Why anomaly-based intrusion detection systems have not yet conquered the industrial market?

S. Seng[1,2], J. Garcia-Alfaro[2], and Y. Laarouchi[1]

[1] EDF R&D, Palaiseau, France
so.seng@free.fr,youssef.laarouchi@edf.fr
[2] Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
joaquin.garcia_alfaro@telecom-sudparis.eu

Abstract. In this position paper, we tackle the following question: why anomaly-based intrusion detection systems (IDS), despite providing excellent results and holding higher (potential) capabilities to detect unknown (zero-day) attacks, are still marginal in the industry, when compared to, e.g., signature-based IDS? We will try to answer this question by looking at the methods and criteria for comparing IDS as well as a specific problem with anomaly-based IDS. We will propose 3 new criteria for comparing IDS. Finally, we focus our discussion under the specific domain of IDS for critical Industrial control systems (ICS).

Keywords: Intrusion Detection System · Anomaly Detection · Explainable Artificial Intelligence · Industrial control system · Critical Infrastructures

1 Introduction

Faced with cybersecurity issues, the implementation of information systems monitoring tools is increasingly needed or a compulsory requirement. Many companies are investing in setting up a SOC (Security Operation Center), equipped with a SIEM (Security Information Management System) for the recognition and management of alerts. The origin of these alerts comes from various sensors, intrusion detection probes or external contextual informations.

There are two main categories of intrusion detection probes. The first category concerns Host-based IDS (HIDS). They use system data such as files or application event logs as input data. The second category concerns Network-based IDS (NIDS) which uses network exchanges as input data. In this paper, we do not distinguish between these two categories. In fact, whether we refer to either HIDS or NIDS, we focus our study on the underlying technology used by the detection engine. Two main representative technologies are often used in the literature: either signature-based or anomaly-based detection.

Signature-based detection, also referred to as *misuse* or *knowledge*-based detection, uses pattern matching classifiers to identify the attacks, i.e., they use

signature databases or heuristics describing the attacks. Early IDS products used this type of detection engines, since it is indeed simple, fast and does not consume much material resources. This type of detection is extremely effective at detecting attacks for which there is a signature, detection heuristic, or possibly an indicator of compromise (IoC). However, due to their operation, this type of detection is incapable of detecting unknown (zero-day) attacks. In addition, it requires a frequent updating of the signature database.

Anomaly-based detection aims at detecting attacks (also unknown ones) by modeling *normal* behaviors and, then, reporting any variations or anomalies deviating from a such model. This type of detection is not very recent. Indeed, the first one was proposed by Denning in 1987 [1]. However, in real life, information systems are often complex and difficult to model. Over the years, several methodologies have been proposed to model malicious behavior. The simplest methodologies are based on statistical methods such as threshold crossings. Today, most existing solutions seem to improve traditional detection rates by using artificial intelligence (AI) algorithms and, in particular, Machine Learning (ML) algorithms.

For nearly 20 years, the scientific literature on IDS has focused on anomaly-based detection engines, in particular on the use of AI algorithms. The majority of these studies on AI-based anomaly detection algorithms present detection rates (i.e., accuracy rates) greater than 95%, with very low false-negative rates, of the order of a few percent [2]. These very good results seem to show that AI algorithms are particularly efficient and suitable for IDS. However, currently on the market, commercial offers are mainly based on signature-based detection engines and ultimately only integrate little AI [3]. This low representativeness of commercial AI-based IDS solutions constitutes a paradox.

In this position paper, we tackle this paradox: why anomaly-based IDS have not yet conquered the industrial market? We will try to answer this question by looking at the methods and criteria for comparing IDS as well as a specific problem with anomaly-based IDS. We focus our discussion under the specific domain of critical Industrial control systems (ICS) and show that this question is particularly important in this context.

The paper is structured as follows. Section 2 provides the background and elaborates further on our problem domain. Section 3 provides our answer to the question. Section 4 discusses the link of our question to the specific domain of critical industrial control systems. Section 5 concludes the work.

2 Low Adoption of ML-based Detection in the Industry

As mentioned in the introduction, there is a vast literature and scientific studies on AI-based anomaly detection engines. Reports like [4] show that between 2000 and 2012, only a 3% of the scientific literature was concerned with signature-based solutions, while almost a 97% of the studies correspond to anomaly-based solutions, from which a high majority relied on AI methods, in particular, ML

methods. We have not found more recent statistics but we are confident that with the craze and the latest advances in AI, the ratio of scientific study has remained very high for AI-based anomaly detection engines. This section will give a quick overview of existing products, both for open source and commercial solutions. Then it will try to identify causes for the low adoption of AI in existing products. Finally, we will look at the evaluation criteria for IDS.

2.1 Omnipresence of Signature-based Detection Engines

OpenSource Products — Successful IDS products in the OpenSource community include NIDS products such as Snort [5], Zeek [6] (formerly called Bro) and Suricata [7]; and HIDS products such as ClamAV/ClamWin [8]. They all use signature-based detection engines. OpenSource IDS using anomaly-based detection engines are mainly at the level of prototypes, derived from research studies [9–14]. Only a few, like Zeek [6] are listed as anomaly-based IDS by some authors. Indeed, Zeek can be used as a development framework which can be easily extended to create new functionalities like anomaly detection. Hence, several research projects use this ability to extend Zeek for proof-of-concept development of anomaly-based algorithms[3]. However, we must note that Zeek shall be considered as a signature-based IDS, since this is its main default mode

Commercial Products — The number of commercial IDS products is considerably larger than OpenSource products [15]. A first observation that can be made on commercial IDS is that almost all of them integrate a signature-based detection engine. Indeed, such engines are generally very effective at detecting known attacks, consuming little material resources and very attractive from a corporate security standpoint.

On the contrary, very few commercial products come with an anomaly-based detection engine. At most, we can find in the market some hybrid designs, promising the two main types of detection. This may also suggest that anomaly-based detection engines are not yet self-sufficient, i.e., they are merely seen as a kind of complement to the more efficient signature-based designs. The inclusion of anomaly-based AI solutions in commercial products can also be seen as a commercial claim [16]. Most commercially available anomaly-based detection solutions are still insufficiently described to be able to assess their capabilities. It is then difficult to estimate whether this is an effective implementation or a cosmetic and marketing argument.

Commercial IDS do not generally use a single intrusion detection probe but a complete solution integrating several additional functionalities [15]. A detection engine can even be provided as a SaaS (Software as a Service), offering hybrid solutions combining multiple detection techniques. We regularly find hybrid solutions containing an intrusion detection probe incorporating a signature engine, coupled with an outsourced service performing an anomaly-based detection. This

[3] For instance, https://www.stratosphereips.org/zeek-anomaly-detector

is notably the case of most antivirus-type HIDS where signature-based detection is performed by the intrusion detection probe itself and the anomaly-based detection is an outsourced service called *CloudAV* [17].

2.2 Anomaly-based challenges for IDS

Some authors in the related literature justify the lack of anomaly-based IDS in the industry, compared to the number of existing studies in the scientific community, by the lack of rigor in such studies [2]. It can be summarized by the following issues: (1) lack of datasets, (2) weak evaluation methods, (3) reproducibility (e.g., lack of data initialization data, replicability of the datasets and hardware configuration), (4) comparability (e.g., different types of attacks needing to be compared separately).

The lack of rigor [18] and the importance of having datasets of quality [19] is in fact a classical issue for the evaluation of AI algorithms, and ML in particular. In the cybersecurity realm, moreover, confidentiality issues can also lead to difficulties to share high quality datasets [4, 20, 21]. This observation particularly affects the evaluation of NIDS products. According to [22], two very old datasets such as KDD99 and NSL-KDD represented in 2020 almost a 71% of the datasets used in scientific literature. Seen by most authors as outdated evaluation datasets, they correspond moreover to a single experiment carried out by DARPA between 1998 and 1999 [23], being the latter a cleaning and improvement of the former, in particular, in terms of data labeling [20]. More recent datasets exist [3], notably CIC-IDS 2017 and CIC-IDS2018 [24] and SWaT [25]. Still, their number remains generally modest and these are still too rarely used. For architectures not covered by KDD99 or by other public datasets, e.g., for industrial architectures, the absence of existing datasets encourages simulation or data generation, even if it means moving away from real constraints.

The aforementioned issues and, more specifically, the difficulties in finding appropriate evaluation datasets, are intrinsic issues in many other AI and ML research domains, such as medicine, where access to data must respect patient privacy. However, they may constitute a major obstacle to consolidate a commercial solution, especially in industrial domains related to critical ICS, in which the incorporation of novel cybersecurity approaches have a certain lack of acceptance.

2.3 Benchmarks and Evaluation Criteria

The expected rate of false positives and false negatives, as well as the processing performance, constitute important criteria to evaluate the quality of an IDS. The processing performance is often related to the number of events per second processed by the detection engine of an IDS. In particular, it is notably used to identify whether the IDS is capable of processing events in real time. The expected rate of false positives and false negatives is often defined as follows:

– False Positive Rate (FPR): $FPR = \frac{FP}{FP+TN}$, where FP is the observed number of false positive events, and TN the true number of negative events.

– False Negative Rate (FNR): $FNR = \frac{FN}{FN+TP}$, where FN is the observed number of false negative events, and TP the true number of positive events.

The two aforementioned indicators are generally used for the evaluation of any classifier used for detection. Receiver Operating Characteristic (ROC) curves are often used to represent binary classifiers based on their FPR and FNR rates [26]. Similarly, a confusion matrix, cf. table 1, can also be used to represent the efficiency of a classifier.

	Actual positives	Actual negatives
Positive Predictions	True positives (TP)	False positives (FP)
Negative Predictions	False negatives (FN)	True negatives (TN)

Table 1. Confusion matrix

In a cybersecurity and IDS context, the primary goal of a classifier is to minimize the number of false negatives (since undetected attacks lead to high risks [27]). This only goal can be a challenge because minimizing the number of false negatives usually involves to increase the number of false positives, which in turn increases the workload of human analysts.

Other criteria to quantify the efficiency of an IDS include [28,29]: (1) accuracy (directly derived from the FPR), (2) performance (i.e., processing capabilities), (3) completeness (i.e., ability to identify all existing attacks and therefore directly derived from the FNR), (4) fault tolerance (i.e., ability of the IDS to resist the attacks itself), and (5) timeliness (i.e., ability to propagate the information, e.g., when a mitigation action must be conducted right after a detection alert has been processed).

2.4 New evaluation Criteria

We think, the aforementioned explanations and evaluation criteria are insufficient to justify the low number of anomaly-based IDS deployed in the market. We propose to define two new concepts or criteria that will be interesting to explore (1) *completeness of knowledge* and (2) *ease of implementation and maintenance*.

Completeness of knowledge differs depending on the detection technique. On one hand, the use of knowledge completeness as a criterion related to a signature-based detection engine would refer to the quality and richness (in the absence of being able to be exhaustive) of the signature database. Since signature-based techniques base their detection on the existence of attack signatures (i.e., attack identification patterns), the higher the number of unique signatures associated to the IDS, the higher as well the completeness of knowledge associated to such an IDS. This criterion may also focus on related properties of the signature database of the IDS, such as the database update mechanism or the language flexibility to define new attacks. On the other hand, the use of knowledge completeness as a criterion related to anomaly-based detection engines rather refers to the

quality of the the training dataset, which is often very domain specific and hard to quantify. This criterion, Completeness of knowledge, is potentially difficult to quantify. A good approach is probably to build a index reference based on the benchmark of several existing solutions.

Ease of implementation and maintenance also depends on the specific detection technique. In fact, signature-based detection is generally agnostic to the use cases or systems they monitor. The general tendency consists in integrating as many attack signatures as possible in the signature database. Its setup and maintenance process is, hence, straightforward. On the contrary, anomaly-based detection is rather specific to use cases. The setup process requires a preliminary step needed to model the normal behavior of the events that will be monitored. The level of expertise required for maintenance and operational conditions (e.g., updates, business knowledge, definition of ML features and samples during the creation of both training and testing datasets, etc.) is definitively much higher than for signature-based detection approaches. This criterion is composed of several subjective elements and therefore difficult to quantify. It would be necessary to look in detail at each of the elements that compose it and identify applicable metrics.

All thoses aforementioned explanation and evaluation criteria lead to a possible explanation for the low adoption of AI and ML techniques in current IDS products. Next, we continue our discussion on the necessity of anomaly-based designs to provide a higher degree of explanability in their predictions, in order to conquer the market.

3 Explainability of IDS Predictions

Regardless of IDS, some machine learning algorithms operate as *black boxes* and offer little explanation of their classification decisions. This lack of explanation or justification of the decision can be a hindrance to confidence in the prediction, in the model and to transparency. This prevents the use of these technologies for certain use cases such as medicine or critical infrastructure. This difficulty in interpreting the predictions of a classifier using machine learning methods can also be a part of the answer to the lack of anomaly-based IDS.

Figure 1, extract from [30], represents an intuitive graph (i.e., not based on accurate values) of the different machine learning algorithms. In the opinion of the author, this figure makes consensus. It illustrates that Neural Network (NN) algorithms offering the best FPR and FNR rates are also those offering the least explanations, and vice versa. This difficulty is well known and has been the topic of a major research focus since 2016. Indeed, in 2016 DARPA launched the eXplainable Artificial Intelligence (XAI) program and funded $2 billion [31]. [32] identified at least 14 workshops or symposia dedicated to this thematic between 2014 and mid-2019. According to Gartner, in 2020, XAI research was among the top 25 trends for artificial intelligence in the *Hype* curve.

The XAI topic is complex and several questions arise:

– What to explain?

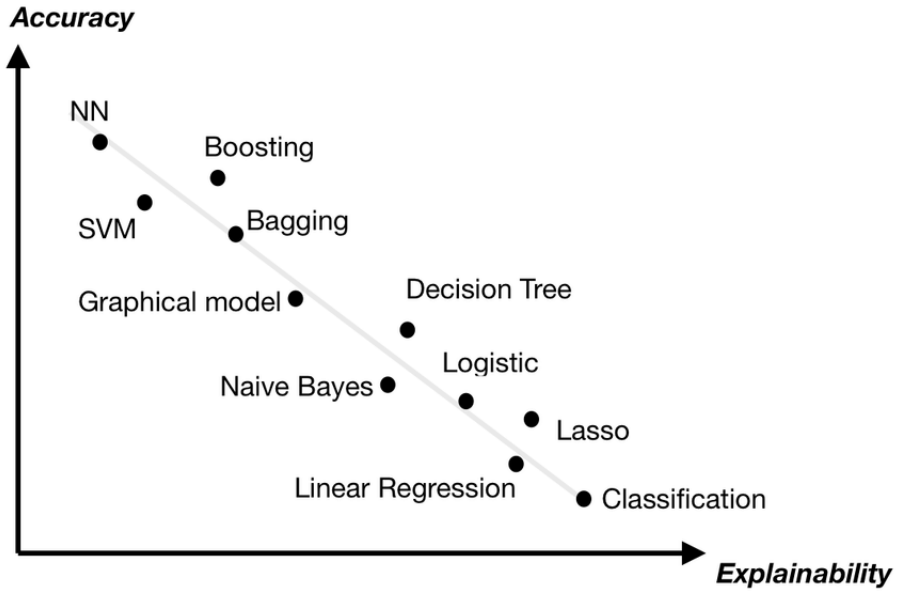

Fig. 1. Accuracy vs. Explainability of the main machine learning algorithms, extracted from [30]

- To whom should explanations be provided?
- How to provide these explanations?
- What explanations can be generated?

The answer to this last question is the one that raises the most scientific challenges. An obvious solution is to use classifiers that can provide explanations, such as decision trees, for use cases that require it. However, this limits the performances to a smaller number of classifiers and potentially the least accurate ones, as illustrated in Figure 1. To try to provide solutions, three main research approaches are studied:

1. Couple an accuracy algorithm with an explanation algorithm
2. Local Interpretation
3. Deep Explanation: Modify the model structure to extract intermediate metrics

Couple an accuracy algorithm with an explanation algorithm The first approach, 1) consists of keeping an existing classifier α, typically a DeepLearning (NN) classifier, and coupling it with a more explanatory classifier β. The latter then takes as input the same data as the classifier α, as well as its output prediction as shown in Figure 2. The β classifier then, having both the input data of the model and the prediction to be obtained, would allow to improve its prediction model and potentially to obtain some explanations. This solution has the advantage

of allowing the use of any α classifier and taking advantage of α's accuracy and β's explanatory capabilities. However, it is not trivial to guarantee that the explanation provided by the β classifier matches the prediction of the α classifier.

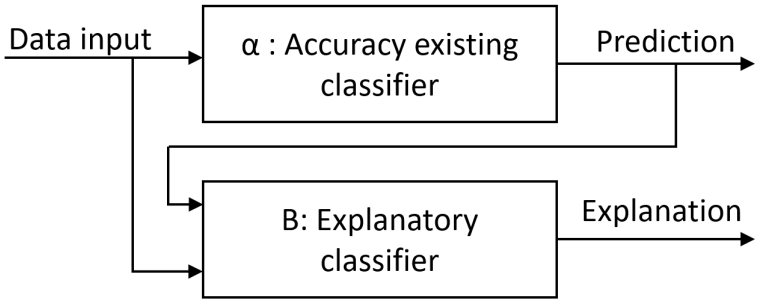

Fig. 2. Couple an accuracy algorithm with an explanation algorithm

Local Interpretation The research approach 2) also uses an already existing classifier and consists, for a given prediction, in slightly varying the input data in order to identify local threshold values from which the classifier modifies its prediction. This method allows to identify the input data that are important for the prediction and to group them into *clusters*. The interpretation that can be made of these clusters can then constitute a possible explanation for the prediction. This interpretation is however difficult to realize and even more difficult to generalize for all possible use cases.

This technique, named LIME (Local Interpretable Model-Agnostic Explanations) was first proposed by Ribeiro et al. in [33]. It seems to be the most studied approach and is particularly efficient for image classification and explanation. The interpretation of the clusters is then assigned to a human who can then evaluate the quality of the prediction.

Deep Explanation Finally, research approach 3) consists in improving existing algorithms or more globally classification models to allow the generation of explanations. An example of this approach is the DeepExplanation cited by Gunning in [31] and described in [34] which aims at extracting intermediate predictions whose semantic association allows the final prediction.

XAI and IDS Research on XAI is a recent topic, the most advanced work seems to be applied to photographic image processing and is most often based on the use of an explanation human-interface which then allows a human to validate or not the prediction. About twenty articles propose to apply the principles of XAI research to IDS. The majority of them uses the LIME method. The results of these studies seem promising but still insufficient. For example, the need to interpret the clusters concept of the LIME method seems indeed appropriate to

detect enumeration attacks such as DDOS or network scans, but seems hardly feasible for other types of attacks.

This explanation issue of AI-based classifiers in anomaly-based IDS does not really appear in signature-based IDS. Indeed, a signature intrinsically contains the detection criteria (the rules) and is often accompanied by descriptive elements such as the name of the associated attack or its references.

4 Discussion about ICS

4.1 Higher cybersecurity risks and impacts

Until recently, Industrial control system (ICS) was a separate and disjointed domain from traditional IT, with little or no communication between these two worlds. However, for cost reasons and complexity, ICS is increasingly adopting IT technologies, especially network communication that are now based on IP technologies. In addition, latest innovations and trends in ICS management and governance, such as Enterprise 4.0, strongly encourage the interconnection between IT and ICS. These two facts offer new opportunities for cybersecurity attacks on ICS.

We believe that ICS, which were until now globally spared, are less well prepared to face cybersecurity attacks. Indeed, some specificities of ICS offer a greater exposure to cyber-attacks. First, industrial equipment designers, industrial solution integrators and operators are still not very aware of cybersecurity, which is why there are rarely effective protection measures against cybersecurity risks. Secondly, ICS are often designed for a much longer lifespan than in IT. It is common to still find ICS in operation 20 to 30 years after their initial setup. However, cybersecurity evolves quickly and requires regular software and hardware updates. But the availability of ICS is often a more important criterion than for IT, the updates of ICS are often grouped during the planned maintenance operations. Thus, a critical vulnerability on a system can sometimes be fixed several months, or even years, after the publication of a patch. This is even more true for critical ICS where a hardware or software update can jeopardize safety qualifications. In these cases, operational safety has priority over cyber security, and operators are reluctant to perform updates. Finally, ICS and especially critical ICS, due to their interaction with the physical world, can have financial, environmental and even human impacts that are much more significant than in IT. All these elements imply that the need for monitoring ICS is probably more important than for IT.

4.2 Potentially effective network monitoring

On another level, some specificities about ICS seem favorable to monitoring solutions. Indeed, compared to IT systems, ICS do not evolve much. They have equipment, especially programmable logic controllers (PLC), that are deterministic in their operation. This provides industrial communication protocols with interesting properties for network monitoring [35]:

- *relatively* simple protocols;
- deterministic communication, based on iterative and continuous polling between, for example, a PLC and its sensors/actuators or between a supervisory console and its PLCs;
- strict timing requirement.

These properties make industrial communications easier and more efficient to monitor than IT communications which are often more complex, evolve rapidly and have a high variability due to human activities [36]. This facilitates the creation of anomaly detection models. However, the heterogeneity of industrial solutions, their low hardware resources and their closed (proprietary) aspects limit the possibilities for Host-based IDS.

4.3 Strong need of anomaly-based IDS for ICS

The two aforementioned points about ICS, comparing to IT, 1) risks and impacts of cybersecurity are potentially much higher and 2) anomaly-based monitoring solutions can be particularly effective, are complementary and make the use of anomaly-based IDS even more important. However, here again, there are several scientific works [3,10,35–51] but few anomaly-based IDS are deployed. The need to explore this paradox becomes even stronger in this context. The XAI issue of anomaly-based IDS may be a part of the problematic.

5 Conclusion

This position paper has addressed why, despite their excellent results and in particular their potential capacity to detect unknown attacks, the use of artificial intelligence (AI) anomaly-based detection in IDS products, e.g., machine learning (ML) approaches, still remain marginal in the cybersecurity industry — compared to other detection approaches, such as the use of signature-based detection.

We have started our discussions by reviewing some existing background and related literature, highlighting specific problems in other AI and ML domains, such as the difficulty of building up and maintaining quality datasets (both for training and operational processing), as well as issues with traditional criteria proposed for the evaluation of IDS. The use of extended criteria, such as *completeness of knowledge* and *ease of implementation and maintenance* led our discussion to claim the necessity of exploring a new criterion, the *explainability of IDS predictions* and positioned some of the necessary rationale to be included by next-generation anomaly-based detection engines, to tackle the problem.

To sum up, we have considered that usual IDS evaluation approaches such as false negative and false positive rates, complemented by additional performance criteria, are not enough for an IDS to adopt new anomaly-based products built upon AI and ML techniques. We think that novel criteria addressing the level of quality and explainability of the predictions derived from anomaly-based detection engines is a must. We have also discussed the importance of handling this

question under the specific domain of critical ICS. Indeed, those systems have increased monitoring needs and have properties that make them more favorable to anomaly detection.

For future work, it would be interesting to identify metrics to quantify the new criteria we have discussed in this paper: *completeness of knowledge*, *ease of implementation and maintenance* and especially *explainability*. Then to measure these metrics on various existing products and thus make a comparison of the existing solutions. Finally, it would be relevant to apply this approach in priority to critical ICS which are particularly adapted to anomaly-based IDS. For the latter case, it will also be necessary to overcome the issue of lack of data sets, which is more pronounced for industrial than for IT.

References

1. Dorothy Denning. An Intrusion Detection Model. In *Proceedings of the Seventh IEEE Symposium on Security and Privacy*, pages 119–131, 1986.
2. Mahbod Tavallaee, Natalia Stakhanova, and Ali Akbar Ghorbani. Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods. *Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods*, 40(5):516–524, 2010. Num Pages: 9 Place: New-York, NY Publisher: Institute of Electrical and Electronics Engineers.
3. Mauro Conti, Denis Donadel, and Federico Turrin. A Survey on Industrial Control System Testbeds and Datasets for Security Research. February 2021. arXiv: 2102.05631.
4. M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita. Network Anomaly Detection: Methods, Systems and Tools. *IEEE Communications Surveys Tutorials*, 16(1):303–336, 2014. Conference Name: IEEE Communications Surveys Tutorials.
5. Snort official web site. Snort - Network Intrusion Detection & Prevention System. https://www.snort.org/, 2021.
6. Zeek official web site. The Zeek Network Security Monitor. https://zeek.org/, 2021.
7. Suricata official web site. Suricata. https://suricata-ids.org/, 2021.
8. ClamavNet official web site. ClamavNet. https://www.clamav.net/, 2021.
9. John Hurley, Antonio Munoz, and Sakir Sezer. ITACA: Flexible, scalable network analysis. In *2012 IEEE International Conference on Communications (ICC)*, pages 1069–1073, June 2012. ISSN: 1938-1883.
10. Shengyi Pan, Thomas Morris, and Uttam Adhikari. A specification-based intrusion detection framework for cyber-physical environment in electric power system. *International Journal of Network Security*, 17:174–188, January 2015.
11. Hamid Bostani and Mansour Sheikhan. Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach. *Computer Communications*, 98:52–71, January 2017.
12. Abdelaziz Amara Korba, Mehdi Nafaa, and Salim Ghanemi. Hybrid Intrusion Detection Framework for Ad hoc networks. *International Journal of Information Security and Privacy*, 10(4):1–32, October 2016.
13. A. Lavin and S. Ahmad. Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 38–44, December 2015.

14. Jiankun Hu. Host-Based Anomaly Intrusion Detection. In Peter Stavroulakis and Mark Stamp, editors, *Handbook of Information and Communication Security*, pages 235–255. Springer, Berlin, Heidelberg, 2010.
15. Lawrences Orans, Jeremy D'Hoinne, and Josh Chessman. Gartner - Market Guide for Network Detection and Response. https://www.gartner.com/doc/reprints?id=1-1Z8C9OAX&ct=200612&st=sb, June 2020.
16. Garner-Hype. 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, 2020, September 2020.
17. wikipedia. Comparison of antivirus software. https://en.wikipedia.org/w/index.php?title=Comparison_of_antivirus_software&oldid=1003484641, January 2021. Page Version ID: 1003484641.
18. Jacques Wainer, Claudia G. Novoa Barsottini, Danilo Lacerda, and Leandro Rodrigues Magalhães de Marco. Empirical evaluation in Computer Science research published by ACM. *Information and Software Technology*, 51(6):1081–1085, June 2009.
19. Alessander Osorio, Marina Dias, and Gerson Geraldo H. Cavalheiro. Tangible Assets to Improve Research Quality: A Meta Analysis Case Study. In Calebe Bianchini, Carla Osthoff, Paulo Souza, and Renato Ferreira, editors, *High Performance Computing Systems*, Communications in Computer and Information Science, pages 117–132, Cham, 2020. Springer International Publishing.
20. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani. A detailed analysis of the KDD CUP 99 data set. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pages 1–6, July 2009. ISSN: 2329-6275.
21. Ali Shiravi, Hadi Shiravi, Mahbod Tavallaee, and Ali A. Ghorbani. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computers & Security*, 31(3):357–374, May 2012.
22. Arwa Aldweesh, Abdelouahid Derhab, and Ahmed Z. Emam. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, 189:105124, February 2020.
23. Darpa. KDD Cup 1999 Data, October 1999.
24. Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization:. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, pages 108–116, Funchal, Madeira, Portugal, 2018. SCITEPRESS - Science and Technology Publications.
25. Singapore University of Technology and Design. Secure Water Treatment. https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/, 2015.
26. Christopher D. Brown and Herbert T. Davis. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24–38, January 2006.
27. M. Szczepański, M. Choraś, M. Pawlicki, and R. Kozik. Achieving Explainability of Intrusion Detection System by Hybrid Oracle-Explainer Approach. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2020. ISSN: 2161-4407.
28. Hervé Debar, Marc Dacier, and Andreas Wespi. A revised taxonomy for intrusion-detection systems. *Annales Des Télécommunications*, 55(7):361–378, July 2000.
29. Ali A. Ghorbani, Wei Lu, and Mahbod Tavallaee. Evaluation Criteria. In Ali A. Ghorbani, Wei Lu, and Mahbod Tavallaee, editors, *Network Intrusion Detection and Prevention: Concepts and Techniques*, Advances in Information Security, pages 161–183. Springer US, Boston, MA, 2010.

30. Alexandre Duval. Explainable Artificial Intelligence (XAI). *MA4K9 Scholarly Report, Mathematics Institute, The University of Warwick*, April 2019.
31. David Gunning. Explainable Artificial Intelligence (XAI). *machine learning*, page 18, 2016.
32. Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, 8(8):832, August 2019. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.
33. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA, August 2016. Association for Computing Machinery.
34. Hui Cheng, Jingen Liu, Ishani Chakraborty, Guang Chen, Qiguang Liu, Mohamed Elhoseiny, Gary Gan, Ajay Divakaran, Harpreet S Sawhney, James Allan, John Foley, Mubarak Shah, Afshin Dehghan, Michael Witbrock, and Jon Curtis. Multimedia Event Detection and Recounting. page 12, 2014.
35. Robert Mitchell and Ing-Ray Chen. A survey of intrusion detection techniques for cyber-physical systems. *ACM Computing Surveys*, 46(4):55:1–55:29, March 2014.
36. S. Cheung, Bruno Dutertre, Martin Fong, Ulf Lindqvist, Keith Skinner, and Alfonso Valdes. Using Model-based Intrusion Detection for SCADA Networks. December 2006.
37. C. Yu, Y. Shen, L. Huang, H. Huang, X. Zhang, S. Jia, and J. Liu. The implementation of IEC60870-5-104 based on UML statechart and Qt state machine framework. In *2015 IEEE 5th International Conference on Electronics Information and Emergency Communication*, pages 392–397, May 2015.
38. C. S. Wickramasinghe, D. L. Marino, K. Amarasinghe, and M. Manic. Generalization of Deep Learning for Cyber-Physical System Security: A Survey. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, pages 745–751, October 2018. ISSN: 2577-1647.
39. Jürgen Beyerer, Alexander Maier, and Oliver Niggemann. *Machine Learning for Cyber Physical Systems: Selected papers from the International Conference ML4CPS 2020*. Springer Nature, January 2021. Google-Books-ID: r8kQEAAAQBAJ.
40. Igor Nai Fovino, Andrea Carcano, Marcelo Masera, and Alberto Trombetta. Design and Implementation of a Secure Modbus Protocol. In Charles Palmer and Sujeet Shenoi, editors, *Critical Infrastructure Protection III*, IFIP Advances in Information and Communication Technology, pages 83–96, Berlin, Heidelberg, 2009. Springer.
41. Fides Aarts, Harco Kuppens, Jan Tretmans, Frits Vaandrager, and Sicco Verwer. Improving active Mealy machine learning for protocol conformance testing. *Machine Learning*, 96(1-2):189–224, July 2014.
42. Hui Lin, Adam Slagell, Zbigniew Kalbarczyk, Peter W. Sauer, and Ravishankar K. Iyer. Semantic security analysis of SCADA networks to detect malicious control commands in power grids. In *Proceedings of the first ACM workshop on Smart energy grid security*, SEGS '13, pages 29–34, Berlin, Germany, November 2013. Association for Computing Machinery.
43. Dina Hadžiosmanović, Robin Sommer, Emmanuele Zambon, and Pieter H. Hartel. Through the eye of the PLC: semantic security monitoring for industrial processes. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC '14, pages 126–135, New Orleans, Louisiana, USA, December 2014. Association for Computing Machinery.

44. Rafael Ramos Regis Barbosa. Anomaly detection in SCADA systems: a network based approach. April 2014.
45. Marco Caselli, Emmanuele Zambon, and Frank Kargl. Sequence-aware Intrusion Detection in Industrial Control Systems. In *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*, CPSS '15, pages 13–24, Singapore, Republic of Singapore, April 2015. Association for Computing Machinery.
46. M. Kerkers. Assessing the Security of IEC 60870-5-104 Implementations using Automata Learning, May 2017. Library Catalog: essay.utwente.nl Publisher: University of Twente.
47. Robert Udd, Mikael Asplund, Simin Nadjm-Tehrani, Mehrdad Kazemtabrizi, and Mathias Ekstedt. Exploiting Bro for Intrusion Detection in a SCADA System. In *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security*, CPSS '16, pages 44–51, Xi'an, China, May 2016. Association for Computing Machinery.
48. Mohamad Kaouk, Jean-Marie Flaus, Marie-Laure Potet, and Roland Groz. A Review of Intrusion Detection Systems for Industrial Control Systems. In *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 1699–1704, April 2019. ISSN: 2576-3555.
49. Izhar Ahmed Khan, Dechang Pi, Pan Yue, Bentian Li, Zaheer Ullah Khan, Yasir Hussain, and Asif Nawaz. Efficient behaviour specification and bidirectional gated recurrent units-based intrusion detection method for industrial control systems. *Electronics Letters*, 56(1):27–30, October 2019. Publisher: IET Digital Library.
50. Habeeb Olufowobi, Clinton Young, Joseph Zambreno, and Gedare Bloom. SAIDuCANT: Specification-Based Automotive Intrusion Detection Using Controller Area Network (CAN) Timing. *IEEE Transactions on Vehicular Technology*, 69(2):1484–1494, February 2020. Conference Name: IEEE Transactions on Vehicular Technology.
51. Robert Mitchell and Ing-Ray Chen. Behavior-Rule Based Intrusion Detection Systems for Safety Critical Smart Grid Applications. *IEEE Transactions on Smart Grid*, 4(3):1254–1263, September 2013. Conference Name: IEEE Transactions on Smart Grid.