Enhancing IoT Security with Machine Learning Models: A Focus on Inference Time in Blockchain-Based Networks

1st Nour ElHouda Bel Haj Fraj
Institut National des Sciences Appliquées et de Technologie (INSAT), Tunis

2nd Nada Jarboui
Institut National des Sciences Appliquées et de Technologie (INSAT), Tunis

3rd Ghofrane Aloui
Institut National des Sciences Appliquées et de Technologie (INSAT), Tunis

4th Shams Ben Mefteh
Institut National des Sciences Appliquées et de Technologie (INSAT), Tunis

5th Arij Aguel
Institut National des Sciences Appliquées et de Technologie (INSAT), Tunis

***Abstract*—Network security, especially in IoT environments, is becoming increasingly complex with the increase in cyber threats.Ensuring the security of IoT networks is a critical challenge, particularly in environments where rapid decision-making is essential. The integration of blockchain into these networks has emerged as a promising solution to enhance security, thanks to its decentralization and immutability characteristics. At the same time, machine learning (ML) has shown its effectiveness in detecting intrusions and improving the performance of security systems.Other researchers have investigated the use of machine learning and blockchain to enable decentralized protection, and privacy. However, one of the key challenges in IoT networks is the inference time of ML models, which needs to be optimized to meet real-time requirements. This paper evaluates the performance of various ML models in blockchain-enabled IoT networks, focusing on prediction accuracy and inference time,two key factors for real-time systems. We opted for the Random Forest model wich outperforms other ML models, achieving an accuracy of 99.41% with an inference time of 0.000026 seconds on the NSL-KDD dataset, and an accuracy of 99.98% with an inference time of 0.000032 seconds on the KDD Cup 99 dataset. These findings confirm the suitability of Random Forest for real-time IoT security, demonstrating its ability to balance high prediction accuracy with minimal inference time.**

***Index Terms*—Machine Learning, Blockchain, IoT, network security**

I. Introduction

The rapid evolution and widespread adoption of Internet of Things (IoT) technologies have profoundly transformed both personal and professional environments. However, this unprecedented connectivity has also introduced significant cybersecurity challenges. Recent statistics reveal an alarming rise in IoT-related threats, with IoT malware attacks increasing by 87% year-over-year in 2022, reaching a record 112.3 million attacks. This surge underscores the growing appeal of IoT devices, such as smart TVs and security cameras, as targets for malicious actors [1]

Blockchain (BC) is establishing itself as an effective technology in this field, thanks to its properties of decentralization, transparency and immutability, which make it possible to strengthen the security of data exchanged within IoT network [2]. Machine learning (ML) plays a vital role in network intrusion detection and prevention. It can quickly analyze large amounts of data and accurately identify anomalies [3]. However, in real-time IoT networks, inference time(the time to produce a prediction)is crucial. An effective model must combine high accuracy with fast response time to meet the needs of critical IoT environments. [4]

This work centers on the evaluation of several machine learning (ML) models in blockchain-enabled environments, specifically targeting their effectiveness in ensuring the security of real-time IoT networks. The analysis focuses on two critical metrics: prediction accuracy and inference time. These factors are crucial in determining the feasibility of deploying ML models in IoT networks, where rapid decision-making is often required to counteract emerging threats.

The combination of BC and ML not only addresses the scalability and response time challenges inherent in IoT networks but also sets a foundation for more secure, efficient and adaptive systems capable of mitigating emerging cyber threats [5].

This paper is structured as follows:

- The introduction provides the context, challenges, and motivations behind integrating blockchain and ML in IoT security, alongside an outline of the paper.
- The Related work discusses traditional and modern approaches, identifying gaps and opportunities in existing research.
- The materials section details the datasets utilized.
- Experimental results describe the simulation setup and analyze performance metrics.
- The discussion to interpret our findings

- The conclusion summarizes the contributions and potential implications of this work.

By addressing the critical issue of inference time, this study aims to provide actionable insights into designing secure and efficient real-time IoT systems.

II. Related Work

This section reviews network security approaches for IoT, starting with **traditional methods** and their limitations. We then explore **new techniques**, such as blockchain and machine learning, their benefits, and the **challenges** they face. Finally, we discuss the potential of combining these technologies to address IoT security challenges.

A. Traditional Approaches

Traditional network security approaches for iot networks have relied on centralized systems with signature-based intrusion detection methods. These methods primarily depend on predefined attack signatures stored in centralized storage networks. While effective in identifying known threats, these techniques face significant limitations when dealing with evolving cyber threats. They lack adaptability and struggle to counter complex and sophisticated cyberattacks, leaving IoT networks vulnerable to new and unforeseen security challenges. [6]

B. New Approaches

Recent research has introduced innovative techniques to enhance IoT security. One of the most promising advancements is **blockchain-based security**, which offers decentralized protection and privacy. Blockchain technology provides rigidity, decentralization, and transparency, enabling secure data transmission and reliable remote connectivity. Additionally, its consensus mechanisms help prevent attacks such as Sybil attacks, ensuring the integrity of the network. [7]

Another major development is integrating **machine learning** into IoT security systems. Machine learning enables automated threat detection and prediction by analyzing large volumes of data. These systems can learn and adapt to new attack patterns, making them particularly effective against rapidly evolving threats. Moreover, machine learning supports real-time intrusion detection, ensuring swift responses to potential security breaches. [8]

C. Challenges and Existing Gaps

Despite the potential of these new approaches, several challenges persist. Blockchain implementations often require high computational power and significant time, which are not always compatible with the limited processing capabilities of most IoT devices. Additionally, developing adaptable and scalable security solutions remains a complex task, particularly given the rapid evolution of sophisticated cyberattack strategies. There is also a pressing need for efficient algorithms capable of processing and analyzing the massive volumes of data generated by IoT networks, which current methods often struggle to handle effectively. [9]

D. Discussion

The convergence of blockchain and machine learning presents a promising avenue for addressing IoT network security challenges. Key considerations in this area include developing resource-efficient security solutions and creating adaptive intrusion detection systems. Ensuring data integrity and confidentiality while balancing computational complexity with security effectiveness is critical. The proposed research leverages the Random Forest algorithm and blockchain technology to address these challenges, offering a more robust and adaptive approach to IoT network security.

III. Materials

Network security researchers rely on comprehensive datasets to evaluate intrusion detection systems. For this study, two pivotal datasets were selected: KDD Cup-99 [10] and NSL-KDD [11] , which provide extensive insights into network traffic patterns and cybersecurity threats.

The selection of KDD Cup-99 and NSL-KDD datasets was strategic, driven by their comprehensive nature and widespread acceptance in the network intrusion detection research community. These datasets offer: Diverse attack scenarios, Comprehensive feature representations, Standardized benchmarking capabilities, and Extensive coverage of network traffic patterns.

A. KDD Cup-99 Dataset

The KDD Cup-99 dataset, developed for the Third International Knowledge Discovery and Data Mining Tools Competition in 1999, represents a landmark resource in network intrusion detection research. Comprising approximately 4.9 million network connection records, it provides a complex landscape of network interactions. The dataset contains 41 features describing network traffic, allowing researchers to analyze intricate network behaviors.

Attack Categories in KDD Cup-99

TABLE I: Attack Categories in KDD Cup-99

Category	Attack Types	Description
DoS (Denial of Service)	Neptune, smurf, teardrop, pod, back, land, loadmodule	Attacks designed to overwhelm or disable systems, causing service unavailability
Probe	ipsweep, portsweep, nmap, satan	Scanning or reconnaissance activities to gather system information
R2L (Remote to Local)	guess_passwd, ftp_write, imap, warezclient, warezmaster, phf, perl, spy	Attempts to gain unauthorized access from remote machines
U2R (User to Root)	rootkit, buffer_overflow, multihop, loadmodule	Attempts to escalate privileges to root level
Normal	normal	Benign network traffic without malicious intent

B. NSL-KDD Dataset

Developed as an improved version of the KDD Cup-99 dataset, NSL-KDD addresses several critical limitations of its predecessor. The dataset offers: Reduced redundancy, more

balanced distribution of attack and normal instances, Elimination of duplicate records, and a More realistic representation of network traffic patterns .

C. Comparative Dataset Characteristics

TABLE II: Comparative Dataset Characteristics

Characteristic	KDD Cup-99	NSL-KDD
Total Instances	4.9 million	Reduced subset
Features	41 network traffic features	41 network traffic features
Attack Categories	4 (DoS, Probe, R2L, U2R)	4 (DoS, Probe, R2L, U2R)
Unique Advantage	Comprehensive historical data	Improved data quality and representation

D. Preprocessing and Experimental Approach

To ensure robust analysis, the following preprocessing steps were implemented: Data cleaning, Removal of redundant records, Feature selection and normalization, and Stratified sampling for training (85%) and validation (15%).

The selection of these datasets allows for a comprehensive evaluation of our proposed intrusion detection approach, providing a standardized framework for assessing machine learning model performance in network security contexts.

IV. Experimental Analysis

This work centers on the evaluation of several machine learning (ML) models in blockchain-enabled environments, specifically targeting their effectiveness in ensuring the security of real-time IoT networks. The analysis focuses on two critical metrics:

1. Accuracy: the proportion of the total number of predictions that were correct. It is given by the relation using the confusion Matrix: [13]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (1)$$

Meaning:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100 \qquad (2)$$

2.Inference time: refers to the time it takes for a machine learning model to process input data and generate a prediction or decision.

$$\text{Inference Time} = \frac{T_{\text{prediction\_End}} - T_{\text{prediction\_Start}}}{\text{Total Samples}} \qquad (3)$$

This research was carried out using the Python programming language. Several widely used Python libraries (Scikit-learn, Pandas, Numpy, Sklearn, Matplotlib, Seaborn, TensorFlow, and Keras) In order to detect intrusions in IoT networks, we developed machine learning models using these libraries. They were implemented with input data from [10] and [11]. For the Nsl KDD testing we merged KDDTest+ and KDDTrain+ and we used (Fields Names.csv) from [12] to fill the column names.

To implement the proposed Models, these steps were followed:

- Loading and Preprocesing the Datasets : which includes removing redundancies, converting categorical data into a numerical form and normalising the numerical data
- Splitting the data into training (85%) and testing (15%) sets.
- Training
- Evaluating the models

Experimental Results:

NSL-KDD (KDDmerged (KDDTrain+(125 973 samples) et KDDTest+(22 544 samples)=148 517 samples) with o.15 test size)

TABLE III: Performance of different algorithms ON NSL-KDD

Algorithm	Accuracy	Inference Time (s)
Random Forest	99.41%	0.000026
Decision Tree	99.17%	0.000003
SVM	97.97%	0.002020
KNN	98.81%	0.000367
Logistic Regression	97.14%	0.000000
Adaboost	81.65%	0.000099
Naïve Bayes	36.64%	0.000022

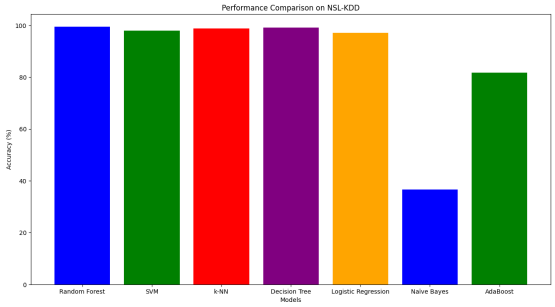

Fig. 1: Attack detection accuracy on NSL-KDD Dataset

Overall Performance of Algorithms:

Random Forest achieves the highest accuracy (99.41%),followed closely by Decision Tree (99.17%) and KNN (98.81%).

Adaboost (81.65%) and Naïve Bayes (36.64%) show relatively poor performance, making them unsuitable for this task.

Inference Time:

Decision Tree has the shortest inference time (0.000003 s), making it ideal for real-time systems.

Random Forest, while slightly slower (0.000026 s), remains competitive due to its exceptional accuracy, making it suitable for high-reliability systems.

SVM (0.002020 s) and KNN (0.000367 s) have relatively higher inference times, limiting their applicability in time-critical environments.

KDD CUP 99: 1 074 992 samples with 0.15 test size

TABLE IV: Performance of different algorithms On KDD Cup 99

Algorithm	Accuracy	Inference Time (s)
Random Forest	99.98%	0.000032
Decision Tree	99.98%	0.000005
SVM	99.91%	0.000983
KNN	99.95%	0.001733
Logistic Regression	99.79%	0.000000
Adaboost	97.40%	0.000061
Naïve Bayes	74.30%	0.000011

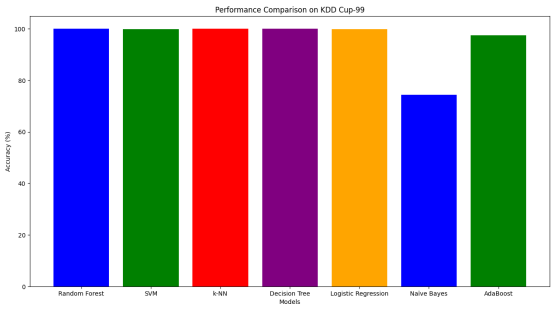

Fig. 2: Attack detection accuracy on KDD Cup 99 Dataset

Overall Performance of Algorithms:

Random Forest and Decision Tree achieve the highest accuracy (99.98%), followed by KNN (99.95%) and SVM (99.91%).

Logistic Regression (99.79%) and Adaboost (97.40%) show slightly lower performance, while Naïve Bayes has the weakest performance (74.30%).

Inference Time:

Decision Tree has the fastest inference time (0.000005 s), making it ideal for systems requiring rapid execution.

Random Forest follows with a slightly higher time (0.000032 s), but remains competitive due to its high accuracy.

Naïve Bayes, despite having a fast inference time (0.000011 s), is ineffective due to its low accuracy.

Although Several studies have been conducted to evaluate the effectiveness of various machine learning techniques for intrusion detection in IoT systems. In this paper,not only did we compare and analyse the reported accuracies of different machine learning algorithms but we also evaluated the inference time of each Model.

V. Discussion

The results of this research offer important perspectives on the possibility of integrating machine learning (ML) algorithms with blockchain technology to improve the security of networks. By prioritizing both prediction accuracy and inference time, our study assesses the practicality of implementing these models in real-time IoT scenarios where quick decision-making is essential.

A. Key findings and model performance:

Our results show the consistent outperformance of Random Forest compared to other models in terms of both accuracy and inference time on both datasets. On the KDD CUP-99 dataset, Random forest achieved an accuracy of 99.98% with an inference time of 0.000032 seconds, while on the NSL-KDD dataset, it reached 99.41% accuracy with an inference time of only 0.000026 seconds. These metrics showcase the strength of Random Forest in combining strong predictive capabilities with low computational latency, making it ideal for real-time intrusion detection.

Alternative models, such as Decision Trees and Support Vector Machines (SVM), showed comparable accuracy levels but differed in inference time. Decision Trees performed similarly to Random Forest, but with slightly lower accuracy. Meanwhile, SVM was notably accurate but demanded increased inference times due to its computational complexity, which restricted its use in latency-sensitive environments.

Naïve Bayes and AdaBoost models demonstrated lower performance due to their dependence on more simpler assumptions and their vulnerability to data imbalance. Their moderate accuracy highlights the necessity for enhanced feature selection and more effective data preprocessing techniques.

B. Comparative Analysis and Insight:

The exceptional efficacy of Random Forest stems from its ensemble learning mechanism, which reduces overfitting by synthesizing predictions from various decision trees. This method guarantees both precision and robustness against data variability. In contrast, SVM's strength resides in its capability to address non-linear decision boundaries, yet its considerable computational demands render it less feasible for real-time applications.

The comparative outcomes also underscore the necessity of selecting models based on specific network conditions. For example, although KNN reached nearly flawless accuracy on the KDD Cup-99 dataset, its inference duration was significantly higher due to its instance-based learning approach.

C. Real-World implications:

The results indicate that Random Forest is exceptionally well-suited for real-time threat detection due to its minimal inference time and high precision. Decision Trees offer a simpler option for devices with limited resources. Conversely, SVM and KNN are more relevant in settings where computing resources are less restricted and instant responses are not as essential.

Integrating machine learning with blockchain fortifies security by merging real-time anomaly detection with mutable logging and decentralized decision-making. This two-tiered system guarantees data clarity and effectively decreases risks.

D. Obstacles and Constraints:

In spite of these encouraging outcomes, various challenges persist. The datasets employed, NSL-KDD and KDD Cup-99, while recognized in the research community, may not

thoroughly capture modern attack types due to their age. This shortcoming underscores the necessity for current and varied datasets that reflect up-to-date cybersecurity challenges.

A further challenge is scalability. Although our models excelled in controlled experiments, their efficacy in larger, real-world network scenarios is still uncertain. Inference time could rise with increasing data volumes, requiring more scalable algorithms and enhanced infrastructure.

VI. Conclusion

By leveraging blockchain's decentralized nature and ML's predictive capabilities, we developed a system capable of detecting anomalies and malicious behavior efficiently.

The accuracy of machine learning models plays a pivotal role in ensuring that potential threats are detected, thereby maintaining the reliability of IoT devices.

Similarly, inference time is equally critical in such real-time environments, where rapid detection and response are necessary to prevent breaches and disruptions in the network. Our results demonstrate that by optimizing both accuracy and inference time, the proposed solution can serve as a robust framework for securing IoT networks: In both the NSL-KDD and KDD CUP 99 datasets, Random Forest and Decision Tree perform the best. Random Forest offers high accuracy, making it ideal for applications requiring reliability, while Decision Tree is faster, making it suitable for real-time systems.

Blockchain further complements this system by ensuring tamper-proof logging, traceability, and trust, which are essential in distributed IoT environments.

References

[1] https://www.eccouncil.org/cybersecurity-exchange/ethical-hacking/the-rise-of-iot-attacks-endpoint-protection-via-trending-technologies/#:~:text=The%20rise%20in%20IoT%20attacks,monitored%20(Petrosyan%2C%20202)

[2] https://www.chakray.com/blockchain-iot-security/#:~:text=For%20IoT%20safety%2C%20the%20blockchain,for%20a%20trusted%20third%20party

[3] https://www.researchgate.net/publication/377115052_Machine_learning_techniques_for_IoT_security_Current_research_and_future_vision_with_generative_AI_and_large_language_models

[4] https://theses.hal.science/tel-04563020v1/file/2023TOU30302.pdf

[5] https://www.researchgate.net/publication/380941640_Security_Privacy_Issues_in_IoT_Using_Blockchain_and_ML

[6] A. M. Ferrag, M. N. Moustafa, "The integration of blockchain and AI for IoT security," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3344–3356, May 2021.

[7] M. Conoscenti, A. Vetro, and J. C. De Martin, "A survey on blockchain for IoT security: Applications, challenges, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1452–1481, 2018.

[8] L. Zhang, H. Luo, and W. Wei, "The role of machine learning in IoT security: An overview," *Scientific Reports*, vol. 11, pp. 3456–3467, 2021.

[9] ISACA, "Security issues in IoT: Challenges and countermeasures,"*ISACA Journal*, vol. 4, pp. 60–70, 2020.

[10] https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[11] https://www.kaggle.com/datasets/hassan06/nslkdd

[12] https://github.com/jeroenvansaane/Deep-Learning-Based-Intrusion-Detection-NSL-KDD/tree/master

[13] https://www.researchgate.net/publication/334276076_Intrusion_Detection_System_Classification_Using_Different_Machine_Learning_Algorithms_on_KDD-99_and_NSL-KDD_Datasets_-_A_Review_Paper