International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

Cyber Hacking Breaches Prediction Using Machine Learning

Dr. Prabakaran Narayan[1], B.Vamshi[2], M.Supraja[3], B.Rama Sudha[4], K. Vijay Sai[5]

[1] Professor, [2,3,4,5]UG Students

[1,2,3,4,5]Department of Computer Science and Engineering, Sai Rajeswari Institute of Technology

Abstract. Cyber-physical systems (cps) have made significant progress in many dynamic applications due to the integration between physical processes, computational resources, and communication capabilities. However, cyber-attacks are a major threat to these systems. Unlike faults that occurs by accidents cyber-physical systems, cyber-attacks occur intelligently and stealthy. Some of these attacks which are called deception attacks, inject false data from sensors or controllers, and also by compromising with some cyber components, corrupt data, or enter misinformation into the system. If the system is unaware of the existence of these attacks, it won't be able to detect them, and performance may be disrupted or disabled altogether. Therefore, it is necessary to adapt algorithms to identify these types of attacks in these systems. It should be noted that the data generated in these systems is produced in very large number, with so much variety, and high speed, so it is important to use machine learning algorithms to facilitate the analysis and evaluation of data and to identify hidden patterns. In this research, the CPS is model as a network of agents that move in union with each other, and one agent is considered as a leader, and the other agents are commanded by the leader. The proposed method in this study is to use the structure of deep neural networks for the detection phase, which should inform the system of the existence of the attack in the initial moments of the attack. The use of resilient control algorithms in the network to isolate the misbehave agent in the leader-follower mechanism has been investigated. In the presented control method, after the attack detection phase with the use of a deep neural network, the control system uses the reputation algorithm to isolate the misbehave agent. Experimental analysis shows us that deep learning algorithms can detect attacks with higher performance that usual methods and can make cyber security simpler, more proactive, less expensive and far more effective.

Keywords: SVM, Decision Tree, Random Forest, Extra Tree Classifier, Cat Boost Classifier and XG Boost Classifier.

I. Introduction

Motivation:
Predicting cyber hacking breaches using machine learning is crucial to proactively defend against cyber threats. By analyzing historical attack patterns and identifying potential vulnerabilities, organizations can fortify their security measures, mitigate risks, and prevent devastating data breaches. Machine learning empowers us to detect emerging attack vectors and adapt defenses rapidly, safeguarding sensitive information and ensuring business continuity. A proactive approach to cyber threat prediction is essential in this constantly evolving digital landscape, enabling organizations to stay one step ahead of cybercriminals and protect their assets, reputation, and customer trust.

Problem Statement:
The problem is to develop a machine learning model that predicts cyber hacking breaches with high accuracy. This model will analyze historical breach data, user behavior, network vulnerabilities, and system logs to identify patterns and indicators of potential attacks. By effectively predicting these breaches, organizations can proactively implement security measures to prevent or mitigate cyber-attacks, safeguarding sensitive data and maintaining business continuity. The ultimate goal is to enhance cyber security posture and protect against emerging threats, minimizing the impact of data breaches and ensuring a secure digital environment for businesses and individuals alike.

Objective of the Project:
The primary goal of this project is to determine the Cyber hacking breaches whether there will be attack or not and to know this we have used the Support Vector , Decision Tree, Random forest, Extra Tree Classifier, Cat Boost Classifier and XG Boost Classifier classification techniques.

Scope:
The scope of Cyber hacking breaches prediction using machine learning involves leveraging advanced algorithms to analyze historical cyber-attack data, network vulnerabilities, and user behavior patterns. Through this, the system aims to identify potential security breaches before they occur, enabling proactive measures for risk mitigation and incident response. Machine learning models can aid in real-time threat detection, anomaly recognition, and predictive analytics, enhancing cyber security posture and safeguarding sensitive data and systems from cyber threats.

Project Introduction:
Recent advances in technology have led to the introduction of cyber-physical systems, which due to their better computational and communicational ability and integration between physical and cyber-components, has led to significant advances in many dynamic applications. But this improvement comes at the cost of being vulnerable to cyber- hacking .These attacks can be detected by system monitoring in the system. But if the attacker can plan a high-level attack to prevent himself from being identified, these attacks are called stealthy deception attacks, and other common methods of counteracting such attacks will not work. Therefore, it is important to be aware of the attacks

that occur in order to respond in a timely manner to attackers. In other words, the security system must be aware of the attack, otherwise it will not be able to detect and control the attack

II. Objective

- To design a machine learning-based system : that can predict and detect cyber hacking breaches from network or system data.
- To collect and analyze cyber security datasets: containing normal and malicious activities for training the detection model.
- To pre-process the data and extract important features: that help in identifying cyber attack patterns.
- To implement multiple machine learning algorithms: such as Decision Tree, Random Forest, Support Vector Machine, or Neural Networks for breach detection.
- To train and test the developed models: in order to classify activities as normal or malicious.
- To evaluate the performance of the models: using metrics like Accuracy, Precision, Recall, and F1-Score.
- To compare the performance of different algorithms and identify: the most effective model for cyber attack detection.
- To develop an efficient system: that can help organizations detect potential cyber threats early and improve network security.

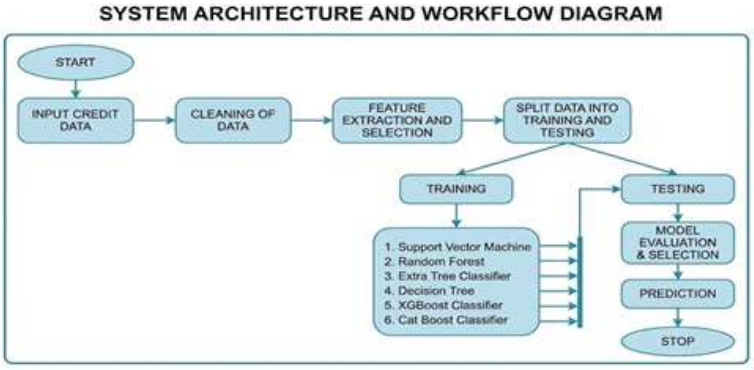
III. Activity Diagram

The activity diagram represents the workflow of the proposed system for predicting and detecting cyber hacking breaches using machine learning techniques. It shows the sequence of activities involved in processing the data and generating predictions. The process begins with collecting cyber security datasets from various sources. The collected data is then pre-processed to remove noise, handle missing values, and prepare it for analysis. After pre-processing, the dataset is divided into training and testing data.

IV. Basic Activity Diagram Symbols and Notations

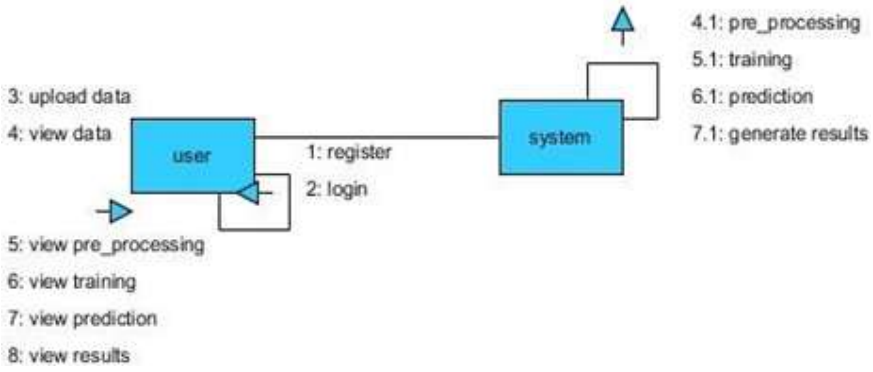
- Initial Node (Start): This symbol represents the starting point of the process in the activity diagram.
- Action State: An action state represents a specific activity or task performed in the system, such as data pre-processing, training the model, or testing.
- Control Flow: Control flow arrows show the direction and sequence of activities from one step to another in the diagram.
- Decision Node: A decision node represents a condition where the process can branch into different paths based on a specific condition.
- Merge Node: A merge node is used to combine multiple alternative paths into a single flow in the activity diagram.
- Object Flow: Object flow represents the movement of data or objects between different activities in the system.
- Final Node (End): This symbol indicates the completion or termination of the process in the activity diagram.

- Activity Partitions: are used to divide the activity diagram into different sections, representing different actors, components or modules of the system.



V. Collaboration Diagram

In collaboration diagram the method call sequence is indicated by some numbering technique as shown below. The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram. The method calls are similar to that of a sequence diagram. But the difference is that the sequence diagram does not describe the object organization whereas the collaboration diagram shows the object organization.



Collaboration Diagram for overall project

VI. Feature Selection Module

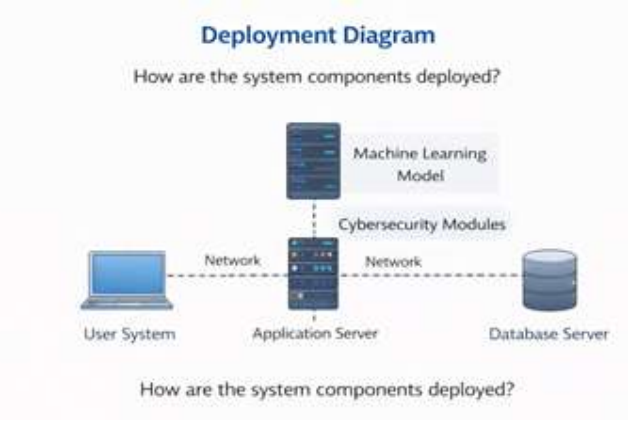
Module is used to identify the most important features from the cyber security dataset that help in detecting and predicting cyber hacking breaches. Since large datasets contain many attributes, some of them may be irrelevant or redundant. Feature selection helps in selecting only the useful features that improve the performance of the machine

learning models. In this project, feature selection methods such as Correlation-based Feature Selection (CFS), Information Gain, and Chi-Square Test are used to analyze the importance of each feature in the dataset. These methods evaluate the relationship between the input features and the target variable to determine which attributes contribute most to cyber attack detection. The selected features are then used to train machine learning algorithms such as Support Vector Machine (SVM), Random Forest, Decision Tree, and Naive Bays. These algorithms help in both prediction and detection of cyber hacking breaches by classifying network activities as normal or malicious. By using feature selection techniques, the system reduces data complexity, improves training efficiency, and increases the accuracy of the prediction and detection process.

VII. Deployment Diagram

Deployment diagram represents the deployment view of a system. It is related to the component diagram. Because the components are deployed using the deployment diagrams. A deployment diagram consists of nodes. Nodes are nothing but physical hardware's used to deploy the application.

- A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.
- The deployment diagram represents the physical architecture of the system,
- It illustrates the communication between different nodes ?
- The diagram shows how the machine learning model ?
- It helps in understanding how the system operates in a real environment.



VIII. Test Plan

Test plan is a general document for entire project, which defines the scope, approach to be taken and the personal responsible for different activities of testing. The inputs for forming test plan are following.

- Project plan.
- Requirements document.

IX. Test Case Specification

he testing process begins with verifying the dataset input to ensure that the system accepts and processes the cyber security data correctly. After that, the data pre-processing module is tested to check whether the system properly cleans and prepares the data for analysis.

Test Case Execution and Analysis
Test case execution is the process of running the defined test cases on the system to verify its functionality. Each test case is executed using the given inputs and the outputs are observed. The results obtained are compared with the expected results to check whether the system works correctly. If the results match the expected output.

X. Implementation

The implementation of the project "Cyber Hacking Breaches Prediction and Detection Using Machine Learning" focuses on developing a system that can analyze cyber security data and identify possible hacking activities. Machine Learning techniques are used to automatically detect malicious behaviour in network traffic and system data. Python is used for implementing the system because it provides powerful libraries for data analysis and machine learning. The system processes large datasets of network activities and uses machine learning algorithms to predict and detect cyber attacks. The implementation helps organizations identify security threats early and take preventive actions to protect sensitive information.

Key functions implemented in the system:
- Anomaly Detection – Identifies unusual patterns in network traffic that may indicate a cyber attack.
- Malware Detection – Classifies files or activities as normal or malicious using machine learning models.
- Intrusion Detection – Detects unauthorized access or suspicious activities in the network.
- Phishing Detection – Identifies fraudulent messages, emails, or malicious links.
- Threat Prediction–Uses historical data to predict possible cyber breaches in the future.

XI. Materials and Methods

Materials:
- Cyber security Dataset – Used to train and test the machine learning models for detecting cyber attacks.
- Python Programming Language – Used to implement the machine learning algorithms.
- Machine Learning Libraries – Libraries such as Numpy, Pandas, Scikit-learn, and Matplotlib are used for data processing and model development.

- Computer System – Used for running the program and processing the dataset.

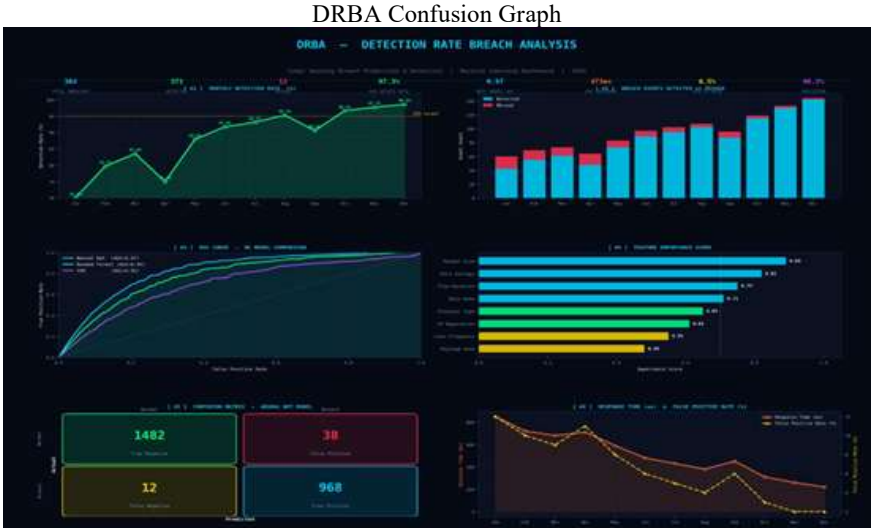
Methods:

- Data Collection – Gathering cyber security datasets containing normal and malicious activities.
- Data Pre-processing – Cleaning the dataset by removing missing or unnecessary data.
- Feature Selection – Selecting important features that help in identifying cyber attacks.
- Model Training – Applying machine learning algorithms such as Decision Tree, Random Forest, and SVM.
- Prediction and Detection – Testing the trained model to predict and detect cyber hacking breaches

XII. Experimental Results and Discussion

Different machine learning algorithms such as Decision Tree, Random Forest, and Support Vector Machine were applied to analyze the dataset. The models were trained using the training data and tested using the testing data to classify network activities as normal or malicious. The experimental results showed that machine learning techniques can effectively detect cyber threats by analyzing patterns in the data. The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1-score. The results indicate that the proposed system can successfully identify potential cyber hacking breaches and improve network security. The discussion highlights that using machine learning algorithms helps in detecting cyber attacks more efficiently compared to traditional security methods.

DRBA Confusion Graph



XIII. Conclusion

In this study, an attempt was made to use the resilient control consensus method in complex discrete cyber-physical networks with a number of local Cyber hacking breaches off. By applying this control method, it was observed that even in the presence of Cyber hacking breaches, the system can remain stable and isolate the Cyber hacking node and the performance of the system is not weakened.

Using the neural network used in this study, it was observed that with a deep neural network, with 7 hidden layers, the system shows better performance. Also in a recurrent neural network integrated with a deep neural network, a deep layer network with a linear function performs better. Therefore, it can be said that the system has less complexity. So With deep learning method, systems can analyze patterns and learn from them to help prevent similar attacks and respond to changing behaviour.

In short, machine learning can make cyber security simpler, more proactive, less expensive and far more effective. After observing the state of the system reported by the neural network, the control system makes decisions based on it and, if there is an Cyber hacking, detects it and isolates it, so as not to have a detrimental effect on the behaviour of other agents. In future research, more attacks on agents can be considered, also data mining and other machine learning methods, such as support vector machine (SVM) algorithms or other types as recurrent Cat Boost to evaluate system performance improvements.

References

1. Kwon, Cheol yeon, Weyi Liu, and Niseko Hwang. ”Security analysis for cyber-physical systems against stealthy deception attacks.” In 2013 American control conference, IEEE (2013): 3344-3349.
2. Pajic, Miroslav, James Weimer, Nicola Bezzo, Oleg Sokolsky, George J. Pappas, and In sup Lee. ”Design and implementation of attack-resilient cyber physical systems: With a focus on attack-resilient state estimators.” IEEE Control Systems Magazine 37, no. 2 (2017): 66-81.
3. Sheng, Long, Ya-Jun Pan, and Xiang Gong. ”Consensus formation control for a class of networked multiple mobile robot systems.” Journal of Control Science and Engineering 2012 (2012).
4. Zeng, Wente, and Mo-Yuen Chow. ”Resilient distributed control in the presence of misbehaving agents in networked control systems.” IEEE transactions on cybernetics 44, no. 11 (2014): 2038-2049.
5. Sun, Hongtao, Chen Peng, Taicheng Yang, Hao Zhang, and Wangli He. ”Resilient control of networked control systems with stochastic denial of service attacks.” Neurocomputing 270 (2017): 170-177.