Advanced Machine Learning-Based Detection of Security Anomalies in Smart Home Networks Using Deep Learning Techniques

[1]Shubi Khajuria,[2] Prof. Ramandeep kaur M Tech (CSE)

Punjab Technical University

Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib

ABSTRACT: The rapid growth of smart home devices has introduced significant convenience but also expanded the attack surface for cyber threats. Traditional signature-based intrusion detection methods often fail to detect novel or evolving attacks targeting these IoT environments. This research investigates the use of advanced machine learning techniques—specifically deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, alongside K-Nearest Neighbors (KNN)—to detect security anomalies in smart home networks. A simulated smart home environment was created to capture both benign and malicious network traffic, including SSH brute-force and botnet attacks. Data preprocessing involved feature extraction and dimensionality reduction techniques to enhance model training. Experimental results demonstrate that LSTM achieves the highest detection accuracy of 98.3% with the lowest false positive rate, effectively capturing temporal patterns in attack behavior. CNN also performs robustly in identifying spatial traffic anomalies, while KNN shows limitations with dynamic datasets. These findings suggest that integrating deep learning models into smart home security systems can significantly improve the detection and mitigation of sophisticated cyber threats.

Keywords: Artificial Intelligence, Machine Learning, cybersecurity, accuracy,VAPT,SSH Brute Force Attacks

I. Introduction:

The rapid adoption of Internet of Things (IoT) devices in smart homes has transformed how people interact with their living environments. Smart devices such as lighting systems, thermostats, security cameras, and entertainment systems offer unprecedented convenience and automation. However, this technological advancement introduces significant cybersecurity challenges. Smart home devices often operate with minimal security protections, outdated firmware, and weak authentication mechanisms, rendering them susceptible to various cyberattacks.

Malicious actors exploit these vulnerabilities by launching attacks such as SSH brute-force attempts, deploying botnets to orchestrate large-scale distributed denial-of-service (DDoS) attacks, or hijacking devices to exfiltrate sensitive data. Given the critical role these devices play in daily life, securing smart homes against evolving cyber threats has become paramount.

II. Literature Review:

Machine learning (ML) Multiple studies benchmark various ML models for intrusion detection. Ahmad et al. (2018) demonstrated that deep learning models outperform traditional classifiers in accuracy and recall. Delplace et al. (2020) highlighted that while CNNs and LSTMs achieve higher detection rates, they require more computational resources and longer training times. There is a trade-off between interpretability, performance, and resource demands.

Smart homes are increasingly connected, comprising heterogeneous devices communicating via Wi-Fi, Zigbee, Bluetooth, or proprietary protocols. While these devices facilitate automation and convenience, their security often remains an afterthought. Common vulnerabilities include default passwords, open ports, and lack of firmware updates. Studies reveal that attackers leverage these weaknesses to launch SSH brute-force attacks, whereby repeated login attempts are made to gain unauthorized access. Botnets such as Mirai exploit vulnerable IoT devices to create massive DDoS attacks, crippling internet infrastructure.

III. Research Methodology for Comparative Analysis

The study adopts an experimental quantitative approach to evaluate ML models under controlled attack simulations. Data was collected from a custom-built smart home testbed designed to mimic typical device interactions and attack scenarios.

The methodology follows six cyclical phases:

The methodology for this research is structured around a modular system architecture designed for the comparative analysis of machine learning models. This architecture follows a sequential data pipeline, with each stage handled by a dedicated component to ensure a systematic and reproducible process. The following interconnected modules drive the research workflow:

Data Collection Environment

The smart home testbed consisted of IoT devices including smart bulbs, IP cameras, a smart plug, and a smart TV connected over a common Wi-Fi network. Attack vectors were introduced using automated scripts: Hydra for SSH brute-force attacks targeting login services, and custom scripts simulating botnet command and control communications.

3.3 Data Preprocessing

Network traffic was captured using Wireshark, logging packet-level data. Python scripts extracted key features such as packet size, flow duration, protocol type, connection frequency, and source/destination IPs. To reduce noise and dimensionality, PCA was applied alongside deep feature extraction using autoencoders, highlighting significant traffic characteristics conducive to classification.

3.4 Model Training and Tuning

Three ML models—CNN, LSTM, and KNN—were trained using a split of 70% training and 30% testing data. Hyperparameters were optimized through grid search and cross-validation, adjusting learning rates (0.001 to 0.01), epoch sizes (20 to 100), and batch sizes (32 to 128) to enhance accuracy and generalization.

3.5 Evaluation Metrics

Models were evaluated on:

- **Accuracy:** Overall correct predictions
- **Precision:** Correct positive predictions among all positives predicted
- **Recall:** Correct positive predictions among all actual positives
- **F1-Score:** Harmonic mean of precision and recall
- **False Positive Rate (FPR):** Proportion of benign instances incorrectly classified as attacks

Confusion matrices were also analyzed to observe misclassification patterns.

Iv Results and Analysis

4.1 Model Performance Summary

Model	Accuracy	Precision	Recall	F1-Score	FPR
CNN	97.6%	96.4%	95.8%	96.1%	2.1%
LSTM	98.3%	97.5%	96.9%	97.2%	1.7%
KNN	93.4%	92.0%	90.1%	91.0%	4.8%

4.2 Confusion Matrix Analysis

The LSTM model achieved the fewest false positives and false negatives, highlighting its strength in capturing temporal patterns inherent to SSH brute-force attacks. CNN showed superior performance in detecting volumetric traffic anomalies, leveraging its ability to learn spatial relationships in packet features. KNN lagged behind, particularly struggling with timely detection in rapidly evolving attack patterns due to its distance-based, non-sequential nature.

4.3 Visualizations

- **ROC Curves:** LSTM reached an AUC of 0.98, CNN 0.97, and KNN 0.91, demonstrating superior discrimination ability of deep learning models.
- **PCA Plots:** Clearly delineated clusters of normal versus anomalous traffic, validating the effectiveness of preprocessing techniques in enhancing separability.

V Discussion and Conclusion

5.1 Summary of Findings

The experimental results affirm that deep learning models, particularly LSTM and CNN, provide robust anomaly detection capabilities in smart home network environments. LSTM's proficiency with sequential data renders it especially effective against login-based intrusion attempts, while CNN excels at recognizing spatial traffic anomalies typical of volumetric attacks. KNN, while useful as a baseline, exhibits limitations with complex, high-dimensional, and dynamic data.

5.2 Implications

The findings support the integration of DL-based anomaly detectors into smart home security architectures. Embedding LSTM and CNN models within edge devices like routers or smart hubs can enable real-time detection and mitigation of cyber threats, thereby protecting user privacy and device integrity.

5.3 Recommendations

- Deploy LSTM models for detecting time-sequenced intrusions such as SSH brute-force attempts.
- Utilize CNNs for spatial traffic pattern recognition and volumetric attack detection.
- Consider hybrid approaches combining KNN's quick filtering capability with deep learning's deeper analysis for optimized performance.

5.4 Future Work

Future research should focus on:

- Applying transfer learning to adapt models for diverse smart home devices.
- Real-time deployment and resource optimization on low-power edge hardware.
- Incorporating adversarial training techniques to harden models against evasion tactics.
- Designing lightweight neural networks tailored for IoT constraints.

5.5 Conclusion

In conclusion, this study demonstrates that advanced deep learning techniques significantly enhance the detection of security anomalies in smart home networks compared to traditional ML methods. LSTM and CNN models deliver high precision and low false positive rates, making them promising candidates for securing the next generation of connected homes. By adopting these technologies, smart home ecosystems can achieve improved resilience against the growing landscape of cyber threats.

References

❖ Alreshidi, N., & Ahmad, I. (2019). Internet of Things (IoT) for smart cities: State of the art, challenges, and open issues. 2019 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), ¬ Ali, S., & Awad, A. I. (2018).

❖ A Review on Internet of Things (IoT) and Its Security Issues. 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE ¬ Delplace, J., Hermoso, R., & Anandita, S. (2020).

❖ A Survey on Anomaly Detection in the Internet of Things: From Data Filtering to Classification Techniques. IEEE Access, 8, 152907-152923.Gonzalez-Manzano, L., Fuentes, L., & Ribagorda, A. (2019). ¬ Thamilarasu, G., & Chawla, V. (2019).

❖ An Intrusion Detection System Using Machine Learning Techniques in Internet of Things Networks. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)