CHAPTER 3

METHODOLOGY

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

Insider threat is a passive attack and one of the emerging security challenges that requires a potential approach for detection and mitigation. It is essential to devise a solution to detect intentional and unintentional insiders using machine learning and mitigation using user authentication with a user profiling mechanism.

This chapter overviews the research methodology, techniques proposed, and outcomes achieved in the present research work.

3.2 PROBLEM SPECIFICATION

The proposed approach uses two publicly available datasets, namely CERT insider threat and CIC Darknet datasets. Since the dataset contains log activities of users generated from diverse sources, it requires further preprocessing. In preprocessing, sampling using Nearmiss2 is tuned to better handle the class imbalance problem. For insider threat detection, the B-SVM algorithm is proposed using Support Vector Machine (SVM) and Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) to detect genuine, intentional, and unintentional insiders. After detection, the unintentional insiders are mitigated using the combination of Clonal Kernel Principal Component Analysis (CKPCA) and Deep Belief Network (DBN). The intentional insiders are mitigated using a user profiling mechanism.

3.3 OVERALL METHODOLOGY

The proposed method attempts to devise an approach for detecting and mitigating both intentional and unintentional insiders. Figure 3.1 shows the overview of the overall methodology.

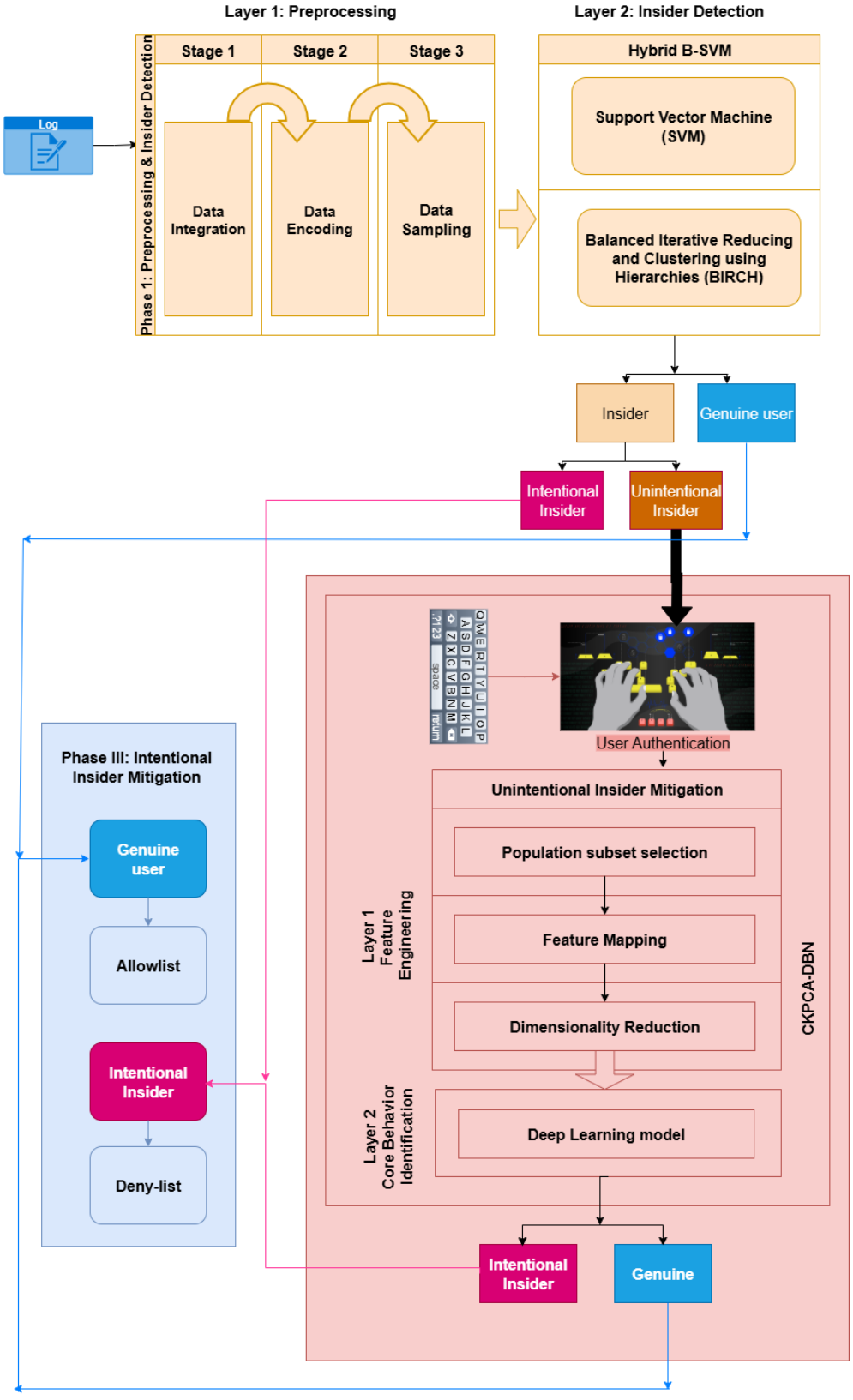

Figure 3.1: Methodology Overview

The entire methodology is divided into three phases, namely, Preprocessing & Insider Detection (P&ID), Unintentional Insider Mitigation (UIM), and Intentional Insider Mitigation (IIM) after analyzing the log activities. Significant contributions are made in the phases of Preprocessing and Insider Detection, Unintentional Insider Mitigation, and Intentional Insider Mitigation.

3.3.1 PREPROCESSING & INSIDER DETECTION (P&ID)

The purpose of P&ID is to preprocess, detect, and classify both intentional and unintentional insider threats while handling class imbalance problems and minimized misclassification rates. It can be attained by proposing a tuned Nearmiss2 sampling technique with B-SVM (SVM and BIRCH) to detect both intentional and unintentional insiders.

The log data obtained from diverse sources are combined using data integration, which contains data representation in categorical and numerical formats. Transforming categorical features of the log data into numerical values is done to ensure data representation. This transformation is facilitated by converting the categorical value into numerical format. Next, the numerical data undergoes sampling through various undersampling and oversampling techniques. Sampling using the Nearmiss2 is fine-tuned to ensure a balanced dataset, which is crucial for accurate classification in subsequent stages.

The processed and sampled data is then given as input to a proposed hybrid B-SVM machine learning algorithm to classify intentional and unintentional insider threats. Initially, an SVM algorithm detects genuine and intentional insiders based on user activity patterns. Then, the BIRCH algorithm is employed to re-evaluate any instances falsely detected as intentional insiders in SVM algorithm, classifying them into genuine, intentional, or unintentional insiders. This approach enhances the accuracy of insider threat detection by refining classification at each stage.

3.3.2 UNINTENTIONAL INSIDER MITIGATION (UIM)

UIM aims to mitigate unintentional insiders detected from the P&ID by enhancing detection accuracy and minimizing the error rates in the user authentication system. It can be achieved by proposing a feature engineering technique, namely Clonal Kernel Principal

Component Analysis with Deep Belief Network (CKPCA-DBN), to mitigate unintentional insiders with low false alarm rates. The recognized unintentional insiders are mitigated using enhanced keystroke biometrics by combining CKPCA with the DBN technique for user authentication.

CKPCA combines population subset selection, feature mapping, and dimensionality reduction. In population subset selection, a clonal selection algorithm is used to isolate the most distinctive behavioral traits of the users, which minimizes the false alarm rate. In feature mapping, the features of the selected population are extended into higher dimensional features with kernel mean embedding. This transformation aims to expand the feature representations and capture the subtle nuances in biometric information that facilitate an enhanced analysis of individual behavioral patterns.

Dimensionality reduction using Principal Component Analysis (PCA) is incorporated to structure the data efficiently. It reduces the complexity of the higher dimensional features while retaining essential variance, making it easier to categorize users effectively. The proposed approach of enhanced feature engineering technique, namely CKPCA, is expected to give intricate keystroke biometric features.

Subsequently, the core behavior of the user is identified using DBN to build the Restricted Boltzmann Machine (RBM) model for uncovering hidden keystroke patterns in complex data. The DBN is trained to analyze individual biometric features through feature extraction, pre-training, and fine-tuning. The users are classified into genuine or intentional insiders through user authentication. The intricated keystroke features are trained using Deep Belief Network to classify genuine and imposter users based on individual biometric features, achieving high accuracy and a low equal error rate.

By integrating this enhanced feature engineering technique with a DBN classifier, precise distinctions based on user-specific biometric data are achieved while minimizing the false alarm rate. This approach effectively distinguishes between genuine and intentional insiders by emphasizing the utilization of biometric authentication in securing sensitive systems.

3.3.3 INTENTIONAL INSIDER MITIGATION (IIM)

The purpose of IIM is to mitigate intentional insiders through the user profiling mechanism. The user profiling mechanism analyzes genuine and intentional insider behavior and profiles them into Allowlist and Denylist. It can be achieved by developing a Decision Tree classifier to assess the risk level of users based on their authentication outcomes and other relevant features. Users are then profiled to an Allowlist or Denylist based on their predicted risk level in the IIM phase.

The detailed profiles of genuine and intentional insiders are created. Initially, genuine insiders are identified and profiled based on the detection outcomes from the P&ID phase and the mitigation outcomes from UIM phase. These profiles are then added to an Allowlist, ensuring that authorized users are recognized and allowed appropriate system access. The profiles of individuals who exhibit behaviors consistent with insider threats are added to Denylist and deemed to restrict their access. It leverages Allowlist and Denylist and enhances the ability of the overall methodology to differentiate between genuine and intentional insiders accurately. As a result, the mitigation strategy strengthens security and streamlines user profiling to ensure that only authorized users maintain access while potential threats are effectively contained.

3.4 TECHNIQUES PROPOSED & OUTCOME ACHIEVED

The techniques proposed, and the outcomes achieved in each phase of the proposed work is visualized in Figure. 3.2.

In phase 1 (P&ID) tuned Nearmiss2 based sampling with B-SVM is proposed for detecting both intentional and unintentional insiders. In phase 2 (UIM), CKPCA-DBN for user behavioral biometrics is proposed to mitigate unintentional insiders. Intentional insiders are mitigated in phase 3 (IIM) by profiling genuine users into Allowlist and insiders into Denylist.

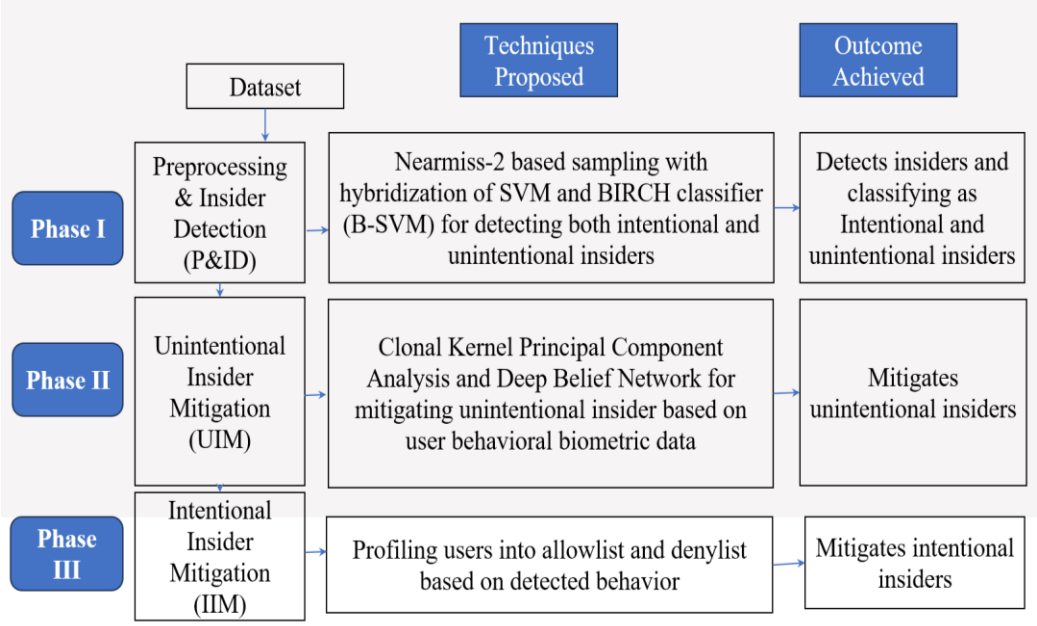

Figure 3.2 Techniques Proposed and Outcome Achieved

3.5 DATASET USED

The proposed methodology is evaluated using two datasets such as CIC darknet dataset and CERT insider dataset. CERT Insider dataset contains the working behavior of both genuine and insider in an organization. CIC Darknet dataset contains the traffic information of hidden and anonymous network. The detail about both datasets are discussed in forthcoming chapters 4.

3.6 TOOLS USED

The working environment for performing experiment is specified below. The proposed methodology used anaconda platform to observe the performance of machine learning models in detecting and mitigating both intentional and unintentional insider threats. The threat models are equipped in Jupyter Notebook containing basic libraries supporting Python 3.9 in Dell Latitude i7 Intel core laptop. Meanwhile, installation can be done if other libraries are required.

3.7 CHAPTER SUMMARY

This chapter discussed the research design. All three phases are based on the five-stage methodology to meet the objectives of the thesis. The overview of all the phases are detailed in this chapter. The proposed work and outcome achieved in each phase are described in this chapter. The forthcoming chapters 4, 5, and 6 elaborately discuss all the phases that are mentioned in this chapter. The next chapter describes the phase 1 (P&ID) in detail.