


i-2NIDS Novel Intelligent Intrusion Detection Approach for a Strong Network Security

Sabrine Ennaji, Sidi Mohamed Ben Abdellah University, Morocco*

 <https://orcid.org/0000-0003-2591-9673>

Nabil El Akkad, National School of Applied Sciences, Morocco

Khalid Haddouch, National School of Applied Sciences, Morocco

ABSTRACT

The potential of machine learning mechanisms played a key role in improving the intrusion detection task. However, other factors such as quality of data, overfitting, imbalanced problems, etc. may greatly affect the performance of an intelligent intrusion detection system (IDS). To tackle these issues, this paper proposes a novel machine learning-based IDS called i-2NIDS. The novelty of this approach lies in the application of the nested cross-validation method, which necessitates using two loops: the outer loop is for hyper-parameter selection that costs least error during the run of a small amount of training set and the inner loop for the error estimation in the test set. The experiments showed significant improvements within NSL-KDD dataset with a test accuracy rate of 99.97%, 99.79%, 99.72%, 99.96%, and 99.98% in detecting normal activities, DDoS/DoS, Probing, R2L and U2R attacks, respectively. The obtained results approve the efficiency and superiority of the approach over other recent existing experiments.

KEYWORDS

Hyper-Parameter Selection, Intrusion Detection System, Machine Learning, Nested Cross-Validation, Network Security

INTRODUCTION

Due to the overuse of the internet and recent technologies revolution, we are drowning in a rampant growth on a massive amount of data (Behera & Bhaskari, 2017). Furthermore, people need to disclose their personal information and exchange sensitive data to be connected, communicate with each other and to benefit from other upsides of the cyberspace like e-commerce, online works, cloud storage, etc. Therefore, the safety and confidentiality of the internet user's information has become more vulnerable towards intrusions and attacks. Many research studies are well carried out to shed light on Intrusion Detection Systems (IDSs), which are a proficient software system of detecting intrusive activities

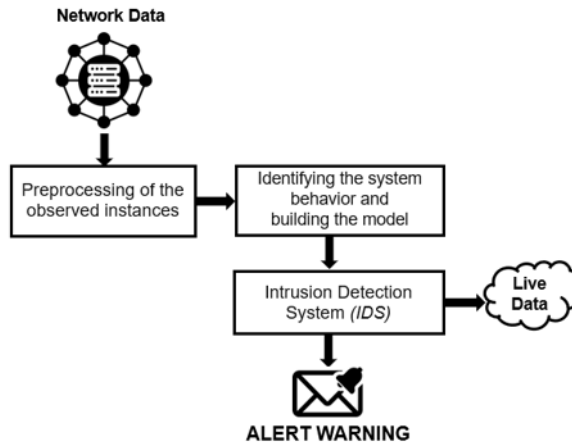
DOI: 10.4018/IJISP.317113

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

by examining all traffic flow over different environments and all internet technologies (Ramdane & Chikhi, 2014; Shukla & Singh, 2019). However, its performance is still need to be updated and improved, as long as an IDS necessitates an additional maintenance effort and human intervention (Ennaji et al., 2021). Additionally, it frequently notifies the users about false positives more than it does to real intrusions (Patel et al., 2012).

Figure 1. Architecture design of IDS based on machine learning



To fill this void, a vast majority of researchers have been opting for machine learning algorithms. The latter are widely applied in dealing with the limitations of intrusion detection systems, since they have a high potential in terms of making better identification and prediction of security threats without any intervention from the user (Stone, 1974). However, an intelligent IDS cannot make good predictions when the parameters are incorrectly selected and also because of the classification issues, such as; underfitting, overfitting, imbalanced data, etc.

For this reason, there is a useful technique, namely; cross-validation. It is considered as a resampling procedure for the determination and the selection of the appropriate parameters, which cost least test error. It is a well-known evaluation method for machine learning models that shows how well the latter will perform to an independent test data that has not been used during the training phase of the model (Stone, 1974). This approach proceeds by splitting the cleaned dataset into k-chunks of equal size. The first partition is considered as a validation set, and the model is fitted on the remaining k-1 partitions that present the training partitions. Then, the analysis is performed on each fold. Finally, it takes the average of scores of all partitions, which presents the overall estimate error. Hence, the cross-validation technique provides a better utilization of the data and it comes in different types. The most commonly used are:

- **Holdout cross-validation:** The simplest type of cross-validation approach. It randomly separates the data into training and test sets. The more data is used for the model's training, the better its performance will be.
- **K-Fold cross-validation:** The dataset is equally split into k folds, then the holdout approach is repeated k-times until each fold is considered as test set and other k-1 folds as training set.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/i-2nids-novel-intelligent-intrusion-detection-approach-for-a-strong-network-security/317113

Related Content

AI and Machine Learning Applications for Preserving Privacy and Data Leakage of E-Commerce Data

Jatin Arora, Gaganpreet Kaur, Monika Sethi and Saravjeet Singh (2025). *Strategic Innovations of AI and ML for E-Commerce Data Security* (pp. 59-78).

www.irma-international.org/chapter/ai-and-machine-learning-applications-for-preserving-privacy-and-data-leakage-of-e-commerce-data/356671

An Effective and Computationally Efficient Approach for Anonymizing Large-Scale Physical Activity Data: Multi-Level Clustering-Based Anonymization

Pooja Parameshwarappa, Zhiyuan Chen and Gunes Koru (2020). *International Journal of Information Security and Privacy* (pp. 72-94).

www.irma-international.org/article/an-effective-and-computationally-efficient-approach-for-anonymizing-large-scale-physical-activity-data/256569

Privacy Preserving Data Mining: Taxonomy of Existing Techniques

Madhu V. Ahluwalia and Aryya Gangopadhyay (2008). *Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions* (pp. 70-93).

www.irma-international.org/chapter/privacy-preserving-data-mining/6862

IMMAESA: A Novel Evaluation Method of IDPSs' Reactions to Cyber-Attacks on ICSs Using Multi-Objectives Heuristic Algorithms

Mhamed Zineddine (2021). *International Journal of Information Security and Privacy* (pp. 65-98).

www.irma-international.org/article/immaesa/273592

Achieving Reconciliation Between Privacy Preservation and Auditability in Zero-Trust Cloud Storage Using Intel SGX

Liangshun Wu, Hengjin Cai and Han Li (2022). *International Journal of Information Security and Privacy* (pp. 1-20).

www.irma-international.org/article/achieving-reconciliation-between-privacy-preservation-and-auditability-in-zero-trust-cloud-storage-using-intel-sgx/284055