Documents

Balla Babiker, A., Hadi Habaebi, M., Mubarak, S., Islam, Md.R.

A detailed analysis of public industrial control system datasets

(2023) *International Journal of Security and Networks*, 18 (4), pp. 245-263. Cited 1 time.

DOI: 10.1504/IJSN.2023.135511

Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, 53100, Malaysia

Abstract
A wide range of critical infrastructures such as power systems, water distribution systems, gas pipelines, and others are controlled and monitored using industrial control systems (ICSs). Recently, security attacks against ICSs are increasing at an alarming rate. These systems cannot afford to lose the availability of service; a cyber-attack can cause catastrophic damage. Intrusion detection systems (IDSs) are the first defence line against such attacks. To develop an effective IDS, a well-designed dataset is a must. In this paper, we present a detailed analysis of public intrusion datasets for ICSs. Focusing on the way security researchers used them to develop an IDS, their results, and the effect of the dataset's drawbacks. We performed exploratory data analysis (EDA), principal component analysis (PCA), and binary classification using random forest (RF) model. We believe this analysis will help the developers of the next generation of ICS-related IDSs.

Author Keywords
cyber security; datasets; ICS; IDS; industrial control system; information security; intrusion detection system

Index Keywords
Computer crime, Control systems, Electric power distribution, Intelligent control, Intrusion detection, Network security, Principal component analysis, Water distribution systems; Catastrophic damage, Cyber security, Cyber-attacks, Dataset, Industrial control systems, Intrusion Detection Systems, Power, Security attacks; Cybersecurity

Funding details
International Islamic University MalaysiaIIUM

We are grateful to the IIUM Tuition Fee Waiver program for sponsoring the tuition fees of Assad Balla.

References

- Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R.K., Ryan, M.
  Deep-IFS: intrusion detection approach for industrial internet of things traffic in fog environment
  (2021) *IEEE Transactions on Industrial Informatics*, 17 (11), pp. 7704-7715.

- Ajiboye, A.R., Abdullah-Arshah, R., Qin, H., Isah-Kebbe, H.
  Evaluating the effect of dataset size on predictive model using supervised learning technique
  (2015) *International Journal of Computer Systems & Software Engineering*, 1 (1), pp. 75-84.

- Alhowaide, A., Alsmadi, I., Tang, J.
  PCA, random-forest, and pearson correlation for dimensionality reduction in IoT IDS
  (2020) *IEMTRONICS 2020 - International IOT, Electronics and Mechatronics Conference, Proceedings*,
  Institute of Electrical and Electronics Engineers Inc

- Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A.B., Alzakari, N., Abou Elwafa, A., Kurdi, H.
  Impact of dataset size on classification performance: an empirical evaluation in the medical domain
  (2021) *Applied Sciences (Switzerland)*, 11 (2), pp. 1-18.

- Alyasiri, H., Clark, J.A., Malik, A., de Frein, R.
  Grammatical evolution for detecting cyberattacks in internet of things environments
  (2021) *Proceedings - International Conference on Computer Communications and Networks*,
  ICCCN, Institute of Electrical and Electronics Engineers Inc

- Asghar, M.R., Hu, Q., Zeadally, S.
  Cybersecurity in industrial control systems: issues, technologies, and challenges
  (2019) *Computer Networks*, 165.
  C

- Beaver, J.M., Borges-Hink, R.C., Buckner, M.A.
  An evaluation of machine learning methods to detect malicious SCADA communications
  (2013) *Proceedings – 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*, pp. 54-59.
  IEEE Computer Society

- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.
  SMOTE: synthetic minority over-sampling technique
  (2002) *Journal of Artificial Intelligence Research*, 16, pp. 321-357.

- Choi, S., Yun, J.H., Kim, S.K.
  A comparison of ICS datasets for security research based on attack paths
  (2019) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 154-166.
  Springer Verlag

- Conti, M., Donadel, D., Turrin, F.
  (2021) *A Survey on Industrial Control System Testbeds and Datasets for Security Research*,
  10 February [online]

- Desprez, M., Zawada, K., Ramp, D.
  Overcoming the ordinal imbalanced data problem by combining data processing and stacked generalizations
  (2022) *Machine Learning with Applications*, 7 (2022), p. 100241.

- Feng, C., Li, T., Chana, D.
  Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks
  (2017) *Proceedings - 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2017*, pp. 261-272.
  Institute of Electrical and Electronics Engineers Inc

- Ferrag, M.A., Maglaras, L., Ahmim, A., Derdour, M., Janicke, H.
  RDTIDS: rules and decision tree-based intrusion detection system for internet-of-things networks
  (2020) *Future Internet*, 12 (3).

- Green, B., Le, A., Antrobus, R., Roedig, U., Hutchison, D., Rashid, A.
  Pains, gains and PLCs: ten lessons from building an industrial control systems testbed for security research
  (2017) *Proceedings of the 10th USENIX Conference on Cyber Security Experimentation and Test, CSET'17*,
  August

- Hindy, H., Brosset, D., Bayne, E., Seeam, A., Bellekens, X.
  (2019) *Improving SIEM for Critical SCADA Water Infrastructures using Machine Learning*, Katsikas, S.K. et al. (Eds): Springer International Publishing, Cham

- Hink, C.R.B., Beaver, J.M., Buckner, M.A., Morris, T., Adhikari, U., Pan, S.
  Machine learning for power system disturbance and cyber-attack discrimination
  (2014) *7th International Symposium on Resilient Control Systems (ISRCS)*,

- Kenyon, A., Deka, L., Elizondo, D.
  Are public intrusion datasets fit for purpose characterising the state of the art in intrusion event datasets
  (2020) *Computers and Security*, p. 102022.

- Koroniotis, N., Moustafa, N., Sitnikova, E., Slay, J.
  Towards developing network forensic mechanism for botnet activities in the IoT based on machine learning techniques
  (2018) *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, pp. 30-44.
  LNICST, Springer Verlag

- Lai, Y., Zhang, J., Liu, Z.
  Industrial anomaly detection and attack classification method based on convolutional neural network
  (2019) *Security and Communication Networks*, pp. 1-11.

- Menze, T.
  (2020) *The State of Industrial Cybersecurity in the Era of Digitalization*,
  [online] (accessed 27 July 2023)

- Morris, T., Vaughn, R., Dandass, Y.S.
  A testbed for SCADA control system cybersecurity research and pedagogy
  (2011) *ACM International Conference Proceeding Series*,

- Moustafa, N., Slay, J.
  UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)
  (2015) *2015 Military Communications and Information Systems Conference, MilCIS 2015 – Proceedings*,
  Institute of Electrical and Electronics Engineers Inc

- Mubarak, S., Hadi Habaebi, M., Rafiqul Islam, M., Balla, A., Tahir, M.A.A., Elsheikh, E., Suliman, M.F.
  Industrial datasets with ICS testbed and attack detection using machine learning techniques
  (2022) *Intelligent Automation & Soft Computing*, 31 (3), pp. 1345-1360.
  [online] (accessed 27 July 2023)

- Pan, S., Morris, T., Adhikari, U.
  A specification-based intrusion detection framework for cyber-physical environment in electric power system
  (2015) *International Journal of Network Security (IJNS)*, 17 (2), pp. 174-188.
  March

- Pan, S., Morris, T., Adhikari, U.
  Classification of disturbances and cyber-attacks in power systems using heterogeneous time-synchronized data
  (2015) *IEEE Transactions on Industrial Informatics*, 11 (3), pp. 650-662.

- Pan, S., Morris, T., Adhikari, U.
  Developing a hybrid intrusion detection system using data mining for power systems
  (2015) *IEEE Transactions on Smart Grid*, 6 (6), pp. 3104-3113.

- Peterson, J.M., Leevy, J.L., Khoshgoftaar, T.M.
  A review and analysis of the Bot-IoT dataset

(2021) *Proceedings – 15th IEEE International Conference on Service-Oriented System Engineering, SOSE 2021*, pp. 20-27.
Institute of Electrical and Electronics Engineers Inc

- (2021) *Real-Time ICS SCADA System Cyber Kit Testbed with Industrial Hacking Scenarios – Mendeley Data*,
[online] (accessed 28 December 2021)

- Revathi, S., Malathi, A.
(2013) *A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection*,
[online] (accessed 27 July 2023)

- Rosa, L., Freitas, M., Mazo, S., Monteiro, E., Cruz, T., Simoes, P.
A comprehensive security analysis of a SCADA protocol: from OSINT to mitigation
(2019) *IEEE Access*, 7, pp. 42156-42168.

- Sharafaldin, I., Habibi Lashkari, A., Ghorbani, A.A.
A detailed analysis of the CICIDS2017 dataset
(2019) *Proceedings of the International Conference on Information Systems Security and Privacy*,
January

- Shiravi, A., Shiravi, H., Tavallaee, M., Ghorbani, A.A.
Toward developing a systematic approach to generate benchmark datasets for intrusion detection
(2012) *Computers and Security*, 31 (3), pp. 357-374.

- Soe, Y.N., Feng, Y., Santosa, P.I., Hartanto, R., Sakurai, K.
Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features
(2020) *Electronics (Switzerland)*, 9, p. 144.

- Stolfo, S.J., Fan, W., Lee, W., Prodromidis, A., Chan, P.K.
Cost-based modeling for fraud and intrusion detection: results from the JAM project
(2000) *Proceedings – DARPA Information Survivability Conference and Exposition, DISCEX 2000*, pp. 130-144.
Institute of Electrical and Electronics Engineers Inc

- Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., Hahn, A.
(2015) *Guide to Industrial Control Systems (ICS) Security*,
Gaithersburg, MD [online] (accessed 19 November 2020)

- Süzen, A.A.
Developing a multi-level intrusion detection system using hybrid-DBN
(2021) *Journal of Ambient Intelligence and Humanized Computing*, 12 (2), pp. 1913-1923.

- (2019) *The Bot-IoT Dataset | UNSW Research*,
[online] (accessed 22 December 2021)

- Morris, Tommy
(2013) *Industrial Control System (ICS) Cyber Attack Datasets*,
[online] (accessed 22 December 2021)

- Wang, W., Harrou, F., Bouyeddou, B., Senouci, S.M., Sun, Y.
A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems
(2021) *Cluster Computing*, 25 (1), pp. 561-578.

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

Publisher: Inderscience Publishers

ISSN: 17478405
Language of Original Document: English
Abbreviated Source Title: Int. J. Secur. Netw.
2-s2.0-85181041783
Document Type: Article
Publication Stage: Final
Source: Scopus