The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems

By: Balla, A (Balla, Asaad) [1] ; Habaebi, MH (Habaebi, Mohamed Hadi) [1] ; Elsheikh, EAA (Elsheikh, Elfatih A. A.) [2] ; Islam, MR (Islam, Md. Rafiqul) [1] ; Suliman, FM (Suliman, F. M.) [2]

View Web of Science ResearcherID and ORCID (provided by Clarivate)

SENSORS

Volume: 23 **Issue:** 2
Article Number: 758
DOI: 10.3390/s23020758
Published: JAN 2023
Indexed: 2023-03-03
Document Type: Article

Jump to

Enriched Cited References

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.

Keywords

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

Author Information

Corresponding Address: Habaebi, Mohamed Hadi (corresponding author)

Int Islamic Univ Malaysia, Dept Elect & Comp Engn, Kuala Lumpur 53100, Malaysia

Addresses:

1 Int Islamic Univ Malaysia, Dept Elect & Comp Engn, Kuala Lumpur 53100, Malaysia

2 King Khalid Univ, Coll Engn, Dept Elect Engn, Abha 61421, Saudi Arabia

E-mail Addresses: habaebi@iium.edu.my

Categories/ Classification

Research Areas: Chemistry; Engineering; Instruments & Instrumentation

Citation Topics: 4 Electrical Engineering, Electronics & Computer Science › 4.61 Artificial Intelligence & Machine Learning › 4.61.1302 Intrusion Detection

Web of Science Categories: Chemistry, Analytical; Engineering, Electrical & Electronic; Instruments & Instrumentation

+ See more data fields

Journal information

SENSORS

eISSN: 1424-8220

Current Publisher: MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND

Research Areas: Chemistry; Engineering; Instruments & Instrumentation

Web of Science Categories: Chemistry, Analytical; Engineering, Electrical & Electronic; Instruments & Instrumentation

3.847

Journal Impact Factor™ (2021)

0.9

Journal Citation Indicator™ (2021)

Citation Network

In Web of Science Core Collection

1
Citation

Create citation alert

1
Times Cited in All Databases

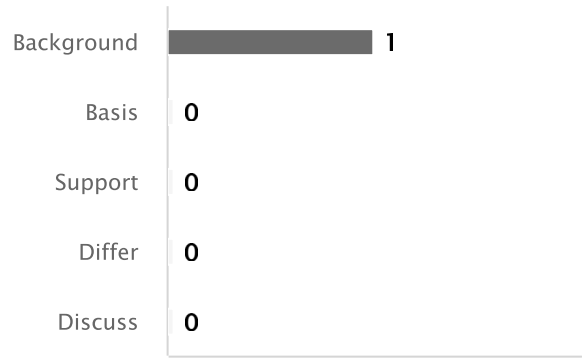
16
Cited References

View Related Records

+ See more times cited

Citing items by classification

Breakdown of how this article has been mentioned, based on available citation context data and snippets from 1 citing item(s).



Most Recently Cited by

Aljebreen, M; Alohali, MA; Abdelbagi, S; et al.
Binary Chimp Optimization Algorithm with ML Based Intrusion Detection for Secure IoT-Assisted Wireless Sensor Networks
SENSORS

Use in Web of Science

Web of Science Usage Count

4
Last 180 Days

4
Since 2013

Learn more

This record is from:

Web of Science Core Collection

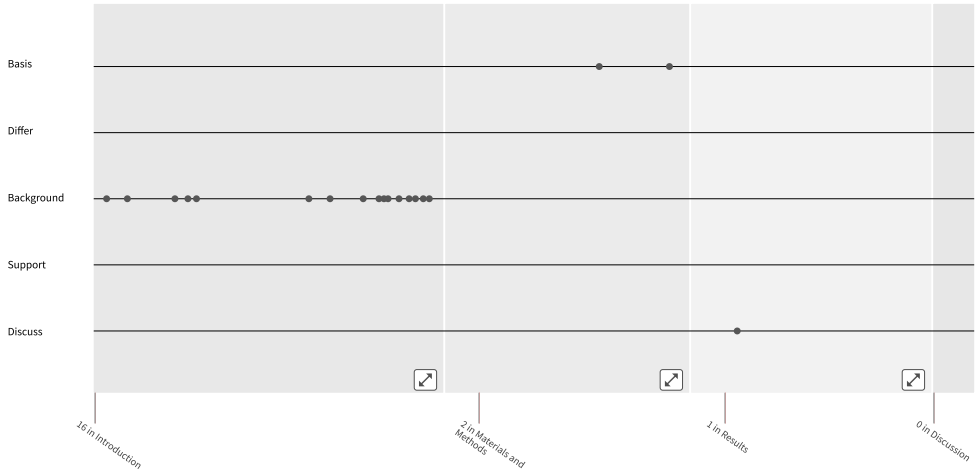
- Science Citation Index Expanded (SCI-EXPANDED)

Suggest a correction

If you would like to improve the quality of the data in this record, please Suggest a correction

16 Cited References

Explore



Showing 16 of 16 **View as set of results** First appearance

(from Web of Science Core Collection)

1

Securing the operations in SCADA-IoT platform based industrial control system using ensemble of deep belief networks

Huda, S; Yearwood, J; (...); Almogren, A
Oct 2018 | APPLIED SOFT COMPUTING 71 , pp.66-77

Full Text at Publisher •••

Cited in Article: 2

52
Citations

45
References

Related records

2

Toward Constructing a Balanced Intrusion Detection Dataset Based on CICIDS2017

Abdulrahman, A.A. and Ibrahem, M.K.
2020 | Samarra J. Pure Appl. Sci 2 , pp.132-142

Cited in Article: 1

6
Citations

0
References

3

On the Effects of Data Sampling for Deep Learning on Highly Imbalanced Data from SCADA Power Grid Substation Networks for Intrusion Detection

Wotawa, F and Muhlburger, H
21st IEEE International Conference on Software Quality, Reliability and Security (QRS)
2021 |
2021 IEEE 21ST INTERNATIONAL CONFERENCE ON SOFTWARE QUALITY, RELIABILITY AND SECURITY (QRS 2021)
, pp.864-872

Enriched Cited References

Full Text at Publisher •••

Cited in Article: 1

2
34 tions
References

Related records

4

Generating Datasets Through the Introduction of an Attack Agent in a SCADA Testbed

Fundin, A.
2021 | Master's Thesis
Linkoping University, Linkoping, Sweden

Cited in Article: 1

1
Citation

0
References

5

A Taxonomy of Supervised Learning for IDSs in SCADA Environments

Suaboot, J; Fahad, A; (...); Drira, K
Apr 2020 | ACM COMPUTING SURVEYS 53 (2)

Free Submitted Article From Repository

View full text

•••

Cited in Article: 1

24
Citations

140
References

Related records

6 Foundations of data imbalance and solutions for a data democracy

55 Citations

25 References

Kulkarni, A; Chong, D and Batarseh, FA
2020 |
DATA DEMOCRACY: AT THE NEXUS OF ARTIFICIAL INTELLIGENCE, SOFTWARE DEVELOPMENT, AND KNOWLEDGE ENGINEERING
, pp.83-106

Free Submitted Article From Repository

View full text

•••

Related records

Cited in Article: 1

7 A Review and Analysis of the Bot-IoT Dataset

5 Citations

73 References

Peterson, JM; Leevy, JL and Khoshgoftaar, TM
15th IEEE International Conference on Service-Oriented System Engineering (SOSE)
2021 |
2021 15TH IEEE INTERNATIONAL CONFERENCE ON SERVICE-ORIENTED SYSTEM ENGINEERING (SOSE 2021)
, pp.20-27

Full Text at Publisher •••

Related records

Cited in Article: 1

8 [Not available]

2 Citations

0 References

Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning
URL:
http://www.cse.wustl.edu/~jain/papers/ftp/imb_isi.pdf

Cited in Article: 2

9 Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network

Jiang, KY; Wang, WY; (...); Wu, HB
2020 | IEEE ACCESS 8 , pp.32464-32476

Free Full Text from Publisher •••

Cited in Article: 1

92 Citations
35 References
Related records

10 Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests

Miah, M. O.; Khan, S. S.; (...); Farid, D. M.
2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)
2019 | 2019 1 INT C ADV SCI , pp.1-5

Cited in Article: 1

11 Citations
0 References

11 Under-sampling class imbalanced datasets by combining clustering analysis and instance selection

Tsai, CF; Lin, WC; (...); Yao, GT
Mar 2019 | INFORMATION SCIENCES 477 , pp.47-54

Full Text at Publisher •••

Cited in Article: 1

128 Citations
36 References
Related records

12 CLUSTERING UNDER-SAMPLING DATA FOR IMPROVING THE PERFORMANCE OF INTRUSION DETECTION SYSTEM

Aziz, MN and Ahmad, T

1 Citation
36 References

Apr 2021 |
JOURNAL OF ENGINEERING SCIENCE AND TECHNOLOGY
16 (2) , pp.1342-1355

•••

Cited in Article: 1

Related records

13 Who starts the trade war? A theory of export controls and quid pro quo

9
Citations

47
References

Wang, Z and Zhou, YL
Oct 2021 | Jan 2021 (Early Access) |
WORLD ECONOMY 44 (10) , pp.2949-2964

View full text •••

Cited in Article: 2

Related records

14 An Intrusion Detection System Based on Convolutional Neural Network for Imbalanced Network Traffic

12
Citations

15
References

Zhang, XX; Ran, J and Mi, JZ
7th IEEE International Conference on Computer Science and Network Technology (ICCSNT)
2019 |
PROCEEDINGS OF 2019 IEEE 7TH INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND NETWORK TECHNOLOGY (ICCSNT 2019)
, pp.456-460

Enriched Cited References

Full Text at Publisher •••

Cited in Article: 1

Related records

15 A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks

134
Citations

47
References

Wu, KH; Chen, ZG and Li, W

2018 | IEEE ACCESS 6 , pp.50850-50859

Free Full Text from Publisher •••

Cited in Article: 1

Related records

16

A detailed analysis of the CICIDS2017 data set

Sharafaldin
2019 |
Communications in Computer and Information Science
977 , pp.172-188
Springer Verlag
URL: https://doi.org/10.1007/978-3-030-25109-3_9

20
Citations

0
References

Cited in Article: 1