CSE 337L: Cryptography Course Project

Submission Date: 08 / 04/ 2025

Project Title:

Insider Threat Detection Using Behavioral Biometrics

Group Members:

1) AP22110010051,Bhargav K,venkatabhargav_kadiyam@srmap.edu.in

2) AP22110010048,Sreemanth V,sreemanth_vaddi@srmap.edu.in

3) AP22110010148,Naresh B,hemnareshreddy_batchu@srmap.edu.in

4) AP22110010439,Bhaskar Y,pardhivsai_prathipati@srmap.edu.in

5) AP22110010785,Pardhiv P,sribhaskar_yendluri@srmap.edu.in

A) Purpose of the project:

To detect insider threats—malicious or unintentional harmful activities carried out by individuals within an organization—by analyzing user behavior patterns using machine learning techniques.

B) Why it is relevant (explain with example):

Insider threats are tough to detect because they come from people we trust—like employees. With so much happening online now, it's easier for mistakes or harmful actions to go unnoticed. This project helps spot unusual behavior early using machine learning, so companies can stay safe and avoid big problems.

C) How you will solve the problem:

We'll use machine learning to analyze user activity data from the CERT dataset. First, we'll clean and organize the data, then extract useful features like login times, file access, and email count. After that, we'll train models like KNN to learn the difference between normal and suspicious behavior. Finally, we'll test the model to see how well it can detect insider threats.

D) What are the other possible ways (if any) to solve it:

1. **Rule-Based Systems**
   Use predefined rules (e.g., "alert if login after midnight")—simple but not adaptive.

2. **Anomaly Detection (Unsupervised Learning)**
   Algorithms like Isolation Forest or One-Class SVM can flag unusual behavior without needing labeled data.

3. **Deep Learning**
   Use RNNs or LSTMs to detect patterns in sequences of user actions—good for time-based behavior tracking.

4. **Behavioral Analytics & UEBA (User and Entity Behavior Analytics)**
   Track and learn normal behavior over time to spot deviations.

5. **Log Analysis & SIEM Tools**
   Use tools like Splunk or ELK stack to analyze system and network logs for suspicious activity.

6. **Hybrid Approaches**
   Combine rule-based, machine learning, and human monitoring for better accuracy and context.

E) Advantages of using the chosen approach:

1. **Simple and Easy to Implement**
   KNN and similar models are straightforward and don't require heavy tuning.

2. **Good for Small to Medium Data**
   They work well when the dataset isn't too large or complex.

3. **No Assumptions About Data**
   Models like KNN don't assume anything about how the data is distributed.

4. **Interpretable Results**
   Easier to understand and explain predictions compared to deep learning models.

5. **Effective with Well-Chosen Features**
   Can perform well if the right behavioral features are selected.