Metadata and metrics for cybersecurity dataset evaluation

Thaísa da Silva Hernandez[1 [0009−0000−8310−796X]],
Caroline Gandolfi[1 [0009−0006−1173−8600]],
Luís Olavo Bonino da Silva Santos[2 [0000−0002−1164−1351]],
Anderson Fernandes Pereira dos Santos[1,3 [0000−0002−6754−4809]], and
Maria Cláudia Cavalcanti[1 [0000−0003−4965−9941]]

[1] Military Institute of Engineering, Rio de Janeiro, RJ, Brazil
{thaisa.silva,caroline.gandolfi29,anderson,yoko}@ime.eb.br
https://www.ime.eb.mil.br/
[2] University of Twente, Twente, The Netherlands
l.o.boninodasilvasantos@utwente.nl http://www.utwente.nl
[3] Venturus, Campinas, SP, Brazil https://www.venturus.org.br/en

Abstract. The rapid increase in digital data generation from various technologies has expanded cyberspace, making it more susceptible to cyberthreats. Intrusion detection systems based on machine learning have been increasingly applied in recent years. However, the effectiveness of the models used in machine learning is strongly related to the quality of the datasets. This paper proposes a set of metadata and metrics to describe and evaluate cybersecurity datasets. These metrics are built on top of generic and specific properties of such datasets, and on their compliance with the FAIR principles. In addition, a FAIR Data Point (FDP) repository was configured to publish the metadata of such datasets in accordance with the FAIR principles, as well as the results of their evaluation based on the corresponding metrics. Finally, to demonstrate the usefulness of such metrics, the Athena evaluator tool was implemented, enabling the user to retrieve the metadata of datasets stored in the FDP repository, evaluate, and analyze them.

Keywords: cybersecurity datasets · dataset metadata · dataset evaluation · FAIR principles.

1 Introduction

In recent years, cyberattacks have caused extensive disruption and loss of information for online organizations, with high-profile incidents in the news [7]. Machine learning-based Intrusion Detection Systems (IDS) have been increasingly adopted, mainly due to their accuracy and lower need for human knowledge. A key focus of IDS based on machine learning techniques is to detect patterns based on cybersecurity datasets [8].

However, attacks evolve at a rhythm that cybersecurity datasets cannot match. According to Kenyon et al. [7], many datasets considered "gold standard" have become obsolete due to lack of maintenance. Furthermore, the lack of a central registry and inconsistent information on composition and provenance contribute to the lack of availability of cybersecurity datasets. In addition, according to Kenyon, there is no standard set of metrics with which to compare the quality of these datasets.

Data management techniques can help locate, integrate, and reuse datasets. In this context, the FAIR principles (Findable, Accessible, Reusable, and Interoperable) have been proposed [14]. They guide the publication of data for reuse and provide support for sharing cybersecurity datasets [14].

We present a set of quality metrics for cybersecurity datasets based on specific cybersecurity properties. In addition, we introduce a metric to evaluate relevance based on the usage of the dataset. Moreover, a FAIR Data Point repository[4] was configured to publish metadata about such datasets. Another contribution is the Athena Evaluator tool, which is able to retrieve that metadata, based on which it evaluates the datasets according to the user's ratings. Finally, it is worth highlighting that the main contribution of this work is the innovative combination of the metrics used, as well as their implementation and visualization.

This paper is structured as follows. Section 2 covers essential concepts used in this research. Section 3 discusses some related works. Section 4 presents the quality metrics. Section 5 describes the architecture for datasets' metadata publishing and evaluation, while Section 6 reports on some experiments and results. Finally, Section 7 summarizes the contributions and suggests future directions.

2 Background

Cybersecurity research and practice are becoming increasingly data-driven [16]. Cybersecurity datasets enable researchers to train, validate, and test proposed models for machine learning-based security solutions. Furthermore, they are essential for reproducing experiments and evaluating performance under comparable conditions using the same data [10].

In [14], the authors propose a set of principles that provide guidelines for publishing digital objects, such as datasets, code, workflows, and controlled vocabularies, in order to make them findable, accessible, interoperable, and reusable. The FAIR principles focus on the ability of machines to automate the use of data, as well as supporting the reuse of this data by humans. In addition, the FAIR principles can assist in comparisons between these digital resources and in assessing their quality or usefulness.

The FAIR Data Point metadata manager is an application server that exposes metadata on the Internet according to FAIR principles [2]. The metadata is represented in Resource Description Framework (RDF) to facilitate computational processing. FAIR Data Point has three main objectives: expose metadata about digital objects in a manner that follows FAIR principles, enable

[4] https://app.fairdatapoint.org/

consumers or users to discover this metadata, and provide this metadata in a machine-processable format.

3 Related Works

This section presents the main works related to this research. Gharib et al. [4] conducted a study of existing cybersecurity datasets between 1998 and 2016, and presented an evaluation framework for cybersecurity datasets with eleven proposed criteria: complete network configuration, complete traffic, labeled dataset, complete interaction, complete capture, available protocols, attack diversity, anonymity, heterogeneity, feature set, and metadata. These eleven criteria are evaluated according to a weight that can be defined based on the organization's request or the type of IDS system selected for the test.

Ring et al. [12] conducted a study focused on cybersecurity datasets and established a collection of fifteen properties compliant with the FAIR principles as a basis to evaluate cybersecurity datasets. These properties are categorized as follows: general information, nature of the data, data volume, recording environment, and evaluation. Although they have defined specific properties, the authors did not develop an evaluation score.

Table 1. Related Work Comparison

Metrics	**attack types cover.**	**labelling**	**relevance**	**network conf. completeness**	**metadata**	**FAIR coverage**
Gharib et al.[4]	X	X		X	X	
Ring et al. [12]	X	X		X	X	X
Mombelli et al. [11]					X	X
Flood et al. [3]	X	X				
Arp et al. [1]	X	X				
Athena Eval.	X	X	X	X	X	X

Flood et al. [3] analysed seven highly-cited benchmark datasets, discovering severe experimental bias. According to their results, those datasets have poor data diversity and murky labelling, weakening their utility as benchmarks. In another recent work, Arp et al. [1] identify common pitfalls in the design, implementation, and evaluation of learning-based security systems. Particularly, two of them are related to dataset evaluation: sampling bias and labelling inaccuracy.

Finally, Mombelli et al. [11] addresses the application of the FAIR Principles and metadata quality in the field of digital forensics. The paper evaluates metadata completeness and compliance with the FAIR Principles in 212 datasets from NIST's Computer Forensic Reference Dataset Portal (CFReDS). The results indicate deficiencies in metadata quality and the need for better data management standards. On the other hand, issues such as relevance, attack coverage, and labeling are not addressed in this work.

Table 1 presents a comparison of related works with respect to their coverage of properties and/or metrics to evaluate cybersecurity datasets. Most studies evaluate datasets based on the diversity of attacks they cover, as well as whether there is inaccuracy in labelling. However, the relevance of the dataset is not explored and, to our knowledge, only three of them include metadata as an important way to characterize and evaluate datasets. Thus, the present study (Athena Eval.) proposes a combination of properties/metrics not yet addressed by these studies. In addition, different from those works, it also configured and developed tools to support publishing and evaluating cybersecurity datasets based on their published metadata and on their compliance with the FAIR principles.

4 Evaluation Metrics

For this work, we consider three sets of metrics. The first two are defined over the properties raised by Ring et al. [12], one to generically evaluate cybersecurity datasets (subsection 4.1), and the other to focus specific properties (subsection 4.2), such as network traffic datasets. The third set aims to evaluate the dataset compliance with the FAIR principles (subsection 4.3).

4.1 Generic cybersecurity dataset metrics

The set of selected properties defined by [12] is described as follows. To each property corresponds a metric that can be applied to evaluate a dataset according to the formulas shown in Table 2. Note that, in addition to the metrics inspired by [12], we introduce a metric to evaluate the relevance of a dataset.

Timeliness: cybersecurity datasets need to be described in terms of the time point at which the cybersecurity data they include was created. Note that it is not the date of the dataset creation but the date of its underlying data. The reason for this is that technical tools evolve rapidly, and data captured in the past may not reflect the current reality. For example, the LL MIT DDoS 1.0[5] is a network traffic dataset that includes interesting attack types and situations, but does not include streaming data (e.g. Netflix), because at the time point of the capture of its data (year 2000), the streaming data service was not yet in use. Thus, the age of the data captured in a cybersecurity dataset plays an important role in the evaluation of that dataset.

To evaluate the *Timeliness* of a dataset, we chose the fuzzy logic method [9] for its wide use and simplicity. To apply this method, we assume that the variable x represents the year of the data in the dataset, and the time interval that this variable can assume is [1998 - 2025]. Then, we define $\widetilde{A}$ as a set of terms, where each element corresponds to a fuzzy set, such that $\widetilde{A} = \{\text{Old, Medium, Recent}\}$, where Old $= \{x \mid 1998 \leq x \leq 2010\}$, Medium $= \{x \mid 2005 \leq x \leq 2020\}$,

[5] http://www.ll.mit.edu/r-d/datasets/

Recent $= \{x \mid 2015 \leq x \leq 2025\}$, and $\forall\, x,\, x \in \mathbb{N}$. Then, a function μ can be applied to each fuzzy set $\tilde{a} \in \widetilde{A}$, such that $\mu_{\tilde{a}}(x) : \mathbb{N} \rightarrow [0, 1]$. This function uses a triangular membership function [9] to calculate the degree of pertinence of x to each fuzzy set $\tilde{a}$. Finally, the *Timeliness* metric will be evaluated based on the highest pertinence degree, as described in Table 2. For example, if the year is 2007, its pertinence degrees for each fuzzy set are: $\mu_{Old}(2007) = 0.25$, $\mu_{Medium}(2007) = 0.16$, and $\mu_{Recent}(2007) = 0$. Thus, the highest pertinence degree for the year 2007 is to the "Old" set ($\tilde{a} = old$), which implies that the *Timeliness* of the year 2007 is 0 ($Y(2007) = 0$), according to Table 2.

Public availability: cybersecurity datasets should be publicly available to serve as a basis for comparing different intrusion detection methods. Furthermore, the quality of datasets can only be verified by third parties if access is unrestricted [12]. However, there may be datasets that are not available for some reason, and some others that may only be available upon request. Thus, this metric evaluates a dataset according to these three situations, assigning 1, 0, and 0.5, respectively.

Table 2. Cybersecurity dataset evaluation metrics

Metric	**Formula**
Timeliness	$Y(x) = \begin{cases} 0, & \mu_{Old}(x) = max(M) \\ \frac{1}{2}, & \mu_{Medium}(x) = max(M) \\ 1, & \mu_{Recent}(x) = max(M) \end{cases}$ where $M = \{\mu_{\tilde{a}}(x), \forall\, \tilde{a} \in \widetilde{A}\}$
Public Availability	$P(b) = \begin{cases} 0, & b = no \\ \frac{1}{2}, & b = upon\, request \\ 1, & b = yes \end{cases}$
Anonymization	$A(t) = \begin{cases} 1, & (t = no \wedge k \in \{E, S\}) \vee (t = yes \wedge k = R) \\ 0, & (t = no \wedge k = R) \vee (t = yes \wedge k \in \{E, S\}) \\ 0, & t = not\, specified \end{cases}$
Labeling	$L(j) = \begin{cases} 0, & j = no \\ \frac{1}{2}, & j = partially\ labeled \\ 1, & j = completely\ labeled \end{cases}$
Relevance	$R(r, Y(x)) = w \times r + (1 - w) \times Y(x)$, where $w \in [0, 1]$, $r = \frac{\mathrm{n}}{\mathrm{m}}$, $n =$ no. dataset's citations, $m =$ no. citations of the most cited dataset.

Anonymization: privacy is compromised when the payload is not encrypted in a dataset with real traffic. So, most datasets have their payloads removed or anonymized, which may decrease the usefulness of the dataset but maintains the privacy of the information [4]. Datasets with synthetic or emulated traffic do not suffer from this issue and can be available. Therefore, the evaluation of this metric is directly related to the type of traffic in the dataset, which can have three values: Real (R), Emulated (E), and Synthetic (S). If a dataset has a real traffic type ($k = R$), it means that the data needs to be anonymized; otherwise, if the traffic type is emulated ($k = E$) or synthetic ($k = S$), it makes no sense to anonymize the data.

Labeling: this metric evaluates if the dataset is labeled, partially labeled, or unlabeled, assigning a higher score to datasets that provide labels.

Relevance: the relevance is evaluated based on a weighted combination of the citedness ratio of the dataset and its *timeliness*. Depending on the user's preference, a dataset may be more relevant if it is widely used and cited. On the other hand, it may be considered more relevant if it is more up-to-date. Thus, a weight (w) constant must be previously defined by the user. In extreme cases, if $w = 1$, only the citedness ratio is considered; conversely, if $w = 0$, the *timeliness* is the only factor considered. The relevance formula shown in Table 2 is used to calculate this metric for a given dataset, where n is the number of citations of this dataset, m is the number of citations of the most cited dataset in its area (related datasets), and $Y(x)$ is the value of the score of the *timeliness* metric for the dataset under evaluation.

4.2 Network Traffic dataset metrics

Besides the generic cybersecurity dataset metrics, for each subtype of cybersecurity dataset, a different set of metrics may be defined. In the case of network traffic datasets, we defined three metrics that are described as follows and summarized in Table 3.

Table 3. Specific cybersecurity dataset evaluation metrics

Metric	Formula
Attack Diversity	$D = \sum_{i=1}^{n} \frac{stride(i)}{n}$, where $n = 6$ and $stride(i) = \begin{cases} 1, & \text{if covered} \\ 0, & \text{if not covered} \end{cases}$
Normal/Attack Traffic Coverage	$N(c) = \begin{cases} 0, & c = no \\ 1, & c = yes \end{cases}$
Network Configuration Completeness	$C(g) = \begin{cases} 0, & g = no \\ 1, & g = yes \end{cases}$

Attack Diversity: In recent years, attack types have been changing and updating daily [4]. Therefore, having the ability to test and analyze Intrusion Detection Systems through a broad coverage of attacks and threat scenarios is one of the most important requirements that a network intrusion detection dataset must support [4; 1]. Given the large number and variations of attacks, to evaluate the attack diversity criterion, attack types were grouped according to the six threat categories of the *Microsoft* STRIDE model [13]: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of

Privilege. The more groups a dataset covers, the higher its score. For instance, if a dataset covers only *Spoofing* and *Tampering* attack types, it scores 2 in 6, which means $D = 0.33$.

Normal and Attack traffic Coverage: cybersecurity datasets should include a broad and accurate representation of normal and attack traffic. This metric indicates whether the dataset contains ($c = yes$) or not ($c = no$) background traffic. Background traffic refers to the traffic that was captured/generated but is not related to the attack.

Network Configuration Completeness: A dataset must include a complete network configuration [4], with several hosts, routers, and other network components, to capture the real effects of attacks [12], and represent a real-world situation. Thus, this metric indicates whether the dataset was created considering a complete network configuration ($g = yes$) or not ($g = no$).

4.3 FAIR Metrics

The compliance with the FAIR principles can also be measured. Each principle is broken into sub-principles, and each of them is associated with a specific metric. In this work, we assumed a subset of the metrics proposed in [15].

Findable: The F1 metric is used to verify the existence of globally unique and persistent identifiers in a dataset. In the FAIR context, globally unique refers to identifiers that are unambiguously linked to a resource, and persistence means that this globally unique identifier is never reused and identifies the resource, even when that same resource is no longer available [6]. In addition, the F2 metric is used to verify the richness of the metadata description. Here, the better the resource is described, both for humans and computers, the easier it will be to find. In addition, the F3 metric indicates whether the metadata explicitly includes the identifier of the data it describes.

Accessible: One of the main objectives of identifying a digital resource is the accessibility of that resource in a given format, using an explicitly established mechanism [6]. Here, a set of metrics is used to verify the level of data accessibility, including authentication/authorization protocols, if necessary (Metrics A1.1 and A1, respectively). In addition, metric A2 is used to verify the accessibility of metadata, even when the data to which this metadata refers is no longer accessible. For FAIR accessibility, it is very important that there is at least access to quality metadata that minimally describes its context and provenance, even when the relevant data is no longer available. According to Jacobsen et al, "There is a continuous focus on keeping relevant digital resources available in the future." [6].

Interoperability: Reaching a "common understanding" of digital resources through a globally understood "language" for machines is the focus of principle I1. To evaluate this principle, we use metrics I1 and I2 to verify the use of a knowledge representation language, vocabularies, and ontologies. In addition, "references to other resources are included to verify that the knowledge representing a resource is linked to that of other resources to create a meaningfully interconnected network of data and services" (metric I3) [6].

Reusable: Digital resources and their metadata must include a license describing the conditions under which the resource may be used, even if that use is "unconditional." Here, metrics are used to verify the presence of an explicit and accessible license (metric R1.1) and a detailed description of the provenance of the dataset (metric R1.2).

5 Athena Evaluator: Architecture and Implementation

The Athena Evaluator was developed to implement the metrics described in the previous section. It takes part as a module in the Athena approach [5], whose architecture modules are shown in Figure 1. The main idea is to allow publishers to describe their datasets with rich metadata, based on which consumers may find, evaluate, select, and reuse them. First, the **Metadata Manager** module works together with the **Schema Generator** to provide an easy and flexible way to store dataset metadata in a **repository**. Based on that metadata, the **Evaluator** module comes into play and applies those metrics. Finally, the enriched metadata, with the evaluation results, becomes publicly available in the repository for reuse. Then, the **evaluator** and **analyzer** modules provide the **consumer** with an interface where he can browse, choose a dataset, and configure (weight) the evaluation metrics, reaching a final score for that dataset. Given a list of data sets of interest, the consumer can compare the datasets according to predefined criteria, making it easier to choose the best data set to use.

The technologies used for the Athena architecture implementation are also indicated in Figure 1 (paralelograms and dashed squares). The FAIR Data Point (FDP) metadata[6] manager was chosen not only for its comprehensive functionality on publishing dataset metadata, but specially for its metadata extension flexibility, allowing for the description of specific metadata according to the publishers' requirements. The Form metadAta Schema ediTor (FAST) complements the publishing functionality. It is a schema generator that facilitates the extension of the original schema provided by the FDP manager. GraphDB[7] was used to implement the metadata repository, storing the metadata captured by FDP in RDF triples, also enabling SPARQL queries.

The Athena Evaluator is a web application designed to support the process of evaluating and analyzing cybersecurity datasets, featuring a client-server architecture. The client-side was developed using the Typescript programming language and the Next.js framework, which is built on top of the React JavaScript library, enabling greater efficiency in rendering, navigation, and organization of the application's source code.

For the server-side, Python was used, responsible for data processing and business logic. When the Publisher submits the published metadata for evaluation, this component communicates directly with the FDP API through GET requests. It retrieves the dataset metadata, and, at this point, these metadata undergo two types of evaluation. First, it extracts the specific metadata required

[6] https://www.fairdatapoint.org/

[7] https://graphdb.ontotext.com/

for the FAIRness evaluation. To do this, it interacts with the FAIR Metrics API[8], which provides the means for automatic assessment of compliance with the FAIR principles. Next, the dataset will be evaluated based on a set of metrics defined according to specific cybersecurity properties, such as timeliness and relevance. Then, the dataset evaluation results are stored in GraphDB for further analysis. Details about user accounts and their respective roles are stored in a MongoDB[9] database.

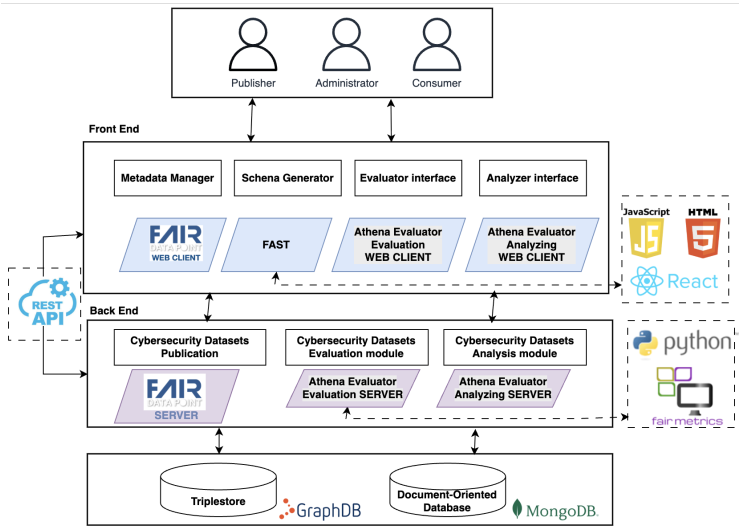

Publisher
Administrator
Consumer
Front End
Metadata Manager
Schena Generator
Evaluator interface
Analyzer interface
FAIR DATA POINT WEB CLIENT
FAST
Athena Evaluator Evaluation WEB CLIENT
Athena Evaluator Analyzing WEB CLIENT
JavaScript
HTML
React
REST API
Back End
Cybersecurity Datasets Publication
Cybersecurity Datasets Evaluation module
Cybersecurity Datasets Analysis module
FAIR DATA POINT SERVER
Athena Evaluator Evaluation SERVER
Athena Evaluator Analyzing SERVER
python
fairmetrics
Triplestore
GraphDB
Document-Oriented Database
MongoDB.

Fig. 1. ATHENA architecture modules and their corresponding implementations.

6 Experiments and Results

For the experiments, a FAIR Data Point repository[10] was configured to disseminate the results and expose information about such datasets. This information has been structured in a specific metadata schema to describe these datasets based on their properties and metrics discussed in Section 4.

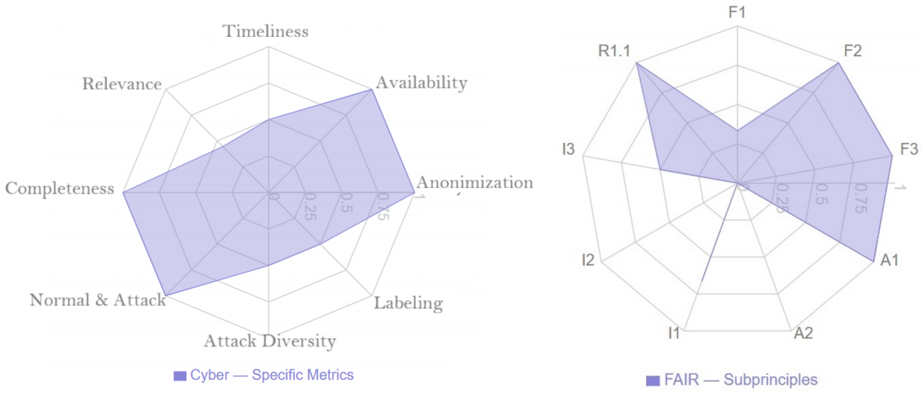
To illustrate the use of the Athena Evaluator, we selected two of the most prominent cybersecurity datasets [10]: ISCX-2012 and CSE-CIC-IDS2018. It is possible to verify that both datasets perform well in relation to general and specific cybersecurity criteria. ISCX-2012 performs slightly worse in terms of attack diversity and labeling, as it is only partially labeled, meaning it does

[8] https://github.com/FAIRMetrics/Metrics

[9] https://www.mongodb.com/

[10] https://app.fairdatapoint.org/

not specify the types of attacks in detail, only whether the traffic is malicious or benign. On the other hand, both datasets are considered average in terms of Timeliness and Relevance metrics. However, upon closer examination of the Relevance metric, we observe that the CSE-CIC-IDS22018 dataset outperforms the ISCX-2012. Furthermore, in addition to being more relevant, the former is more recent than the latter, contradicting the idea that older datasets tend to be more relevant due to the number of citations. This situation demonstrates that newer datasets can quickly become relevant and surpass older ones.



(a) (b)

Fig. 2. ISCXIDS2012 dataset

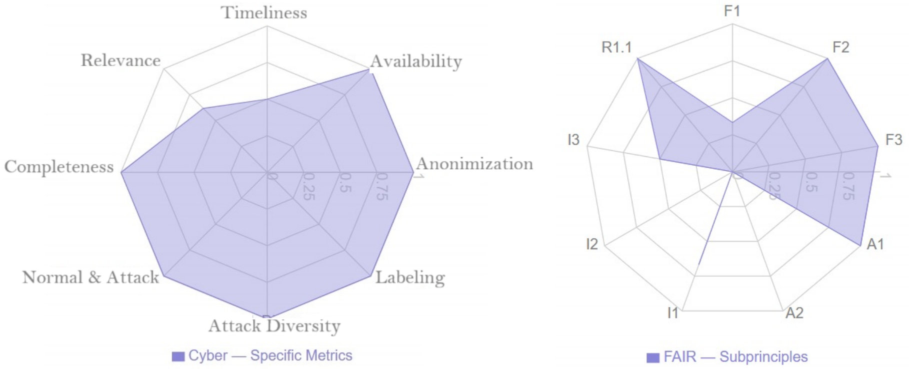
Now looking at it from the perspective of the FAIR Principles, it is possible to observe that the publication of datasets in the FAIR Data Point contributes mainly to the principles related to metadata richness. In addition, metadata records are available in RDF format, and widely used vocabularies such as DCAT[11], Dublin Core Terms[12], and Unified Cyber Ontology (UCO)[13] for specific cybersecurity metadata, contributing to the principle of interoperability.

These features improve the FAIRness of the datasets and expand the criteria for evaluating their quality and usefulness [14]. The aim is not to score on all the principles, but to encourage the community to provide more accessible, interoperable, and reusable datasets to advance cybersecurity research.

7 Conclusion

This paper presents a set of metadata and quality metrics for cybersecurity datasets based on their properties. In particular, it includes a metric to evaluate

[11] https://www.w3.org/TR/vocab-dcat-3/
[12] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[13] https://unifiedcyberontology.org/



(a) (b)

Fig. 3. CSE-CIC-IDS2018 dataset

relevance based on the usage of the dataset, increasing the evaluation criteria. In addition, it also includes metrics to evaluate the compliance of the dataset to the FAIR principles. This combination of metrics and their implementation brings a relevant contribution. Moreover, the implementation choices, such as the use of the FDP repository and the Athena evaluator tool, provide a flexible and harmonious architecture and interface that support the dataset publishers and consumers in their tasks. Furthermore, a special metadata schema for cybersecurity and quality metrics evaluations was created, allowing researchers to publish metadata about their datasets according to the FAIR principles, thereby expanding the criteria for comparing and evaluating dataset quality and usefulness. Next steps in this research, we aim to expand the set of properties used to describe cybersecurity datasets. Athena Evaluator is available on GitHub[14] for reproduction of the results presented here.

One limitation of this approach is that the publisher must provide all the metadata for the dataset being published, which are not always available or known. Future work may seek a more flexible and proactive approach, allowing some metadata to be disregarded or including guidance for publishers on how to improve datasets' metadata.

Acknowledgements

This research has been partially funded by FINEP/DCT/FAPEB (ref.: 2904/20 contract no 01.20.0272.00) under the "System of Systems of Command and Control" project.

[14] https://github.com/comp-ime-eb-br/S2C2-IME/tree/main/deliverables/AthenaEvaluator

Bibliography

[1] D. Arp, E. Quiring, F. Pendlebury, et al. Dos and don'ts of machine learning in computer security. In *Proc. of USENIX Security Symposium*, 2022.

[2] L. O. B. da Silva Santos, K. Burger, R. Kaliyaperumal, and M. D. Wilkinson. Fair data point: a fair-oriented approach for metadata publication. *Data Intelligence*, 5(1):163–183, 2023.

[3] R. Flood, G. Engelen, D. Aspinall, and L. Desmet. Bad design smells in benchmark nids datasets. In *2024 IEEE 9th European Symposium on Security and Privacy (EuroSP)*, pages 658–675, 2024.

[4] A. Gharib, I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani. An evaluation framework for intrusion detection dataset. In *2016 Int. Conf. on Information Science and Security (ICISS)*, pages 1–6. IEEE, 2016.

[5] T. Hernandez, C. Gandolfi, P. Bulcão, L. O. B. Santos, A. Santos, and M. C. Cavalcanti. Athena: A fair approach to publish and evaluate cybersecurity datasets. In *Proc. 18th Sem. on Ontology Research in Brazil*, 2025.

[6] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, et al. Fair principles: interpretations and implementation considerations, 2020.

[7] A. Kenyon, L. Deka, and D. Elizondo. Are public intrusion datasets fit for purpose? characterising the state of the art in intrusion event datasets. *Computers & Security*, 99:102022, 2020.

[8] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman. Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 2(1):1–22, 2019.

[9] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic*. Prentice Hall, 1995. vol.4.

[10] M. Macas, C. Wu, and W. Fuertes. A survey on deep learning for cybersec.: Progress, challenges, and opportunities. *Comp. Netw.*, 212:109032, 2022.

[11] S. Mombelli, J. Lyle, and F. Breitinger. Fairness in digital forensics datasets' metadata–and how to improve it. *Forensic Science Intern.*, 48:301681, 2024.

[12] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho. A survey of network-based intrusion detection data sets. *Computers & Security*, 86: 147–167, 2019.

[13] R. Scandariato, K. Wuyts, and W. Joosen. A descriptive study of microsoft's threat modeling technique. *Requirements Engineering*, 20(2):163–180, 2015.

[14] M. Wilkinson, M. Dumontier, I. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.

[15] M. D. Wilkinson, S.-A. Sansone, E. Schultes, P. Doorn, L. O. B. Santos, and M. Dumontier. A design framework and exemplar metrics for fairness. *Scientific Data*, 5(180118), 2018.

[16] M. Zheng, H. Robbins, Z. Chai, P. Thapa, and T. Moore. Cybersecurity research datasets: Taxonomy and empirical analysis. In *11th USENIX Workshop on Cyber Security Experimentation and Test*, 2018.