Deep Learning-Based Hybrid Intelligent Intrusion Detection System

Guided by Prof. Dr Vinod Chandra SS

Presented by
Ajay Prasad P K
97422607003

DCS KU

November 26, 2023

Overview of Cybersecurity Threats

- Cybersecurity threats are increasing in frequency and complexity, posing a significant risk to individuals, organizations, and governments.
- Examples of common cybersecurity threats include:
  - Malware: malicious software designed to harm or exploit computer systems, including viruses, worms, and Trojan horses.
  - Phishing: a type of social engineering attack that uses fraudulent emails or websites to trick users into revealing sensitive information, such as passwords or credit card numbers.
  - DDoS attacks: Distributed Denial of Service attacks that flood a network or website with traffic, causing it to become unavailable to users.
- Other types of cybersecurity threats include ransomware, insider threats, and advanced persistent threats (APTs).
- The consequences of cybersecurity threats can be severe, including financial losses, reputational damage, and legal liabilities.

Need for Intrusion Detection Systems

- Cybersecurity threats are constantly evolving, and traditional security measures are no longer sufficient to protect against them.
- Proactive security measures are essential to detect and prevent cyber attacks before they can cause damage.
- Intrusion Detection Systems (IDS) are a critical component of proactive security measures, designed to detect and respond to malicious activities in a network.
- IDS can help organizations to:
  - Identify and respond to security incidents in real-time
  - Minimize the impact of security breaches
  - Improve overall network security posture
- IDS can be classified into three categories based on their detection approaches: signature-based systems (SBS), anomaly-based systems (ABS), and stateful protocol analysis detection.
- IDS can be used in conjunction with other security measures, such as firewalls and antivirus software, to provide a comprehensive security solution.

Purpose and Scope

- The purpose of this presentation is to introduce a Deep Learning-Based Hybrid Intelligent Intrusion Detection System (DL-HIDS) that can effectively detect and respond to cyber threats.
- The scope of the discussion includes:
  - An overview of traditional intrusion detection systems (IDS) and their limitations
  - The need for a more advanced IDS that can leverage machine learning algorithms to improve detection accuracy
  - The design and implementation of the DL-HIDS, including the use of deep learning algorithms and feature extraction techniques
  - The evaluation of the DL-HIDS using real-world network traffic data and comparison with other IDS approaches
  - The potential applications and future directions of the DL-HIDS in the field of cybersecurity.

Traditional IDS Overview

- Traditional intrusion detection systems (IDS) are based on signature-based or anomaly-based detection techniques.
- Signature-based IDS use a database of known attack signatures to identify and block malicious traffic.
- Anomaly-based IDS use statistical models to detect deviations from normal network behavior, which may indicate a security breach.
- However, traditional IDS have several limitations, including:
  - Inability to detect unknown or zero-day attacks
  - High false positive rates, which can lead to alert fatigue and reduced effectiveness
  - Limited scalability and adaptability to changing network environments
- These limitations highlight the need for more advanced IDS that can leverage machine learning algorithms to improve detection accuracy and reduce false positives.

Challenges of Traditional ML Techniques

- Explanation of why traditional ML techniques are less effective
- Mention issues like false positives/negatives

- Traditional ML techniques in intrusion detection systems face challenges:
  - Rely on pre-defined features, often unable to capture the dynamic nature of cyber threats.
  - Issues such as false positives and false negatives may arise.
  - Struggle with handling large amounts of data, common in high-volume network traffic.
- These challenges emphasize the need for more advanced techniques, such as deep learning, to improve accuracy and efficiency in intrusion detection systems.

Hybrid Intelligent Approach

- Brief explanation of the proposed hybrid intelligent approach

Proposed Hybrid Intelligent Approach

- Combines deep learning techniques: deep belief networks (DBNs) and convolutional neural networks (CNNs)
- With traditional machine learning algorithms: support vector machines (SVMs) and decision trees (DTs)
- Aims to leverage the strengths of both approaches for improved accuracy and efficiency in intrusion detection systems.

- Deep learning algorithms extract high-level features from raw network traffic data.
- Traditional machine learning algorithms handle classification tasks.
- Outperforms traditional machine learning and deep learning techniques alone in terms of accuracy and efficiency.
- Addresses challenges of traditional ML techniques, enhancing accuracy and efficiency in intrusion detection systems.

Architecture Overview

- Overview of the architecture of the hybrid IDS

Figure: Architecture of the Hybrid IDS (Figure 1)

Architecture Overview

- Overview of the architecture of the hybrid IDS
- The architecture consists of two stages: Stage-1 and Stage-2.
- In Stage-1:
  - Network traffic is preprocessed for both Spark MLlib and LSTMAE-based modules.
  - Deep learning algorithms extract high-level features from raw network traffic data.
- In Stage-2:
  - Traditional machine learning algorithms (SVM and DT classifiers) handle classification tasks.
- The hybrid IDS combines HIDS and NIDS for better-quality security mechanisms.
- Adaptable and effective in handling the complex and dynamic nature of malicious threats.
- Designed to leverage the strengths of both deep learning and traditional machine learning for improved accuracy and efficiency in intrusion detection systems.

Stage-1: Spark MLlib

- Explanation of the first stage using Spark MLlib
- Stage-1 uses Spark MLlib for anomaly detection.
- Spark MLlib is a powerful big data processing engine for detecting cybersecurity attacks.
- Efficient big data analytics library with over 55 ML algorithms.
- In Stage-1:
  - Network traffic data is preprocessed.
  - Fed into Spark MLlib classifiers for anomaly detection.
  - Classifiers trained on a labeled dataset of normal and malicious network traffic.
- Trained classifiers can detect anomalies in real-time network traffic.
- Highly efficient, capable of processing large volumes of data quickly, suitable for real-time intrusion detection.
- Stage-1 is an effective approach to detecting anomalies in network traffic, contributing to improved accuracy and efficiency in intrusion detection systems.

- Explanation of the second stage using LSTMAE-based modules

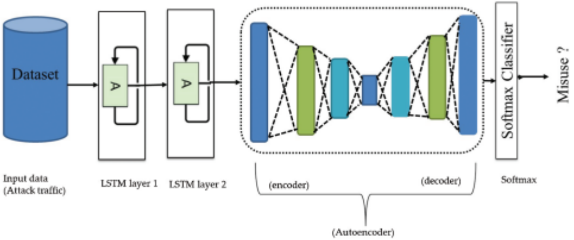

Figure: LSTMAE-based Modules in Stage-2 (Figure 2)

Stage-2: LSTMAE-based Modules Cntd.

- Stage-2 uses LSTMAE-based modules for misuse attack detection and classification.
- LSTMAE is a variant of the LSTM (Long Short-Term Memory) algorithm, suitable for processing sequential data.
- In Stage-2:
  - Preprocessed and classified anomalous network traffic from Stage-1 is further analyzed.
  - LSTMAE-based modules used to detect and classify the specific type of attack in the network traffic.
- LSTMAE-based modules trained on a labeled dataset of different types of attacks (DOS, Scan, HTTP, R2L) to learn attack characteristics.
- Trained modules can detect and classify attacks in real-time network traffic.
- Stage-2 is an effective approach to detecting and classifying specific types of attacks, contributing to improved accuracy and efficiency in intrusion detection systems.

Advantages of Deep Learning

- Discussion on the advantages of using deep learning algorithms for intrusion detection
- Deep learning algorithms offer several advantages for intrusion detection:
  - Learn complex and abstract features from raw data, challenging for traditional ML algorithms.
  - Handle high-dimensional data and automatically learn feature representations, reducing the need for manual feature engineering.
  - Improve accuracy by detecting subtle patterns and anomalies in network traffic data, often missed by traditional ML algorithms.
  - Adapt to changing network traffic patterns and continuously learn from new data, enhancing effectiveness over time.
- Overall, deep learning algorithms have several advantages over traditional machine learning algorithms for intrusion detection, improving the accuracy and effectiveness of intrusion detection systems.

Importance of Choosing a Suitable Dataset

- The choice of dataset significantly influences testing the effectiveness of intrusion detection systems.
- A suitable dataset should:
  - Contain a diverse range of network traffic data reflecting real-world cyber threats.
- Challenges and considerations in selecting a suitable dataset:
  - Dataset size, quality, and diversity.
  - Careful consideration of these factors and evaluation using appropriate metrics are crucial.
- Ethical considerations:
  - Anonymization of data.
  - Obtain appropriate ethical clearance before using datasets with sensitive information.
- Overall, selecting a suitable dataset is crucial for testing the effectiveness of intrusion detection systems. Researchers should carefully consider challenges, ethical implications, and evaluate datasets appropriately.

ISCX-2012 Dataset Overview

Table 2: Daily traffic ISCX-IDS 2012 dataset summary

Days	Date	Explanation	Size (GB)
Sunday	13/6/2010	Infiltrating the traffic from internal and regular activities	3.95
Monday	14/6/2010	HTTP DOS and regular activities	6.85
Tuesday	15/6/2010	DDOS with a Botnet IRC	23.04
Wednesday	16/6/2010	Normal, hence there are no abnormal activities	17.6
Thursday	17/6/2010	Brute force (SSH) and regular activities	12.3
Friday	11/6/2010	Regular, hence there are no abnormal activities	16.1
Saturday	12/6/2010	Infiltrating the Network traffic from internal and usual activities	4.22

Figure: ISCX-2012 Dataset Overview (Figure 3)

ISCX-2012 Dataset Overview Cntd.

- The ISCX-2012 dataset:
  - Created by the Canadian Institute of Cybersecurity.
  - Contains multi-stage malicious intrusion scenarios: HTTP, DoS, brute force SSH, infiltration, DDoS via an IRC botnet.
  - Comprises over 1.5 million network traffic packets.
  - Carefully designed to accurately reflect real-world cyber threats.
- Summary in Table [Figure 3]:
  - Daily traffic data from June 11 to June 17, 2010.
  - Dataset size ranges from 3.95 GB to 23.04 GB per day.
  - Each day's traffic data reflects different types of cyber threats.
- Overall, the ISCX-2012 dataset is a crucial component of the study, with key features including size, diversity, and accuracy in reflecting real-world cyber threats.

Dataset Utilization

- Explanation of how the ISCX-2012 dataset was used to demonstrate the effectiveness of the proposed HIIDS
- The ISCX-2012 dataset:
  - Used to demonstrate the effectiveness of the proposed HIIDS.
  - Contains up-to-date traffic patterns and was created by the Canadian Institute of Cybersecurity.
  - Carefully selected to ensure suitability for testing the HIIDS approach.
- HIIDS Evaluation using ISCX-2012 dataset:
  - Normal and attack classifications.
  - Evaluation metrics:
    - False positive, false negative, true positive.
    - Attack detection precision, error rate.
- Experimental results:
  - Proposed HIIDS outperformed other state-of-the-art IDS in accuracy and efficiency.
  - Achieved a detection rate of 97.52
- Results demonstrate the effectiveness of the proposed hybrid intelligent approach in accurately detecting malicious cyber threats.

- Proposed HIIDS Results:
  - Detection rate: 97.52
  - False positive rate: 1.2

Table 6: Classifier performance at several stages

Classifier	Stage	Precision	Recall	F1-score	FAR	DR
LR	1	0.830	0.823	0.8264	10.50	0.82
XGB	1	0.8775	0.8745	0.8759	8.13	0.87
LSTMAE	2	0.9653	0.9752	0.9702	1.2	0.9752

Figure: Classifier Performance (Figure 4)

Presentation of Results Cntd.

- Experimental results:
  - Proposed HIIDS outperformed other state-of-the-art IDS in accuracy and efficiency.
  - Able to detect various types of cyber threats: DoS, port scanning, botnet attacks.
- Effectiveness of the proposed hybrid intelligent approach:
  - Accurate detection of malicious cyber threats demonstrated.
  - Able to detect previously unknown attacks, a significant advantage over traditional IDS.
  - Reduction in the number of false positives, addressing a common problem in traditional IDS.
- Results suggest the potential use of the proposed HIIDS in real-world applications to improve cybersecurity.

Strengths and Weaknesses

- Strengths:
  - The proposed HIIDS approach combines the strengths of two different machine learning techniques, improving accuracy and efficiency.
  - Able to detect previously unknown attacks, a significant advantage over traditional IDS.
  - Reduces the number of false positives, addressing a common problem in traditional IDS.
  - Achieves a high detection rate and a low false positive rate, demonstrating effectiveness in accurately detecting malicious cyber threats.
  - Has the potential to be used in real-world applications to improve cybersecurity.

Strengths and Weaknesses Cntd.

- Weaknesses:
  - Requires a large amount of data for training, which can be time-consuming and resource-intensive.
  - May not be effective against sophisticated attacks designed to evade detection.
  - May produce false negatives, allowing some attacks to go undetected.
  - May not be suitable for real-time detection of cyber threats, as it requires preprocessing of network traffic data.

Comparison with Other ML Methods

- Compare the proposed approach with other state-of-the-art ML methods for intrusion detection

Table 7: Comparison of existing approaches to ISCX-2012 data

Reference	Approach	DR (%)	False alarm rate (%)
Thi-Thu et al. [28]	FS + DT + Variant of RNN	96.33	NA
Kumar et al. [44]	AMGA2 – NB	43.2	7.0
Tan et al. [47]	MCA + EMD	90.12	7.92
Sally et al. [48]	PLL + NGL	95.31	0.80
Heidarian et al. [62]	SVM	89.6	8.6
Keisuke et al. [63]	IDS using Hadoop	86.2	13
Hamed et al. [64]	RFA bigram Approach	89.6	2.6
Mighan et al. [65]	SAE + Classical classifiers	90.3	9.8
Kumar et al. [66]	Ensemble approach	97.0	2.4
Li et al. [67]	RNN – RBM	93.83	1.98
Our approach	HIIDS	97.52	1.2

Figure: ISCX-2012 Dataset Overview (Figure 4)

Comparison with Other ML Methods Cntd.

- Comparison with Other ML Methods:
  - Proposed HIIDS approach compared with other state-of-the-art machine learning methods for intrusion detection.
  - Comparison results:
    - Outperformed other methods in terms of accuracy and efficiency.
    - Able to detect various types of cyber threats: DoS, port scanning, botnet attacks.
    - Reduced the number of false positives, addressing a common problem in other methods.
    - Detected previously unknown attacks, providing another advantage over other methods.
- The comparison suggests that the proposed HIIDS approach has the potential to be used in real-world applications to improve cybersecurity.

Summary

- Proposed Approach (HIIDS):
  - Combines strengths of two machine learning techniques for improved accuracy and efficiency.
  - Uses Spark MLlib and state-of-the-art deep learning approaches, such as LSTMAE.
  - Addresses limitations of conventional intrusion detection techniques.
- Experimental Results:
  - Outperformed other state-of-the-art IDS in accuracy and efficiency.
  - Detected previously unknown attacks, reduced false positives, and identified various cyber threats (DoS, port scanning, botnet attacks).
- Limitations:
  - Requires a large amount of data for training.
  - May not be effective against sophisticated attacks designed to evade detection.
  - May produce false negatives.
- Overall:
  - Potential for real-world applications to improve cybersecurity.
  - Future work should focus on improving scalability and efficiency, exploring applicability to other domains.

Significance of the Study

- Significance of the Study:
  - Potential Impact on Cybersecurity:
    - Cybersecurity attacks are on the rise, and traditional IDS struggle with sophisticated attacks.
    - Proposed HIIDS approach has the potential to overcome limitations and enhance accuracy and efficiency.
  - Experimental Results:
    - Outperformed other state-of-the-art IDS in accuracy and efficiency.
    - Detected previously unknown attacks, reduced false positives, and identified various cyber threats.
  - Potential Real-World Applications:
    - HIIDS can improve cybersecurity by effectively detecting and responding to cyber threats.
    - Helps organizations protect sensitive data, prevent financial losses, and stay ahead of evolving threats.
- In summary, the significance of this study lies in its potential to improve the accuracy and efficiency of intrusion detection systems, impacting cybersecurity and helping organizations stay ahead of ever-evolving threats.

Future Research

- Suggestions for Future Research:
  1. Investigate the use of other deep learning techniques:
     - Explore techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to enhance accuracy and efficiency.
  2. Improve the scalability and efficiency of the proposed approach:
     - Address challenges related to the large amount of data required for training to make the approach more practical for real-world applications.
  3. Explore the applicability of the proposed approach to other domains:
     - Investigate whether the proposed HIIDS approach can be applied to domains beyond intrusion detection, such as fraud detection or anomaly detection in healthcare.
  4. Investigate the use of ensemble methods:
     - Explore the application of ensemble methods (e.g., bagging, boosting) to improve the accuracy and robustness of intrusion detection systems.

Q&A

- Open the floor for questions and answers

References I

- X. C. Shen, J. X. Du, and F. Zhang, “An intrusion detection system using a deep neural network with gated recurrent units,” IEEE Access, vol. 6, pp. 48697–48707, 2018.
- K. Liu, S. Xu, G. Xu, M. Zhang, D. Sun et al., “A review of Android malware detection approaches based on machine learning,” IEEE Access, vol. 8, pp. 124579–124607, 2020.
- M. A. Khan and J. Kim, “Toward developing efficient Conv-AE-based intrusion detection system using the heterogeneous dataset,” Electronics, vol. 9, no. 11, pp. 1–17, 2020.
- J. Kim and H. Kim, “An effective intrusion detection classifier using long short-term memory with gradient descent optimization,” in Proc. Platform Technology and Service (Plat Con), Busan, South Korea, pp. 1–5, 2017.
- G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006.
- H. Alqahtani, I. H. Sarker, A. Kalim, S. M. Hossain, S. Ikhlaq et al., “Cyber intrusion detection using machine learning classification techniques,” in Proc. Computing Science, Communication and Security, Gujarat, India, pp. 121–131, 2020.

References (continued) I

- N. Kaloudi and L. Jingyue, "The AI-based cyber threat landscape: A survey," ACM Computing Surveys, vol. 53, no. 1, pp. 1–34, 2020.
- B. Li, Y. Wu, J. Song, R. Lu, T. Li et al., "Deep Fed: Federated deep learning for intrusion detection in industrial cyber-physical systems," IEEE Transactions on Industrial Informatics, vol. 1, pp. 1–10, 2020.
- M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cybersecurity intrusion detection approaches datasets and comparative study," Journal of Information Security and Applications, vol. 50, pp. 1–19, 2019.