ScienceDirect

Computers and Electrical Engineering

Volume 105, January 2023, 108519

Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment

Asha S [a], Shanmugapriya D (Assistant Professor and Head) [b], Padmavathi G (Professor and Dean) [a]

[a] Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamilnadu 641043, India

[b] Department of Information Technology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamilnadu 641043, India

Received 11 August 2022, Revised 19 November 2022, Accepted 27 November 2022, Available online 2 December 2022, Version of Record 2 December 2022.

https://doi.org/10.1016/j.compeleceng.2022.108519

Get rights and content

Abstract

Machine learning (ML) techniques have currently been exploited for malicious insider threat (MIT) detection. The data variation between malicious and genuine user influences the ML model to misinterpret a malicious insider. Hence, the class imbalance problem (CIP) remains a challenging one. Regardless of the CIP in MIT detection, past research has a significant shortfall in deploying diverse sampling methods. i.e., undersampling and oversampling approach. This study proposed a novel double-layer architecture for MIT detection. The initial layer involves integration, transformation, and sampling system of data. In the sampling system, an efficient sampling approach is adopted to depreciate CIP among eight sampling techniques, depending on the performance of support vector machine (SVM) classifier. Nearmiss2 (NM-2) excels and is considered an optimal sampling technique. In the second layer, sampled data of NM-2 is employed in an anomalous MIT detection model using various anomaly detection techniques and evaluated with performance metrics. The main focus is to validate the solution for CIP in anomaly detection techniques with previous research. The proposed double-layer architecture with NM-2 and One-class SVM obtained recall and f-score of 100% and 78.72%. In contrast, it exhibits an accuracy of 82.46%, with a reasonable detection rate for MIT detection

Introduction