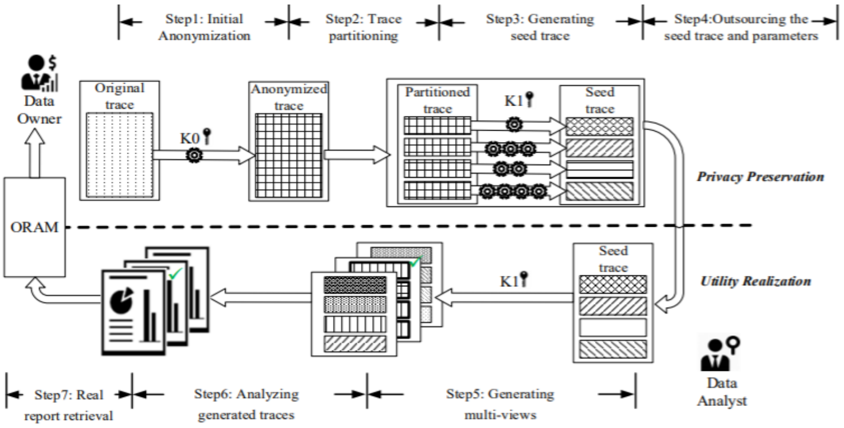
Step1: Initial Anonymization
Step2: Trace partitioning
Step3: Generating seed trace
Step4:Outsourcing the seed trace and parameters
Data Owner
Original trace
K0
Anonymized trace
Partitioned trace
K1
Seed trace
Privacy Preservation
ORAM
Seed trace
K1
Utility Realization
Data Analyst
Step7: Real report retrieval
Step6: Analyzing generated traces
Step5: Generating multi-views

Figure 1: MV approach (Mohammady, et al., 2018).

4.3 Operations Required for IDS

The accuracy of the required analyses must be guaranteed alongside its confidentiality because the first motivation of such outsourcing models is to benefit from more accurate analysis tasks. For this purpose, we believe the MV approach is an ideal candidate. In contrast to other solutions, MV allows third-party analysts to receive and interact with the raw data. Nevertheless, as MV anonymization entails, the original versions of some of the attributes must be replaced with a cryptographic prefix preserving version. We now shed light on the impact of such transformation over the outcomes of certain classes of IDS algorithms. Since our focus is network traces, and we only apply PP on the IP address attributes, all traffic-level analyses (packet level) are expected to return trustworthy results, e.g., the traffic of the entire network, or the traffic on a certain port. With a similar reasoning, the flow level analyses also remain intact. The main issue regards the graph-level set of analyses, e.g., subnet-based analyses, throughputs of a subnet, and reachability analyses.

Recently, solutions with Deep Neural Networks (DNN) and optimisation algorithms have undertaken machine learning approaches in various applications (Abolghasemi, et al., 2022; Ghezelji, et al., 2022; Tohidi, et al., 2022), due to their success (Tohidi and Rustamov, 2020), and IDS is not an exception.

In (Kumar, et al., 2020), a misuse-based intrusion detection system was proposed to detect 5 categories in a network called: Probe, Exploit, DOS, Generic and Normal. This system could act as a firewall with some extra information added to it. Moreover, unlike most related works, in their paper UNSW-NB15 dataset was considered as the offline dataset to design own integrated classification-based model for detecting malicious activities in the network. Plus, they generated their own real-time data set at NIT Patna CSE lab (RTNITP18) that acted as the working example of their intrusion detection model.

In (Cao, et al., 2022), a network intrusion detection model was introduced which fused a CNN and a gated recurrent unit. Their main was to tackle the problems associated with the low accuracy of existing intrusion detection models for the multiple classification of intrusions and low accuracy of class imbalance data detection. In their method, a hybrid sampling technique combining adaptive synthetic sampling and repeated edited nearest neighbours was applied for sample processing to solve the positive and negative sample imbalance issue in the original dataset. The Feature Selection (FS) was carried out by combining RF algorithm and Pearson correlation analysis to address the feature redundancy problem. Afterwards, the spatial features were extracted by using a CNN, and then extracted by fusing Average-pooling and Max-pooling, by using attention mechanism to assign different weights to the features, therefore, reducing the overhead and improving the model performance. At the same time, a Gated Recurrent Unit (GRU) was applied to extract the long-distance dependent information features to achieve comprehensive feature learning. Eventually, a Softmax function was used for classification.

We consider employing such NN-IDS in our ZTA-IDS, and therefore, we will elaborate on the accuracy of these special types of IDS.

4.4 Criteria on Security and Accuracy

Clearly, the characteristics of the framework, rely on the IDS task. If the IDS task is independent of the prefix relation of the IP addresses, e.g., counting the number of packets with the size of larger than 300KB, then we can strengthen the confidentiality aspect of the solution arbitrarily. On the other hand, if the IDS task depends on the trustworthiness of the prefix relations, an appropriate degree of MV must be applied. Interpreting one IDS task's requirements is one of the main contributions of our ZTA_IDS framework. For an NN-IDS, depending on the type of intrusion, the learning module may require fingerprinting a larger number of attributes to predict the malicious activities from the benign ones. Therefore, partitioning algorithm, number of partitions, and number of attributes involved in the partitioning must be carefully selected to guarantee a maximal level of accuracy for an NN-IDS. These are the most important parts of our experiments which we will elaborate on in the next report.

4.4.1 Privacy Preservation

On the data owner side, ZTA_IDS performs like the MV approach. The only modification is that the partitioning algorithm (mainly the parameters) is defined based on both security and accuracy requirements. An appropriate choice of partitioning could vary from distinct IP based partitioning for the prefix groups of with length three octets when conducting anomaly-based IDS, to distinct IP based partitioning for the prefix groups of with length only one or even less when conducting packet/flow level analyses. The overview is depicted in Figure 2.

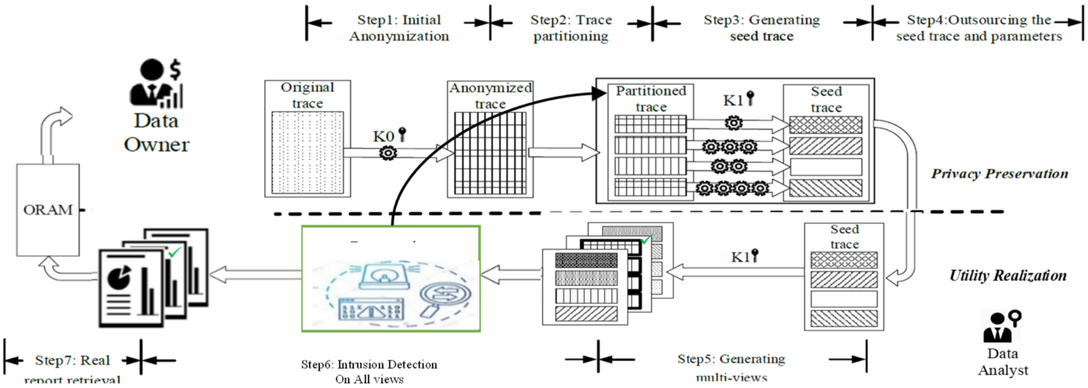
Step1: Initial Anonymization
Step2: Trace partitioning
Step3: Generating seed trace
Step4:Outsourcing the seed trace and parameters
Data Owner
Original trace
K0
Anonymized trace
Partitioned trace
K1
Seed trace
Privacy Preservation
ORAM
Seed trace
K1
Utility Realization
Data Analyst
Step7: Real report retrieval
Step6: Intrusion Detection On All views
Step5: Generating multi-views

Figure 2: Overview of the ZTA-IDS approach.

4.4.2 Actions on Data Analyst Side

On this side, ZTA-IDS follows the MV approach to generate different dataset views. Next, the analyst pre-processes all views to a format suitable for the operations. Plus, the analyst may need to reorder the entire trace based on time after generating the views. The prefix-to-prefix communication is another set of queries for which the analyst must run some pre-processing to reduce the impact of anonymization. He/she will then run the IDS algorithms on each of views and returns the outcomes to the data owner.

5 THE PROPOSED MODEL

The architecture consists of 4 modules: AE feature extractor, Convolution blocks, Attention mechanism and LSTM layers. The proposed model is inspired from AE-CNN for binary NID. Thus, we use the extracted features as the input to our neural network for categorical intrusion detection. We also utilize Attention module to focus on important features and LSTM layers to handle temporal dynamics.

5.1 AE Feature Extraction

The compressed bottleneck features of network traces are extracted via a pre-trained deep AE. The AE bottleneck layer maps the original input into a compressed representation where the input features are more correlated. AE consists of 2 components: the encoder that compress input features, and the decoder which is discarded after pre-training. Thus, a deep AE can be used to extract a combined and compressed feature from network trace attributes. In the AE, the bottleneck feature z is extracted using the encoder function from the original data X. The decoder function - maps the bottleneck z to the output $\hat{X}$ The decoder is expected to reconstruct the input as Eq. (1).

$$\psi = X \Rightarrow z\,,$$
$$\phi = z \Rightarrow \hat{X}\,, \tag{1}$$
$$\psi\,,\phi = argmin\left\|\hat{X} - (\phi(\psi(X)))\right\|^2$$

Mean Squared Error (MSE) loss function of the AE is as Eq. (2). Where $X - \hat{X}$ is usually averaged over a mini-batch input training set. W, W_0 are weight matrices and b, b_0 are bias vectors for encoder and decoder, respectively. Bias is not used for encoder part to aggregate input feature only.

$$L(X,X') = \left\|X - \hat{X}\right\|^2 = \|X - \sigma(\,W_0(\sigma(WX+b)) + b_0)\|^2 \tag{2}$$

In the structure of the AE (Figure 3), dotted lines are discarded after training the AE. The bottleneck features of the trained AE, which are more spatially related, are used as input to CNN-LSTM.

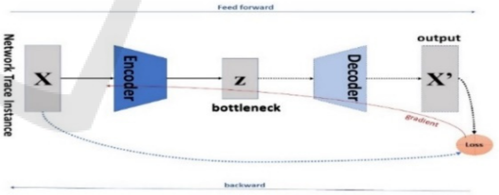

Figure 3: AE feature extractor for network traces.

Since AE with a bottleneck layer accepts any numerical value and compresses the information available in the input numerical values, pre-processing and FS is not needed.

5.2 Convolutional Neural Network (CNN)

We use a CNN to consider spatially related features extracted using the AE. The classifier is applied on a bottleneck feature extracted from a trained AE for NID because CNNs work well with data that has a spatial relationship. The CNNs are also known to be good feature extractors because of local convolution filters, repetitive filters among whole input data, and pooling layers which make it robust. Here, we also use a tuned 1D CNN to handle spatial dependencies within traces of data. Our proposed CNN structure is shown in Figure 4, and LeakyReLU [4] with 0.2 negative slope is considered as activation function for hidden layers. In convolution layers, the first number is for filters and the number in parentheses is the convolution filter size, e.g., first layer has 128 filters, where 11 is the convolution filter size. A pooling with size of 2 is only applied on the first convolution layer. In the CNN layer output, we have 256*5 features, which its knowledge needs to be aggregated together, since it has a high dimension of feature vector to feed into any regular layer.

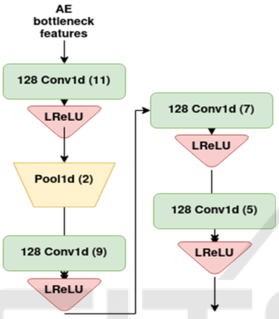

Figure 4: The proposed CNN to handle spatial dependencies of network traces.

5.3 Attention Module

We apply a multi-head Self Attention module to aggregate the information available in extracted features and handle the relation between the CNN features and LSTM components (subsequent layers). The attention module dimension is the number of channels from the last CNN layer (256) and uses 8 heads. The output is mapped into 64 dimension to limit the features. The attention learns to focus on intrusion related features. The alternative structure for multi-head self-attention on top of CNN, would be a linear Flatten layer, which maps the CNN multi-dimensional features into one large dimension. The total number of features in this layer, is the same as the total number of CNN features in all dimensions. We also report the results of the proposed method with a linear Flatten layer instead of Attention mechanism. Plus, we use BiLSTM layers after the Attention to handle the temporal dynamics between the sequences of network traces.

5.4 BiLSTM Classifier

Since LSTMs can hold or forget information for a long time, we propose to use LSTMs to handle the temporal dynamics. Also, BiLSTM is able to take forward and backward sequences into consideration which can be important in handling temporal dynamics. We use 2 BiLSTM layers with 128-dimensional representations. A dropout with probability of 0.2 is applied between 2 layers of BiLSTM. Finally, a linear layer with the number of neurons same as target categories is applied (from 2 to 10). In the subsequent paragraphs, we will review the training and test conditions along with the evaluation results.

6 EVALUATION METHODS

In this section, we explain the chosen evaluation metrics and data.

6.1 Training Setup

All experiments are implemented in PyTorch and conducted on Colab platform with a batch size of 32. The AE is trained to minimize MSE criterion as loss function, which is also known as reconstruction error. Both encoder and decoder parameters are considered and trained independently. The dimension of the bottleneck features is considered as 64, which is compact enough to compress input features. We used the Adam optimizer with learning rate of 1e-4 and weight decay of 1e-5 to minimize the reconstruction loss. The model is trained until no more improvement is possible according to validation results. All data attributes are normalized to numerical values between 0 and 1. Thus, non-numerical attributes are converted into numerical values using one-hot encoding. The training dataset is set to prevent over-fitting.

The bottleneck features extracted from the trained AE are fed into the CNN for further training and processing while the AE is frozen. Since the compact features of input attributes are available in the bottleneck layer with 64 neurons, the CNN input spatial dimension is 64 and the sequence number equals the size of mini batch.

The CNN modules are trained with learning rate of 1e-3, while the Attention and BiLSTM modules are trained with learning rate of 1e-4 for a maximum 50

[4] https://pytorch.org/docs/stable/generated/torch.nn.LeakyReLU.html

epochs. We used cross-entropy loss as the classifier loss function and Adam optimizer.

6.2 Data

We evaluate the proposed method on the UNSW-NB15 dataset (Moustafa and Slay, 2015), which is comprised of a hybrid of real modern normal activities and synthetic contemporary attack behaviours. This is an upgraded version of the KDD cup dataset which is a more balanced dataset. The UNSW-NB15 dataset contains 10 classes, namely: Normal, Fuzzers, Analysis, Back-doors, DoS, Exploits, Generic, Reconnaissance, Shell Code, and Worms. We use the train and test subsets of this dataset with 175343 and 82337 records respectively. In this dataset each record has a 42-dimensional feature, which 3 features of them are non-numerical values and need pre-processing to be fed into Neural Networks since the input should be a digital matrix. These 3 features are protocol, service, and state with 133, 13, and 11 symbol attributes, respectively.

One-hot encoding is used to map non-numerical attributes of the data set to numerical feature vectors. In total, the pre-processed input feature size would be 196. Then, the features are normalized between 0 and 1 which are used to train the AE unsupervised. The 64-dimensional bottleneck features extracted from the trained AE are used for next experiments.

Since classes of attacks are unbalanced, most studies reduced it by merging or removing some categories. Binary classification means 9 categories are merged into 1 class as Intrusion, consequently the classes would be Intrusion/Non-Intrusion in this scenario. However, some works tried to merge the categories that are not far from each other. Some others reduced the imbalance data by removing the categories with fewer number of existing items including Backdoor, Analysis, Shellcode, Worms and sometimes Fuzzers which make the data imbalance.

We compare the results of 10 categories classification with related works and report the result of removing imbalanced data attributes to have a fair assessment. To show the advantage of the proposed structure we report the binary classification results using the same data structure, which train, and test data are used in reverse, so, have a few training samples. Additionally, we propose a nearly balanced sampling procedure to enhance the detection of the categories with fewer samples in the CNN module. Due to the sequential nature required to train the LSTM, we cannot use any sampling strategy to train it. The UNSW-NB15 dataset is highly imbalanced, the Normal category has 56000 samples for training while the Worms category only has 130 samples. We reduce the impact of this imbalance by sampling based on a smoothing probability function as Eq. 3.

$$P(cl_i) = \frac{\#cl_i - (1 - \frac{\min \#cl_i}{\#cl_i} + \epsilon) median}{\sum_{j=1}^{10} \#cl_i - (1 - \frac{\min \#cl}{(\#cl_j} + \epsilon) median} \qquad (3)$$

In this equation, cl_i means ith category, so $P(cl_i)$ is the probability of choosing a sample from ith category, calculated using number of samples in each category $(\#cl_i)$ and median of the number of samples per category. We use a small ϵ (0.1) to prevent zero addition for the category with minimum number of samples. The minimum number of samples is 130 associated with category Worms, and the median is 11378. The proposed sampling strategy keeps the ordering of the number of the categories but make the sampling more balanced by reducing the distance between number of items in each category.

7 EXPERIMENTAL RESULTS

In this section, we report the experimental results of NID methods for ten, six, and two categories.

7.1 Ten Categories CNN-BiLSTM Data Classification

For hyper parameter optimization, we explored the optimal number of layers and neurons for each part of CNN-BiLSTM with Attention module. To evaluate the effect of each module, the results of BiLSTM, CNN-BiLSTM, and Attention-based CNN-BiLSTM are compared to related methods in Table 1 for 10 categories data classification. Since the results of the baseline CNN model using AE features are not available for categorical classification, we implemented it and reported the results for

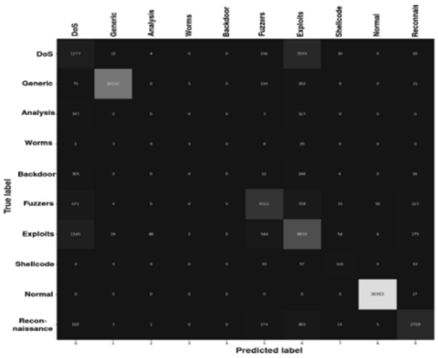

Figure 5: Confusion matrix of ten categories for CNN-BiLSTM with AE feature extractor.

comparison. Attention-based CNN-BiLSTM using AE bottleneck features out-performed other related works using deep learning methods for 10 categories NID. The confusion matrix is shown in Figure 5.

Accordingly, most errors of Analysis, Backdoor and Exploits Attacks are misclassified as DoS. More, Fuzzers and Exploits are misclassified interchangeably. None of Analysis and Backdoor records are predicted correctly. Only 3 records of Worms class are predicted correctly. Consequently, removing imbalanced data attributes including Backdoor, Analysis, Shellcode and Worms should obviously improve the accuracy. Thus, removing Fuzzers may lead to better accuracy.

7.2 Pre-Train CNN Using Balanced Data Sampler

We use a balanced data sampler to pre-train the CNN for later usage in CNN-BiLSTM with Attention module instead of reducing the number of categories. Our goal is to improve the discrimination of the model to learn discriminate the data better even when the number of training samples are imbalanced. Hence, we use the trained CNN in CNN-BiLSTM with Attention module to enhance the detection of network intrusion. The accuracy results of CNN-BiLSTM with Attention module with and without balanced sampler are compared in Table 1.

Table 1: Accuracy of 10 categories NID on test data.

Method	Accuracy
BiLSTM	77.46%
CNN	78.23%
CNN-BiLSTM	78.76%
CNN-Attention + BiLSTM	87.76%
CNN-Attention	88.13%
Proposed method (Standard Sampler)	87.76%
Proposed method (Balanced Sampler)	**91.72%**

We compare Recall, Precision and F-measure of the proposed method (balanced sampler) in Table 2.

Table 2: Results of NID on test data.

Method	Precision	Recall	F-measure
Decision Tree C5 (Kumar, et al., 2020)	-	75.8%	75.54%
Rule-based (Kumar, et al., 2020)	-	65.21%	68.13%
FS + ANN (Kasongo and Sun, 2020)	79.50%	77.53%	77.28%
Proposed method	60.24%	78.5%	62.62%

As Table 2 presents, the recall of the proposed model outperforms related works which means that it detects most of the relevant results.

It is seen that pre-training CNN using balanced sampler outperforms related works in terms of accuracy. The confusion matrix of the proposed method (10 categories) is shown in Figure 6. The number of misclassifications for each category is low, especially for Normal category.

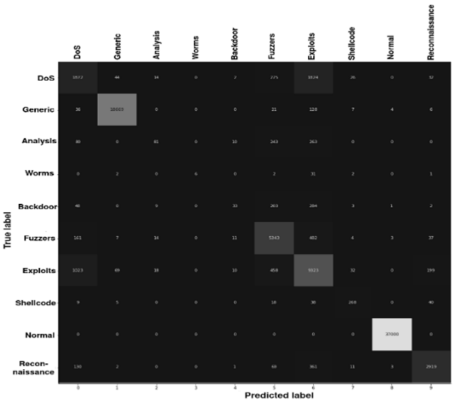

Figure 6: Confusion matrix of pre-trained CNN via balanced sampler, Attention and BiLSTM with AE feature extractor.

7.3 Six Categories CNN-BiLSTM Data Classification

Since recent works removed imbalanced data and reported 6 categories, we have removed 4 imbalanced categories too. The results are reported in Table 3.

Table 3: Accuracy of 6 categories NID on test data.

Method	Accuracy
CNN	82.01%
BiLSTM	83.11%
MLP + IGRF-RFE (Yin, et al., 2022)	84.24%
Rule-based (Kumar, et al., 2020)	84.84%
CNN-GRU + RFP (Cao, et al., 2022)	86.25%
CNN-BiLSTM	86.28%
CNN-Attention	87.54%
CNN-Attention + BiLSTM	**89.79%**

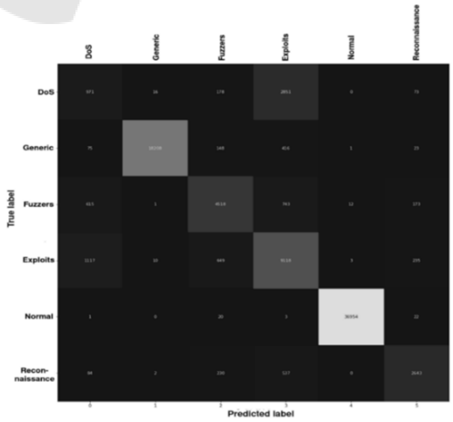

Figure 7: Confusion matrix of 6 categories for CNN-BiLSTM with AE feature extractor.

As seen, the proposed method outperforms related methods for 6 categories classification (confusion matrix in Figure 7).

7.4 Binary CNN-BiLSTM Data Classification

The hyper parameters for binary classification model have been kept same as the multi-class model. To evaluate each module, the results of BiLSTM, CNN-BiLSTM, and Attention-based CNN-BiLSTM are compared to the most related work using CNN with AE bottleneck features for NID in Table 4.

For a fair comparison, our data should be similar. We used train and test data interchangeably to have a fair comparison with CNN and AE method.

As Table 4 depicts, using BiLSTM decreases the accuracy, especially in combination of CNN. It can be due to the high dimension of CNN output, which is fed into the BiLSTM layers. However, using an Attention module on CNN to aggregate the CNN features for feeding into BiLSTM layers outperformed CNN and BiLSTM models. Since other binary classification methods using original train and test dataset for NID have reached almost 100% accuracy, more experiments are not needed.

Table 4: Accuracy of NID (binary classification).

Method	Accuracy
CNN-BiLSTM	78.93%
FS + DNN (Kanimozhi and Jacob, 2019)	89%
BiLSTM	90.84%
CNN	92.23%
CNN-Attention + BiLSTM	**93.01%**

According to the results, CNN and BiLSTM both perform well for NID using AE bottleneck features. However, an Attention module is needed to handle the relation between the components of these two structures and compose them together.

8 CONCLUSIONS

In this paper, we have introduced a new CMM with Zero Trust. The proposed model effectively protects various systems, crucial infrastructure, networks, data, services, end-users from critical security risks, achieving different security requirements.

Further, we have proposed an Attention CNN with Bi-directional Long Short Term Memory (CNN-BiLSTM) using Auto-Encoder Bottleneck features for Network Intrusion Detection System. We utilized the compressed bottleneck features of the Auto-Encoder. We also used a CNN to consider the spatial relation between extracted features. A multi-head Self Attention module is applied on CNN to aggregate the features and attend to the most important parts of the CNN feature maps for BiLSTM in the next layer. Finally, two BiLSTM layers are used for classification. To reduce the problem of data imbalance, we also propose a balanced sampler for pre-training the CNN. Our experimental results showed that our proposed approach outperforms state-of-the-art methods for 2, 6 and 10 categories with classification accuracy of 93.01%, 89.79% and 91.72% on test set of UNSW-NB15 dataset.

As future works, we will apply transfer learning methods and metaheuristic methods to reduce the number of parameters in such a complicated structures which can be run on edge devices.

REFERENCES

Alalmaie, Abeer Z.; Nanda, Priyadarsi; He, Xiangjian. (2022). Zero Trust-NIDS: Extended Multi-View Approach for Network Trace Anonymization and Auto-Encoder CNN for Network Intrusion Detection. *IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. Wuhan, China: IEEE.

Abolghasemi, Majid; Dadkhah, Chitra; Tohidi, Nasim. (2022). HTS-DL: Hybrid Text Summarization System using Deep Learning. *The 27th International Computer Conference, the Computer Society of Iran*. Tehran, Online. Retrieved from https://ieeexplore.ieee.org/abstract/document/9780395

Bazi, Hamid reza; Hassanzadeh, Alireza; Moeini, Ali. (2017). A comprehensive framework for cloud computing migration using Meta-synthesis approach. *Journal of Systems and Software, 128*, 87-105.

Cao, Bo; Li, Chenghai; Song, Yafei; Qin, Yueyi; Chen, Chen. (2022). Network Intrusion Detection Model Based on CNN and GRU. *Applied Sciences, 12*(9). Retrieved from https://doi.org/10.3390/app12094184

Dabbaghzadeh, Hamed; Falahati, Abolfazl; Sanandaji, Nader;. (2021). CBTC Systems Safety Enhancement by Employing an IoT-Based Spread Spectrum Technique. *2021 International Conference on Circuits, Controls and Communications (CCUBE)* (pp. 1-5). Bangalore, India: IEEE.

Eusgeld, Irene; Nan, Cen; Dietz, Sven. (2011). “System-of-systems” approach for interdependent critical infrastructures. *Reliability Engineering & System Safety, 96*(6), 679–686.

Ghezelji, Mazyar; Dadkhah, Chitra; Tohidi, Nasim; Gelbukh, Alexander. (2022). Personality-Boosted Matrix Factorization for Recommender Systems. *International Journal of Information and Communication Technology Research (IJICTR), 14*(1), 48-55.

Gourisetti, S.N.G.; Mylrea, M.; Patangia, H.; . (2019). Cybersecurity Vulnerability Mitigation Framework Through Empirical Paradigm (CyFEr): Prioritized Gap Analysis. *IEEE Systems Journal, 14*(2), 1897-1908.

Kanimozhi, V.; Jacob, Prem. (2019). UNSW-NB15 dataset feature selection and network intrusion detection using deep learning. *Journal of Recent Technology and Engineering, 7*(1), 443–446.

Karabacak, Bilge; Yildirim, Sevgi Ozkan; Baykal, Nazife. (2016). A vulnerability-driven cyber security maturity model for measuring national critical infrastructure protection preparedness. *International Journal of Critical Infrastructure Protection, 15*, 47-59.

Kasongo, Sydney M.; Sun, Yanxia;. (2020). Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. *Journal of Big Data, 105*(7). doi:https://doi.org/10.1186/s40537-020-00379-6

Kindervag, J. (2010). *Build Security Into Your Network's DNA: The Zero Trust Network Architecture.* For Security & Risk Professionals, Forrester Research.

Kumar, Vikash; Sinha, Ditipriya; Das, Ayan Kumar; Pandey, Subhash Chandra; Goswami, Radha Tamal. (2020). An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset. *Cluster Computing, 23*, 1397–1418. Retrieved from https://doi.org/10.1007/s10586-019-03008-x

Mettler, T.; Blondiau, A.;. (2012). HCMM-a maturity model for measuring and assessing the quality of cooperation between and within hospitals. *25th IEEE International Symposium on Computer- Based Medical Systems (CBMS).*

Mohammady, M.; Wang, L.; Hong, Y.; Louafi, H.; Pourzandi, M.; Debbabi, M.;. (2018). Preserving both privacy and utility in network trace anonymization. *The 2018 ACM SIGSAC Conference on Computer and Communications Security.*

Moustafa, Nour; Slay, Jill. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *Military Communications and Information Systems Conference (MilCIS).* Canberra, ACT, Australia: IEEE. doi:10.1109/MilCIS.2015.7348942

Nahari H.; Krutz R. L.,. (2003). *System Security Engineering Capability Maturity Model (SSE-CMM) Model Description Document. Version 3.0.* Carnegie Mellon University.

Newhouse, William; Keith, Stephanie; Scribner, Benjamin; Witte, Greg. (2017). *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework.* Gaithersburg: National Institute of Standards and Technology.

Niazi, Mahmood; Saeed, Ashraf Mohammed; Alshayeb, Mohammad; Mahmood, Sajjad; Zafar, Saad. (2020). A maturity model for secure requirements engineering. *Computers & Security, 95.*

Nussbaum, B.; Park, S., . (2018). A tough decision made easy? local government decision-making about contracting for cybersecurity. . *The 19th Annual International Conference on Digital Government Research: Governance in the Data Age.*

Renteria, C.; Gil-Garcia, J.R.; Pardo, T.A.;. (2019). Toward an Enabler-Based Digital Government Maturity Framework: A Preliminary Proposal Based on Theories of Change. *The 12th International Conference on Theory and Practice of Electronic Governance.*

Saleem, D.; Sundararajan, A.; Sanghvi, A.; Rivera, J.; Sarwat, A.I.; Kroposki, B.; . (2019). A Multidimensional Holistic Framework for the Security of Distributed Energy and Control Systems. *IEEE Systems Journal, 14*(1), 17-27.

Sanandaji, Nader; Falahati, Abolfazl;. (2021). Lattice-Based Coding to Enhance Error Performance of the Hidden Direct Sequence Spread Spectrum. *IETE Journal of Research, 67*(6), 754-766.

Tohidi, Nasim; Rustamov, Rustam B. (2022). Short Overview of Advanced Metaheuristic Methods. *International Journal on Technical and Physical Problems of Engineering (IJTPE), 14*(51), 84-97.

Tohidi, Nasim; Rustamov, Rustam B.;. (2020). A review of the machine learning in gis for megacities application. In R. B. Rustamov, *Geographic Information Systems in Geospatial Intelligence* (pp. 29-53). London: IntechOpen.

White, G. (2011). The community cyber security maturity model. *IEEE international conference on technologies for homeland security (HST).*

Yin, Yuhua; Jang-Jaccard, Julian; Xu, Wen; Singh, Amardeep; Zhu, Jinting; Sabrina, Fariza; Kwak, Jin. (2022). IGRF-RFE: A Hybrid Feature Selection Method for MLP-based Network Intrusion Detection on UNSW-NB15 Dataset. *arXiv:2203.16365*. Retrieved from https://doi.org/10.48550/arXiv.2203.16365.