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 27, 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 (SBS).
    - Anomaly-based (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 Traditional ML Techniques

- Challenges with traditional ML 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.

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

Hybrid Intelligent Approach

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.

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

Architecture Overview

- Overview of the architecture of the hybrid IDS

Figure: Architecture of the Hybrid IDS (Figure 1)

Architecture Overview Cntd.

- Hybrid IDS architecture overview.
- 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 (SVM, DT) 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

- Explanation of the second stage using LSTMAE-based modules

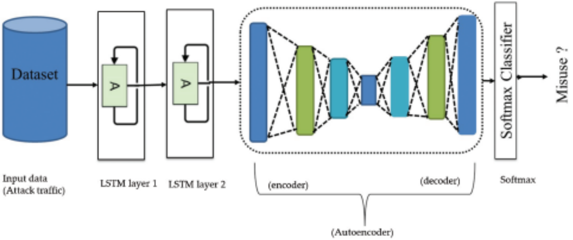

Figure: LSTMAE-based Modules in Stage-2 (Figure 2)

Stage-2: LSTMAE-based Modules Cntd.

- Stage-2 uses LSTMAE-based modules for misuse attack detection and classification.
- LSTMAE: 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, HTTP, 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.

Advantages of Deep Learning

- Advantages of using deep learning for intrusion detection:
  - Learns complex features from raw data, challenging for traditional ML.
  - 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.

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

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.

- ISCX-2012 dataset:
  - Created by the Canadian Institute of Cybersecurity.
  - Features multi-stage malicious intrusion scenarios.
  - Includes scenarios like HTTP, 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 [Figure 3]:
  - 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.

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.

Presentation of Results

- 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:
  - 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.

Weaknesses

- HIIDS Weaknesses:
  - 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.

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 Methods:
  - HIIDS compared with state-of-the-art ML methods for intrusion detection.
  - Results:
    - Outperformed others in accuracy and efficiency.
    - Detected various threats: DoS, port scanning, botnet attacks.
    - Reduced false positives, addressing a common issue.
    - Detected unknown attacks, an advantage over other methods.
- Comparison suggests HIIDS has potential for real-world cybersecurity improvement.

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.

Significance of the Study

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

Future 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.

Q&A

- the floor for questions and Suggestions

References

1. 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.
2. N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," Information Security Journal: A Global Perspective, vol. 25, no. 13, pp. 18–31, 2016.
3. J. A. Khan and N. Jain, "A survey on intrusion detection systems and classication techniques," International Journal of Scientific Research in Science, Engineering and Technology, vol. 2, no. 5, pp. 202–208, 2016.
4. N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system," Procedia Computer Science, vol. 89, no. 1, pp. 213–217, 2016.
5. S. S. S. Sindhu, S. Geetha and A. Kannan, "Decision tree-based lightweight intrusion detection using a wrapper approach," Expert Systems with Applications, vol. 39, no. 1, pp. 129–141, 2012.
6. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cybersecurity intrusion detection," IEEE Communications Surveys and Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
7. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P. A. Manzagol et al., "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," Journal of Machine Learning Research, vol. 11, no. 12, pp. 3371–3408, 2010.
8. G. Kim, S. Lee and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," Expert Systems with Applications, vol. 41, no. 4, pp. 1690–1700, 2014.
9. N. Paulauskas and J. Auskalnis, "Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset," in Proc. Open Conf. of Electrical Electronic and Information Sciences IEEE, Vilnius, Lithuania, pp. 1–5, 2017.
10. S. Zaman and F. Karray, "Features selection for intrusion detection systems based on support vector machines," in Proc. of Consumer Communications and Networking IEEE, Las Vegas, NV, USA, pp. 1–8, 2009.

References

11. N. Kaloudi and L. Jingyue, “The AI-based cyber threat landscape: A survey,” ACM Computing Surveys, vol. 53, no. 1, pp. 1–34, 2020.
12. 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,” EEE Transactions on Industrial Informatics, vol. 1, pp. 1–10, 2020.
13. 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.
14. W. Zong, Y. W. Chow and W. Susilo, “Interactive three-dimensional visualization of network intrusion detection data for machine learning,” Future Generation Computer Systems, vol. 102, no. 4, pp. 292–306, 2020.
15. 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.
16. N. Moustafa and J. Slay, “The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set,” Information Security Journal: A Global Perspective, vol. 25, no. 13, pp. 18–31, 2016.
17. J. A. Khan and N. Jain, “A survey on intrusion detection systems and classification techniques,” International Journal of Scientific Research in Science, Engineering and Technology, vol. 2, no. 5, pp. 202–208, 2016.
18. N. Farnaaz and M. A. Jabbar, “Random forest modeling for network intrusion detection system,” Procedia Computer Science, vol. 89, no. 1, pp. 213–217, 2016.
19. S. S. S. Sindhu, S. Geetha and A. Kannan, “Decision tree-based lightweight intrusion detection using a wrapper approach,” Expert Systems with Applications, vol. 39, no. 1, pp. 129–141, 2012.
20. A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cybersecurity intrusion detection,” IEEE Communications Surveys and Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.

References

21. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, M. Ghogho et al., "Deep learning approach for network intrusion detection in software-defined networking," in Proc. Int. Conf. on Wireless Networks and Mobile Communications, Fez, Morocco, pp. 258–263, 2016.
22. M. A. Qatf, Y. Lasheng, M. A. Habib and K. A. Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," IEEE Access, vol. 6, pp. 52843–52856, 2018.
23. F. Farahnakian and J. Heikkonen, "A deep auto-encoder based approach for an intrusion detection system," in Proc. 20th Int. Conf. on Advanced Communication Technology, Chuncheon-si Gangwon-Do, South Korea, pp. 178–183, 2018.
24. S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han et al., "Enhanced network anomaly detection based on deep neural networks," IEEE Access, vol. 6, pp. 48231–48246, 2018.
25. S. Bandyopadhyay, R. Chowdhury, A. Roy and B. Saha, "A step forward to revolutionize intrusion detection system using deep convolution neural network," Preprints, vol. v1, pp. 1–13, 2020.
26. M. A. Albahar and M. Binsawad, "Deep autoencoders and feedforward networks based on a new regularization for anomaly detection," Security and Communication Networks, vol. 2020, no. 8, pp. 1–9, 2020.
27. M. Monshizadeh, V. Khatri, B. G. Atli, R. Kantola and Z. Yan, "Performance evaluation of a combined anomaly detection platform," IEEE Access, vol. 7, pp. 100964–100978, 2019.
28. L. T. Thu, Y. Kim and H. Kim, "network intrusion detection based on novel feature selection model and various recurrent neural networks," Applied Sciences, vol. 9, no. 7, pp. 1–29, 2019.
29. S. N. Mighan and M. Kahani, "Deep learning-based latent feature extraction for intrusion detection," in Proc. Electrical Engineering (ICEE), Iranian Conf. on IEEE, Mashhad, Iran, pp. 1511–1516, 2018.
30. W. Wang, Y. Sheng, J. Wang, X. Zeng, X. Ye et al., "HAST-IDS learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection," IEEE Access, vol. 6, pp. 1792–1806, 2017.

Thank You!

Thank you for your attention!

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

Feel free to reach out for further discussion or information.