Deep Learning-Based Hybrid Intelligent Intrusion Detection System

Guided by Prof. Dr Vinod Chandra SS

Presented by

Ajay Prasad P K

97422607003

Department of Computer Science

November 30, 2023

Overview of Cybersecurity Threats

- Rising Threat Landscape:
  - Increased frequency and complexity
- Common Threats:
  - **Malware:** Viruses, worms, Trojan horses
  - **Phishing:** Fraudulent emails/websites for sensitive data
  - **DDoS Attacks:** Flood network, causing unavailability
- Additional Threats:
  - Ransomware, insider threats, APTs
- Consequences:
  - Financial losses, reputational damage, legal liabilities

Need for Intrusion Detection Systems

- Cyber threats evolve, surpassing traditional security measures
- Proactive security is vital to prevent damage from cyber attacks
- Intrusion Detection Systems (IDS):
  - Detect and respond to malicious activities
  - Benefits:
    - Real-time incident response
    - Minimize impact of breaches
    - Enhance overall network security
  - Categories:
    - Signature-based System(SBS)
    - Anomaly-based System(ABS)
    - Stateful protocol analysis
- Use IDS in conjunction with firewalls and antivirus for comprehensive security

Traditional IDS Overview

- Traditional IDS: Signature-based or anomaly-based detection
- Signature-based: Uses known attack signatures for identification
- Anomaly-based: Detects deviations from normal behavior using statistical models
- Limitations:
  - Inability to detect unknown or zero-day attacks
  - High false positive rates, leading to alert fatigue
  - Limited scalability and adaptability to changing network environments
- Need for advanced IDS with ML to improve detection accuracy and reduce false positives

Challenges of Existing Techniques in IDS

- Challenges with traditional Machine Learning in intrusion detection:
  - Reliance on pre-defined features, limiting adaptability to dynamic threats
  - Issues with false positives/negatives
  - Struggles with large data volumes in high-traffic networks
- Emphasizes the need for advanced techniques like deep learning for improved accuracy and efficiency

Motivation and Significance of the Study

- Significance:
  - Potential Impact on Cybersecurity:
    - Cybersecurity attacks rising, traditional IDS struggling
    - Potential to overcome limitations and enhance accuracy
  - Expected Experimental Results:
    - To Outperform other IDS in accuracy and efficiency.
    - To Detect unknown attacks, reduced false positives, identify various threats
  - Real-World Applications:
    - improved cybersecurity, protects data, prevents financial losses
    - Helps organizations stay ahead of evolving threats
- In summary, The study's significance lies in improving IDS accuracy, impacting cybersecurity, and aiding organizations against evolving threats

Advantages of Deep Learning

- Advantages of using deep learning for intrusion detection:
  - Learns complex features from raw data
  - Handles high-dimensional data, reducing manual feature engineering
  - Improves accuracy by detecting subtle patterns missed by traditional ML
  - Adapts to changing traffic patterns, continuously learning from new data
- Deep learning outperforms traditional ML, enhancing accuracy and effectiveness in intrusion detection

Deep Learning-Based Hybrid Intelligent Intrusion Detection System

Combining Unsupervised and Supervised Learning Techniques for Accurate Detection of Cyber Threats

Muhammad Ashfaq Khan and **Yangwoo Kim**
Department of Information and Communication Engineering, Dongguk University, Seoul, 100-715, Korea
Department of Electronics Engineering, IoT and Big-Data Research Center, Incheon National University, Incheon, Korea
* Corresponding Author: Yangwoo Kim. Email: ywkim@dongguk.edu

Computers, Materials & Continua
Tech Science Press
DOI: 10.32604/cmc.2021.015647
January 2021

Hybrid Intelligent Approach

- Deep learning extracts high-level features from raw network data
- Traditional ML handles classification
- Should Outperform standalone techniques in accuracy and efficiency
- Addresses challenges of traditional ML, enhancing intrusion detection systems

Proposed Hybrid Intelligent Approach

- Combines Logistic Regression (LR), Extreme Gradient Boosting (XGB), with with Spark MLlib, and Long Short-Term Memory Autoencoder (LSTMAE)
- Aims for enhanced accuracy and efficiency in intrusion detection

Architecture Overview

- Overview of the architecture of the hybrid IDS

Figure: Architecture of the Hybrid IDS

Architecture Overview Cntd.

- Consists of two stages: Stage-1 and Stage-2
- **Stage-1:**
  - Preprocesses network traffic for Spark MLlib and LSTMAE modules
  - Deep learning extracts high-level features
- **Stage-2:**
  - Traditional ML algorithms (LR, XGB) handle classification
- Combines HIDS and NIDS for enhanced security.
- Adaptable and effective against dynamic threats.
- Leverages both deep learning and traditional ML for improved accuracy and efficiency

Stage-1: Spark MLlib

- Spark MLlib for anomaly detection in Stage-1
- Powerful big data processing engine for cybersecurity attacks
- Over 55 ML algorithms for efficient analytics.
- In Stage-1:
  - Preprocesses network traffic data
  - Uses Spark MLlib classifiers for real-time anomaly detection
  - Trained on labeled datasets of normal and malicious traffic
- Efficiently processes large data volumes for real-time intrusion detection
- Contributes to improved accuracy and efficiency in intrusion detection systems

Stage-2: LSTMAE-based Modules

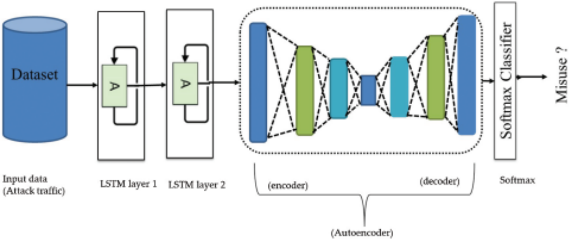

Figure: LSTMAE-based Modules in Stage-2

Stage-2: LSTMAE-based Modules Cntd.

- Stage-2 uses LSTMAE-based modules for misuse attack detection and classification
- LSTMAE: Variant of LSTM for processing sequential data
- In Stage-2:
  - Analyzes preprocessed anomalous traffic from Stage-1
  - LSTMAE-based modules detect and classify specific attack types
- Trained on labeled datasets of various attacks (DOS, Scan, DDos, R2L) for learning attack characteristics
- Detects and classifies attacks in real-time network traffic
- Effective in classifying specific attack types, improving accuracy and efficiency in intrusion detection systems

Importance of Choosing a Suitable Dataset

- Dataset choice crucial for testing intrusion detection systems
- Suitable dataset should:
  - Contain diverse, real-world cyber threat traffic
- Challenges and considerations:
  - Size, quality, and diversity
  - Evaluation using appropriate metrics is crucial
- Ethical considerations:
  - Anonymization of data
  - Obtain ethical clearance for datasets with sensitive information
- Overall, choosing a suitable dataset is crucial Researchers must consider challenges, ethical implications, and evaluate datasets appropriately

ISCX-2012 Dataset Overview

- ISCX-2012 dataset:
  - Created by the Canadian Institute of Cybersecurity
  - Features multi-stage malicious intrusion scenarios
  - Includes scenarios like DoS, brute force SSH, infiltration, DDoS via IRC botnet
  - Comprises 1.5 million+ network traffic packets
  - Carefully designed to reflect real-world cyber threats accurately
- Summary :
  - Daily traffic data from June 11 to June 17, 2010.
  - Dataset sizes range from 3.95 GB to 23.04 GB per day.
  - Each day's data reflects different cyber threats
- ISCX-2012: Crucial for its size, diversity, and accurate representation of real-world cyber threats

ISCX-2012 Dataset Overview Cntd.

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: Daily Traffic ISCX-IDS-2012 Dataset summary

Dataset Utilization

- ISCX-2012 dataset used to demonstrate HIIDS effectiveness
  - Up-to-date patterns, created by the Canadian Institute of Cybersecurity
  - Carefully selected for HIIDS testing suitability
- HIIDS Evaluation:
  - Normal and attack classifications
  - Metrics: False positive, false negative, true positive, precision, error rate
- Experimental Results:
  - HIIDS outperformed other IDS in accuracy and efficiency
  - Achieved 97.52%
- Results demonstrate HIIDS effectiveness in accurately detecting malicious cyber threats

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

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 at several stages

Results Cntd.

- Experimental Results:
  - HIIDS outperformed other state-of-the-art IDS in accuracy and efficiency
  - Detected various cyber threats: DoS, port scanning, botnet attacks
- Effectiveness of HIIDS:
  - Accurate detection of malicious threats demonstrated.
  - Detects unknown attacks, a significant advantage over traditional IDS
  - Reduces false positives, addressing a common issue in traditional IDS
- Results suggest the potential real-world application of HIIDS to enhance cybersecurity

Strengths

- HIIDS Strengths:
    - Combines strengths of two ML techniques, improving accuracy and efficiency
    - Detects unknown attacks, a significant advantage
    - Reduces false positives, addressing a common issue
    - High detection rate, low false positive rate, demonstrating effectiveness
    - Potential for real-world applications in cybersecurity improvement

Areas of Improvement

- HIIDS Areas of Improvement:
    - Requires substantial data for training, which can be time and resource-intensive
    - May struggle against sophisticated attacks designed to evade detection
    - Potential for false negatives, allowing some attacks to go undetected
    - May not be suitable for real-time detection due to the need for preprocessing network traffic data

Scopes

1. **Advanced Deep Learning Algorithms:**
   - Integrate advanced deep learning models (e.g., CNNs, RNNs) for improved intrusion detection accuracy
2. **Enhanced Preprocessing:**
   - Refine preprocessing with advanced feature extraction and data cleaning techniques
3. **Diverse Dataset Evaluation:**
   - Evaluate system performance on diverse datasets to ensure effectiveness across various cyber threats
4. **Real-Time Detection System:**
   - Develop a real-time intrusion detection system for prompt threat detection and response

These future scopes aim to elevate the capabilities of HIIDS, addressing evolving cybersecurity challenges

Summary

- Proposed HIIDS Approach:
  - Combines ML techniques for enhanced accuracy and efficiency
  - Utilizes Spark MLlib and deep learning (LSTMAE)
  - Addresses limitations of conventional intrusion detection
- Experimental Results:
  - Outperformed other IDS in accuracy and efficiency
  - Detected unknown attacks, reduced false positives, identified various threats (DoS, port scanning, botnet)
- Limitations:
  - Requires large training data
  - May struggle against sophisticated attacks, producing false negatives
- Overall:
  - Potential for real-world cybersecurity improvement
  - Future work: Enhance scalability, explore applicability in other domains

Q&A

- the floor for questions and Suggestions

References

1. 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.
2. 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.
3. C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," *Computers and Security*, vol. 70, no. 2, pp. 255–277, 2017.
4. Rhishabh Hattarki, Shruti Houji, and Manisha Dhage, "Real-Time Intrusion Detection System For IoT Networks," 2021 6th International Conference for Convergence in Technology (I2CT) Pune, India, Apr 02-04, 2021.
5. Jedrzej Bieniasz and Krzysztof Szczypiorski, "Dataset Generation for Development of Multi-Node Cyber Threat Detection Systems," *Electronics* 2021, 10, 2711.

Thank You!

Thank you for your attention!

- Contact Information: `ajayprasad0008@gmail.com.com`

Feel free to reach out for further discussion or information.