Scopus

Documents

Balla, A.[a] , Habaebi, M.H.[a] , Elsheikh, E.A.A.[b] , Islam, M.R.[a] , Suliman, F.M.[b]

The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems
(2023) *Sensors*, 23 (2), art. no. 758, . Cited 2 times.

DOI: 10.3390/s23020758

[a] Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, 53100, Malaysia
[b] Department of Electrical Engineering, College of Engineering, King Khalid University, Abha, 61421, Saudi Arabia

Abstract
Integrating IoT devices in SCADA systems has provided efficient and improved data collection and transmission technologies. This enhancement comes with significant security challenges, exposing traditionally isolated systems to the public internet. Effective and highly reliable security devices, such as intrusion detection system (IDSs) and intrusion prevention systems (IPS), are critical. Countless studies used deep learning algorithms to design an efficient IDS; however, the fundamental issue of imbalanced datasets was not fully addressed. In our research, we examined the impact of data imbalance on developing an effective SCADA-based IDS. To investigate the impact of various data balancing techniques, we chose two unbalanced datasets, the Morris power dataset, and CICIDS2017 dataset, including random sampling, one-sided selection (OSS), near-miss, SMOTE, and ADASYN. For binary classification, convolutional neural networks were coupled with long short-term memory (CNN-LSTM). The system's effectiveness was determined by the confusion matrix, which includes evaluation metrics, such as accuracy, precision, detection rate, and F1-score. Four experiments on the two datasets demonstrate the impact of the data imbalance. This research aims to help security researchers in understanding imbalanced datasets and their impact on DL SCADA-IDS. © 2023 by the authors.

Author Keywords
cyber security; ICS; IDS; imbalanced datasets; SCADA

Index Keywords
Computer crime, Cybersecurity, Integrated circuits, Intrusion detection, Network security, SCADA systems; Cyber security, Data collection, Data imbalance, Data-transmission, ICS, Imbalanced dataset, Intrusion Detection Systems, Performance, SCADA, Transmission technologies; Long short-term memory;
algorithm, benchmarking, information processing, Internet, long term memory; Algorithms, Benchmarking, Data Collection, Internet, Memory, Long-Term

References
- Huda, S., Yearwood, J., Hassan, M.M., Almogren, A.
  Securing the operations in SCADA-IoT platform based industrial control system using ensemble of deep belief networks
  (2018) *Appl. Soft Comput*, 71, pp. 66-77.

- Abdulrahman, A.A., Ibrahem, M.K.
  Toward Constructing a Balanced Intrusion Detection Dataset Based on CICIDS2017
  (2020) *Samarra J. Pure Appl. Sci*, 2, pp. 132-142.

- Wotawa, F., Muhlburger, H.
  On the Effects of Data Sampling for Deep Learning on Highly Imbalanced Data from SCADA Power Grid Substation Networks for Intrusion Detection
  Proceedings of the IEEE International Conference on Software Quality, Reliability and Security (QRS),
  Haikou, China, 6–10 December 2021

- Fundin, A.
  Generating Datasets Through the Introduction of an Attack Agent in a SCADA Testbed
  (2021) *Master's Thesis*,
  Linköping University, Linköping, Sweden

- Suaboot, J., Fahad, A., Tari, Z., Grundy, J., Mahmood, A.N., Almalawi, A., Zomaya, A., Drira, K.
  A Taxonomy of Supervised Learning for IDSs in SCADA Environments
  (2020) *ACM Comput. Surv*, 53, pp. 1-37.

- Kulkarni, A., Chong, D., Batarseh, F.A.
  Foundations of data imbalance and solutions for a data democracy
  (2020) *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*, pp. 83-106.
  Batarseh F.A., Yang R., (eds), 1st ed., Elsevier, Amsterdam, The Netherlands

- Peterson, J.M., Leevy, J.L., Khoshgoftaar, T.M.
  A Review and Analysis of the Bot-IoT Dataset
  Proceedings of the 15th IEEE International Conference on Service-Oriented System Engineering, SOSE 2021, pp. 20-27.
  Oxford, UK, 23–26 August 2021

-
  Available online

- Jiang, K., Wang, W., Wang, A., Wu, H.
  Network Intrusion Detection Combined Hybrid Sampling with Deep Hierarchical Network
  (2020) *IEEE Access*, 8, pp. 32464-32476.

- Miah, O., Khan, S., Shatabda, S., Dewan, M.F.
  Improving Detection Accuracy for Imbalanced Network Intrusion Classification using Cluster-based Under-sampling with Random Forests
  Proceedings of the International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT),
  Dhaka, Bangladesh, 3–5 May 2019

- Tsai, C.F., Lin, W.C., Hu, Y.H., Yao, G.T.
  Under-sampling class imbalanced datasets by combining clustering analysis and instance selection
  (2019) *Inf. Sci*, 477, pp. 47-54.

- Aziz, M.N., Ahmad, T.
  Clustering Under-Sampling Data for Improving the Performance of Intrusion Detection System
  (2021) *JESTEC*, 16, pp. 1342-1355.

- Wang, Z., Jiang, D., Huo, L., Yang, W.
  An efficient network intrusion detection approach based on deep learning
  (2021) *Wirel. Netw*, 5, pp. 1-14.

- Zhang, X., Ran, J., Mi, J.
  An Intrusion Detection System Based on Convolutional Neural Network for Imbalanced Network Traffic
  Proceedings of the International Conference on Computer Science and Network Technology (ICCSNT),
  Dalian, China, 19–20 October 2019

- Wu, K., Chen, Z., Li, W.
  A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks
  (2018) *IEEE Access*, 6, pp. 50850-50859.

- Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.
  (2019) *A Detailed Analysis of the CICIDS2017 Data Set*, 977, pp. 172-188.

Mobasheri A., (ed), Communications Computer and Information Science, Springer, Cham, Switzerland

Correspondence Address
Habaebi M.H.; Department of Electrical and Computer Engineering, Malaysia; email: habaebi@iium.edu.my