HPAC-IDS: A Hierarchical Packet Attention Convolution for Intrusion Detection System

Anass GRINI[1], Btissam EL KHAMLICHI[1], Abedllatif EL AFIA[3], Amal EL FALLAH SEGHROUCHNI[1,2]

[1] Ai movement, The International Artificial Intelligence Center of Morocco, UM6P, Rabat, Morocco

[2] Lip6, Sorbonne University, Paris, France

[3] ENSIAS, Mohammed V University, Rabat, Morocco

23 April 2024

Outline

1. Introduction

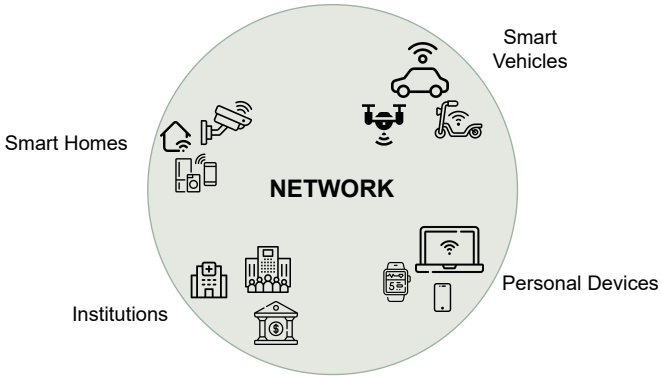
2. Adversarial Attacks Overview
   - Traditional Techniques
   - Machine Learning Techniques
   - Deep Learning Techniques

3. Proposed Work
   - Network Packet Processing
   - Network Packet Emedding
   - Classifier

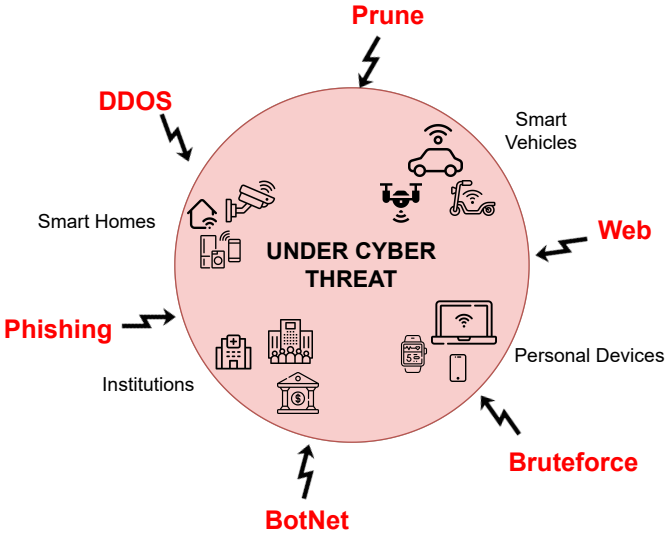
4. Experiment Results

5. Conclusion and Perspectives

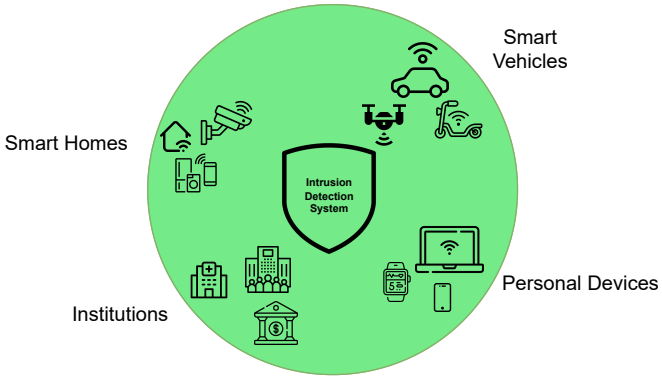
Introduction



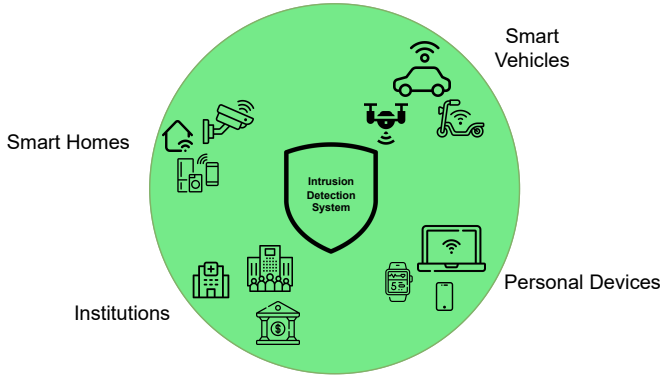
Introduction



Introduction



Introduction



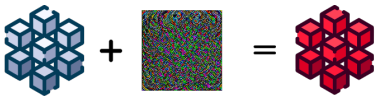
⟶ The Intrusion Detection System (IDS) can be **fooled** using *Adversarial Samples*.

Adversarial Attacks Overview

- Szegedy et al. (2013) could fool neural networks with a small perturbation, which unveiled an *'intriguing'* behavior NNs, which represents a real threat to the numerous intended applications of these models.
- An adversary aims to change the classifier's predicted class by injecting an adversarial sample into the classifier.

Adversarial Perturbation

The essence of an adversarial sample lies in imperceptible perturbations, a crucial component of an adversarial attack.



Adversarial Attacks Overview

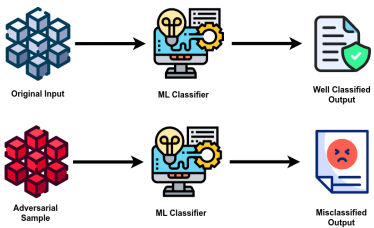


Figure 2.1: Adversarial Samples Induce Misclassification

- In the case of a benign input, the ML classifier gets the predicted class correct
- Adversarial sample causes the ML classifier to make an erroneous prediction.
- In the case of IDSs, the objective is to induce **misclassification** specifically in the identification of **malicious network traffic**.

$\longrightarrow$ To enhance the security of IDSs, it is imperative to bolster their robustness through **adversarial training** and the development of **resilient DL architectures**.

Traditional Techniques

- **Signature-based**: Rely on predefined signatures of known attacks to detect anomalies, but are limited against zero-day threats not yet in the database.
- **Anomaly-based**: makes use of machine learning and statistical approaches to classify traffic. They are able to detect zero-day attacks but generate a high false positive rate when handling benign traffic

Machine Learning Techniques

Machine learning algorithms improve IDS detection accuracy, adaptability, and resilience against evolving threats. Here are some key contributions:

- **Improved pattern recognition:** ML algorithms enhance IDS accuracy in detecting known and unknown threats by recognizing complex patterns in data.
- **Scalability and adaptability:** ML models, trained on extensive datasets, can adapt to the evolving threat landscape. By updating and retraining models as new threats surface, IDS effectiveness is maintained.

Deep Learning Techniques

DL approaches provide automatic feature extraction without manual engineering. However, they require much data and time to build an effective model against network attacks.

- **Recurrent Neural Net.** : LSTM-based architectures enable packet-level IDS classification, identifying malicious traffic semantics and improving processing efficiency.
- **Natural Language** : *Packet2Vec* adapts *Word2Vec* (a neural network model for word embeddings) to analyze network packets, representing hexadecimal values as words to capture relationships.

Proposed Work

HPAC-IDS: Hierarchical Packet Attention Convolution for Intrusion Detection System

- A Hierarchical Attention Convolution System tailored for network packets.
- Implemented a segmenter to break down packets for easier analysis, similar to processing sentences and words.
- The HPAC-IDS effectively detects malicious traffic compared to state-of-the-art deep learning based systems.
- HPAC-IDS remains robust against adversarial attacks, making it a strong candidate for network security.

General Architecture

The HPAC-IDS contains 3 main parts:

- *Network Packet Processing*
- *Network Packet Embedding*
- *Classifier*

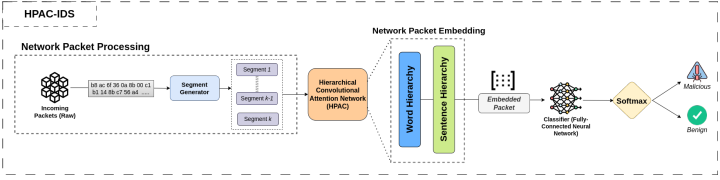


Figure 3.1: General Structure of the HPAC architecture for Malicious Network Packet Detection

Packet Segmenter

- View incoming packets as raw hexadecimal strings.
- Packet segmenter divides each packet into smaller, fixed-size segments.
- Segments contain single-byte hexadecimal data.
- Segmented packets represented as $P = S_1,\ S_2,\ ..., S_m$.
- $m = \lceil \frac{n}{k} \rceil$, where n is packet length and k is segment size.

$\longrightarrow$ **Analogy to Natural Language Processing (NLP)**: Each segment resembles a *'sentence'* of distinct *'words'* (1-byte hexadecimal values).

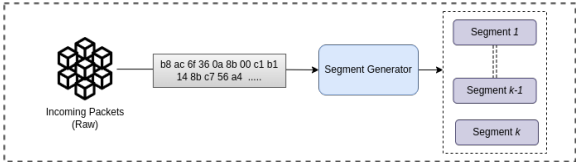


Figure 3.2: Network Packet Segmenter Unit

Network Packet Embedding

- Each word (1-byte hex) in segments is embedded, then fed to the word hierarchy of the HCAN model.
- Each segment is embedded first using segment' word embedding. We recover the sentence-level vector representation.
- The same structure is repeated for the sentences, and sentence embeddings are fed to the sentence hierarchy of the model.
- We get at the end a vectorized representation of the incoming packet, the *Packet Embedding*.

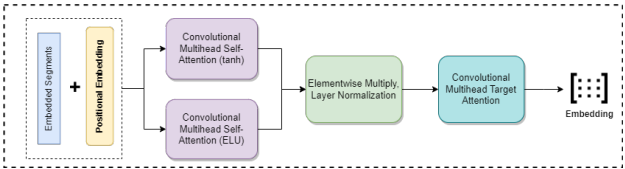


Figure 3.3: Hierarchy structure in HPAC

Classifier

- After getting the embedding of the raw network packet. The embedding tensor is given to an MLP classifier
- Followed by a Softmax that gives packet's class, categorizing it as either *'malicious'* or *'benign'*

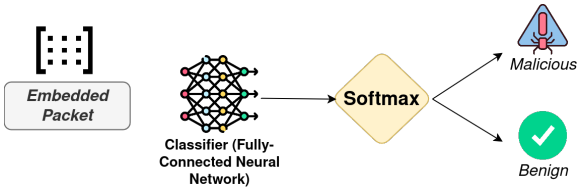


Figure 3.4: Classifier

Experiment Results

Data Understanding

- *CIC-IDS2017* is a comprehensive collection of network traffic data.
- Addresses the need for reliable test and validation datasets.
- Resembles up-to-date real-world network traffic captured in PCAPs format.

Experiment Results

- We compared our proposed *Hierarchical Packet Attention Convolution System (HPAC-IDS)* with the *Extended Byte Segment Neural Network (EBSNN)* Xiao et al. (2021) based on validation and test metrics.

Experiment Results

Hyperparameter's name	Values
Epochs	40
Loss function	Focal Loss
Optimizer	Adam
Learning Rate	10^{-3}
Segment Size	20
Batch Size	40
Embedding Size	96
Number of Heads (*for HPAC-IDS model*)	8

Table 1: HPAC-IDS training hyperparameters

- Training hyperparameters were carefully selected for optimal performance of the HPAC-IDS and EBSNN models.

Experiment Results

Metrics	EBSNN	HPAC-IDS
Validation Accuracy	0.99500	**1.00**
Validation DR	0.90624	**0.99999**
Validation F1 Score	0.94959	**0.99999**
Validation FPR	0.08333	**0.00**
Validation Loss	0.61511	**3 $\times$ 1e-6**
Validation Precision	0.90412	**0.99999**
Test Accuracy	0.99905	**0.99970**
Test DR	0.99937	**0.99987**
Test FPR (%)	0.22499	**0.02499**

Table 2: Performance comparison between EBSNN and HPAC-IDS with segment size 32

Experiment Results

Work	Model	Acc %	DR %	FPR %
Sun et al. Sun et al. (2020)	CNN+LSTM	98.67	97.21	0.47
Azzaoui et al. Azzaoui et al. (2022)	DNN	99.43	80.33	**0.0007**
Yin et al Yin et al. (2023)	Birch + MLP	99.73	-	0.15
Proposed Method	**HPAC**	**99.79**	**99.99**	0.02

Table 3: Comparaison of similar work

Experiment Results

Effect of Segment Size

Segment size	Test Accuracy	Test FPR (%)
8	0.99965	0.1499%
20	0.99895	0.4999%
32	*0.9997*	*0.0249%*
39	0.9674	16.299%

Table 4: Test Results on different *Segment size*

Assessing Adversarial Robustness

- *Severity* measures how badly an attack disrupts the model's decision-making.
- Unlike accuracy, severity captures the degree of the impact.
- High severity = model is highly vulnerable, even if it sometimes gets the classification right.

Assessing Adversarial Robustness: HPAC-IDS vs. EBSNN

Evaluated HPAC-IDS and EBSNN on different segment sizes against:

- *Projected Gradient Descent (PGD)*
- *Fast Gradient Sign Method (FGSM)*
- *Wasserstein GAN (WGAN)*

Model	PGD	FGSM	WGAN
HPAC (seg_size 8)	**0%**	**0%**	**10%**
HPAC (seg_size 20)	**5%**	**10%**	**5%**
HPAC (seg_size 32)	**0%**	**0%**	15%
EBSNN (seg_size 8)	95%	80%	100%
EBSNN (seg_size 20)	95%	80%	65%
EBSNN (seg_size 32)	5%	10%	**0%**

Table 5: Severity results under different adversarial attacks on HPAC-IDS and EBSNN models

Assessing Adversarial Robustness

HPAC-IDS Robustness:

- Extremely robust at segment sizes 8 and 32, especially against PGD attacks (0% severity).
- Minor vulnerability at segment size 20 (PGD: 5%, FGSM: 10%).

EBSNN Vulnerabilities:

- Highly susceptible at segment sizes 8 20, especially with PGD (up to 95% severity) and FGSM (up to 80% severity).
- Improved resistance at segment size 32 against PGD FGSM, but completely resistant to WGAN at this segment size.

Conclusion and Perspectives

- **HPAC-IDS Outperforms Benchmarks:** Experiments demonstrate the superiority of HPAC-IDS in intrusion detection due to its novel packet segmentation approach and hierarchical attention / CNN architecture.
- **Hierarchical NLP-inspired Design:** HPAC-IDS leverages a natural language-inspired design to understand complex traffic patterns and the relationships between network packets.
- **Robustness Against Adversarial Attacks:** HPAC-IDS effectively resists adversarial attacks (PGD, FGSM, WGAN) due to its ability to identify intricate network traffic features that these attacks cannot fully disguise.
- **Optimal Segment Size is Key:** Carefully selecting the segment size in HPAC-IDS allows it to capture long-term traffic patterns, significantly enhancing detection accuracy.
- **Future Directions:** Exploring alternative packet representations (e.g., time series, new NLP-inspired embeddings) could further enhance IDS robustness against evolving threats.

References

Azzaoui, H., Boukhamla, A. Z. E., Arroyo, D., and Bensayah, A. (2022). Developing new deep-learning model to enhance network intrusion classification. *Evolving Systems*, 13(1):17–25.

Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., and Chen, J. (2020). Dl-ids: Extracting features using cnn-lstm hybrid network for intrusion detection system. *Security and communication networks*, 2020:1–11.

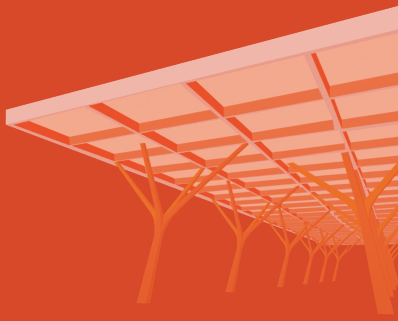
Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Xiao, X., Xiao, W., Li, R., Luo, X., Zheng, H., and Xia, S. (2021). Ebsnn: Extended byte segment neural network for network traffic classification. *IEEE Transactions on Dependable and Secure Computing*, 19(5):3521–3538.

References

Yin, Y., Jang-Jaccard, J., Sabrina, F., and Kwak, J. (2023). Improving multilayer-perceptron (mlp)-based network anomaly detection with birch clustering on cicids-2017 dataset. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 423–431. IEEE.

Thank You.



Anass Grini