applied sciences

MDPI

Article

NTARC: A Data Model for the Systematic Review of Network Traffic Analysis Research

Félix Iglesias *, Daniel C. Ferreira and Gernot Vormayr, Maximilian Bachl and Tanja Zseby

Communication Networks Group, Institute of Telecommunications, TU Wien, 1040 Vienna, Austria; daniel.ferreira.1@gmail.com (D.C.F.); gernot.vormayr@nt.tuwien.ac.at (G.V.); maximilian.bachl@tuwien.ac.at (M.B.); tanja.zseby@tuwien.ac.at (T.Z.)

* Correspondence: felix.iglesias@tuwien.ac.at

Received: 28 April 2020; Accepted: 19 June 2020; Published: 23 June 2020

Abstract: The increased interest in secure and reliable communications has turned the analysis of network traffic data into a predominant topic. A high number of research papers propose methods to classify traffic, detect anomalies, or identify attacks. Although the goals and methodologies are commonly similar, we lack initiatives to categorize the data, methods, and findings systematically. In this paper, we present Network Traffic Analysis Research Curation (NTARC), a data model to store key information about network traffic analysis research. We additionally use NTARC to perform a critical review of the field of research conducted in the last two decades. The collection of descriptive research summaries enables the easy retrieval of relevant information and a better reuse of past studies by the application of quantitative analysis. Among others benefits, it enables the critical review of methodologies, the detection of common flaws, the obtaining of baselines, and the consolidation of best practices. Furthermore, it provides a basis to achieve reproducibility, a key requirement that has long been undervalued in the area of traffic analysis. Thus, besides reading hard copies of papers, with NTARC, researchers can make use of a digital environment that facilitates queries and reviews over a comprehensive field corpus.

Keywords: network traffic analysis; data curation; reproducible science; meta-analysis

1. Introduction

Communication networks are perhaps the technology breakthrough that has caused major impacts in the worldwide socio-economic structure in modern times. As the object of scientific curiosity, communication networks constantly generate overwhelming volumes of data assets that can be analyzed and studied. If we focus on the network level, we refer to this investigation as Network Traffic Analysis (NTA). Taken as an applied science, NTA research is relevant to improve, optimize, and reduce failures in communication infrastructures and services; nevertheless, security aspects clearly stand out as the principal focus of NTA research from the very beginning [1]. RFC 3917 [2] lists usage-based accounting, traffic profiling, traffic engineering, attack/intrusion detection, and QoS monitoring as key application fields for traffic capturing. While accounting and QoS monitoring are currently usually covered by mature standards, research works tend to focus on traffic classification, anomaly detection, or specific attack identification. In addition to the practical usefulness, the attention of the research community is fully justified since, from a data science perspective, NTA is one of the most challenging fields due to its intrinsic peculiarities, for instance big data, the high variety of feature representations, evolving scenarios, stream data, adversarial environments, encryption, or limitations imposed by privacy concerns.

In this regard, the compendium of publications that tackle NTA increases year-by-year. A search by topic of papers on the Web of Science (https://apps.webofknowledge.com/) using "network traffic" as the keywords found 12,085 publications (consulted on 5 February 2020), from which 21.8% were published as of 2017. Specifically, NTA at the network and transport layer attracts a considerable part of the research attention as it is:

- low-intrusive (i.e., privacy respectful),
- fast and lightweight,
- applicable to big volumes of traffic,
- suitable for embedding in network middle-boxes.

Surprisingly, in spite of the high number of related publications, there are no standardized methods, algorithms, or steps to dig into network traffic data from analytical perspectives. This deficiency has been emphasized several times. For instance, Kim et al. [3] claimed that "recent research on Internet traffic classification algorithms has yielded a flurry of proposed approaches for distinguishing types of traffic, but no systematic comparison of the various algorithms". On this matter, we hypothesize that the fast evolution of communications and the push for new applications have complicated the characterization of network traffic and the achievement of unified criteria about how to analyze it. However, a significant part of the research corpus shows repeated structures, i.e., many papers follow the recurrent scheme summarized in Figure 1, which is also supported by several field surveys [1,4–6].

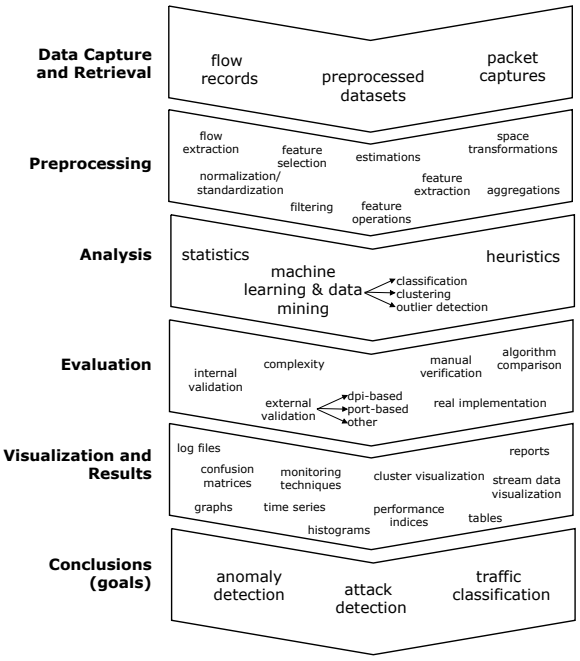


Figure 1. Common scheme in NTA papers and the underlying structure of NTARC objects.

In this work, we present the Network Traffic Analysis Research Curation (NTARC) model. NTARC is a data model that stores key aspects of NTA research publications. A database of NTARC objects is intended to increase the value of past research as it enables the automated retrieval, reuse, comparison, and analysis of published papers. Additionally, it facilitates reproducibility and the consolidation of standardized methodologies and best practices. NTARC emerges because the current way of reusing past field research is obsolete, manual, subjective, does not facilitate reproducibility, and misses opportunities opened by modern advances in data sharing. This problem is not specific to NTA, but generalized in science and getting more and more attention. There are no similar approaches to automatize the study of past research in NTA with the deepness pursued by NTARC, which also plays

the role of a methodology template. Its adoption can help to settle best practices since researchers become aware of methodology deficiencies and errors and are encouraged to create more reproducible works. Curating research and creating NTARC objects comprise a significant effort, but science demands it as it demands methods to increase credibility, quality, and efficiency in the knowledge accumulation [7]. NTARC is designed specifically for research works that propose analysis methods for network traffic captured at the network level, but it could be similarly developed for other scientific fields provided they share extended, common experimental and methodological structures as the one shown in Figure 1, which is characteristic of the field here under review.

Finally, we show the potential of NTARC by revisiting the NTA research conducted during the last two decades. To this end, we explore the papers included in the latest release of the NTARC database [8]. We aim to offer a representative, overall snapshot of the main trends observed in the field, also emphasizing the drawbacks and reasons that could be hampering the research efficacy and making novel proposals unfeasible and far from satisfying the requirements of real-life applications. Our field review focuses on NTA's main goals, research foci, selected features, datasets used, analysis approaches, evaluation methods, claimed contributions, and reproducibility.

The rest of this paper is organized as follows: Section 2 explores previous data sharing initiatives that have been developed to improve scientific research. Section 3 provides a detailed description of NTARC's internal structures (fields and subfields). Section 4 lists and explains a set of tools developed for the edition, revision, verification, sharing, and analysis of NTARC objects and databases. Section 5 introduces early initiatives to expand NTARC and encourage its use in the research community. In Section 6, we present the current release of the NTARC database with a collection of 80 curated papers from 2002 to 2020. Section 7 elaborates on a systematic review of the NTA research embraced by the NTARC database. Conclusions are provided in Section 8.

2. Scientific Data Repositories and Data Sharing Initiatives

There are several projects aiming to host and promote research data repositories with a general, non-field-specific purpose; for example: B2SHARE

[9], Figshare [10], the Globus Data Publication service [11], 4TU.Researchdata[12], the Zenodo platform [13], Dataverse [14], and Dryad [15].

Assante et al. [16] wondered if generalist scientific data repositories were able to cope with the requirements of research data publishing and concluded that generalist repositories suffer from dealing with a multiplicity of data formats and topologies, a highly varied and multidisciplinary community of data owners and consumers, and a lack of consolidated and shared practices. Repositories were found viable, but conservative and in need of evolving. Such intrinsic heterogeneity seems to be a pressing problem to tackle in the near future to prevent underused repositories. When the research scope of a given repository is more specific, the heterogeneity problem is minimized. For instance, NASA's Common Metadata Repository (CMR) for Earth Science Data Information is a good example of a system developed to standardize and solve a past inefficient data retrieval situation. The use of Earth science data involves a community of Earth scientists, educators, government agencies, decision makers, and the general public. The CMR initiative has highly increased the value of past research and metadata [17].

The success of data repositories partially lies in creating metadata structures able to categorize and identify datasets and research objects effectively. In [18], Devarakonda et al. defined metadata as structured information that describes data content. Metadata explains the definition of measured and collected variables, their units, precision, accuracy, layout, transformations, limitations, etc. In addition, it should also clarify the data lineage, i.e., how data are measured, acquired, and preprocessed. Hence, metadata facilitates data sharing, access, and reuse. In addition—as claimed in [18]—metadata must be accessible in a format that is easily adaptable to technology changes, e.g., XML and JSON (used in NTARC).

However, the adoption of data sharing practices and detailed metadata descriptions is still immature in the scientific community. Harrison et al. [19] recognized several challenges to face, mainly concerning the evolution of researchers' mindsets and habits. Nevertheless, the authors foresaw a near future in which data resources would be assessed similarly to journal publications in the scientist portfolio, increasing therefore both the quality and quantity of published data. Within the context of environmental sciences, Harrison et al. also presented a workflow to publish datasets, models, and model outputs, enabling the access, reuse, and citation of data products [19].

Scientific journals play an important role and endorse or directly develop platforms to improve how scientific publishable material is shared, managed, and reused. Dryad [15] publishes datasets related to peer-reviewed journal articles and scientifically reputable sources. In [20], Bardi and Manghi explored the concept of enhanced publications, meaning digital publications that incorporate ways to access and disseminate research materials beyond papers. They found that such initiatives are hindered by the fact that researchers must face several difficulties, e.g., manual efforts in curating data, preparing the material, or acquiring new skills to adapt their data, obtaining no obvious, direct benefits. Therefore, enhanced publications are more common when scientific journals demand such materials with clear policies. Similar conclusions and claims were exposed in the data journal survey in [21]. In [22], tools and digital environments were proposed to reduce costs and facilitate the creation of enhanced publications. In this line, Science Object Linking and Embedding (SOLE) [23] is a tool intended to enable reproducible research by linking research papers with associated science objects. Here, science objects are linked with tags in a bibliography-like form, making their reference easy. In this respect, Scientific Data, launched by the Nature Publishing Group in 2014, is a peer-reviewed journal that focuses on data descriptors, defined as "a new type of publication that combines the narrative content characteristic of traditional journal articles with structured, curated metadata that outline experimental workflows and point to publicly archived data records"[24].

Making research data, results, and materials more profitable is a necessary process that involves manual effort. By analyzing the relationship of institutional repositories with small science, Cragin et al. [25] claimed the necessity of redefining and standardizing the understanding of "data sharing", as well as promoting the establishment of data curation policies to empower the use of data repositories and protect against data misuse. In a similar line, the USA National Research Council recently published a study about digital curation with the title *Preparing the Workforce for Digital Curation* [26]. In this work, digital curation was deeply analyzed, discussing the current status and practices, society requirements, career paths, professional opportunities, derived benefits, and importance for the scientific advancement. Digital curation is defined as "the active management and enhancement of digital information assets for current and future use". In the conclusions, the authors first emphasized the limitations and missed opportunities due to the current immaturity and ad-hoc nature of digital curation. Digital curation is not well understood, but the application of digital curation in organization practices is expected to reduce costs and increase benefits. Some examples of organizations focused on promoting and developing digital curation are: the Digital Curation Center (DCC), the National Digital Stewardship Alliance (NDSA), Research Data Alliance (RDA), and the Committee on Data for Science and Technology (CODATA).

Several examples exist that directly show how using data curation and metadata models can benefit science; for instance, the Linking Open Drug Data (LODD) project, which is a task force within the World Wide Web Consortium's (W3C) Health Care and Life Sciences Interest Group (HCLS IG). LODD gathered and connected reliable information about drugs that are publicly available on the Internet, uncovering relevant questions for the science and the industry, and providing recommendations for best practices [27]. As for the NTA field, the main precedents have been developed by the Center for Applied Internet Data Analysis (CAIDA), i.e., DatCat, an Internet Measurement Data Catalog (IMDC), which is "a searchable registry of information about network measurement datasets" [28], and the Internet Traffic Classification [29], which is a collection of 68

curated metadata of NTA papers published between 1994 and 2009. Worth mentioning also is the IMPACTproject [30], but in a wider perspective.

The NTARC model goes a step further and proposes a detailed collection of metadata that fits the structure shown in Figure 1. The goal is to improve the reuse of previous research by enabling the use of statistics on data summaries and metadata or meta-analysis (understanding meta-analysis in a broad sense). Meta-analysis consists of bringing together different studies about the same research question and applying statistics and analysis methods to obtain global conclusions and general perspectives. Meta-analysis is a perfect procedure to glue together small science in a global context and transform independent works into more profitable parts of the complete science building. This is specially true when the same research question is repeatedly faced by different teams in different places. Meta-analysis has been actually determinant in fields like medicine, pharmacology, epidemiology, education, psychology, business, or ecology. A well-known introduction to meta-analysis was offered by Borenstein et al. [31]. Specifically for medical research, we address the reader to [32]. Meta-analysis can also be satisfactorily applied in technical research and engineering, but detailed data models must be previously created. Such models will pave the way toward standardized procedures, which are required for reliable meta-analyses.

3. NTARC Data Structures

An NTARC object is a digital summary of a peer-reviewed NTA scientific publication. NTARC publications are required to fit the scheme shown in Figure 1. Additionally, every NTARC object must be compliant with the NTARC model, which follows the structure depicted in Figure 2.

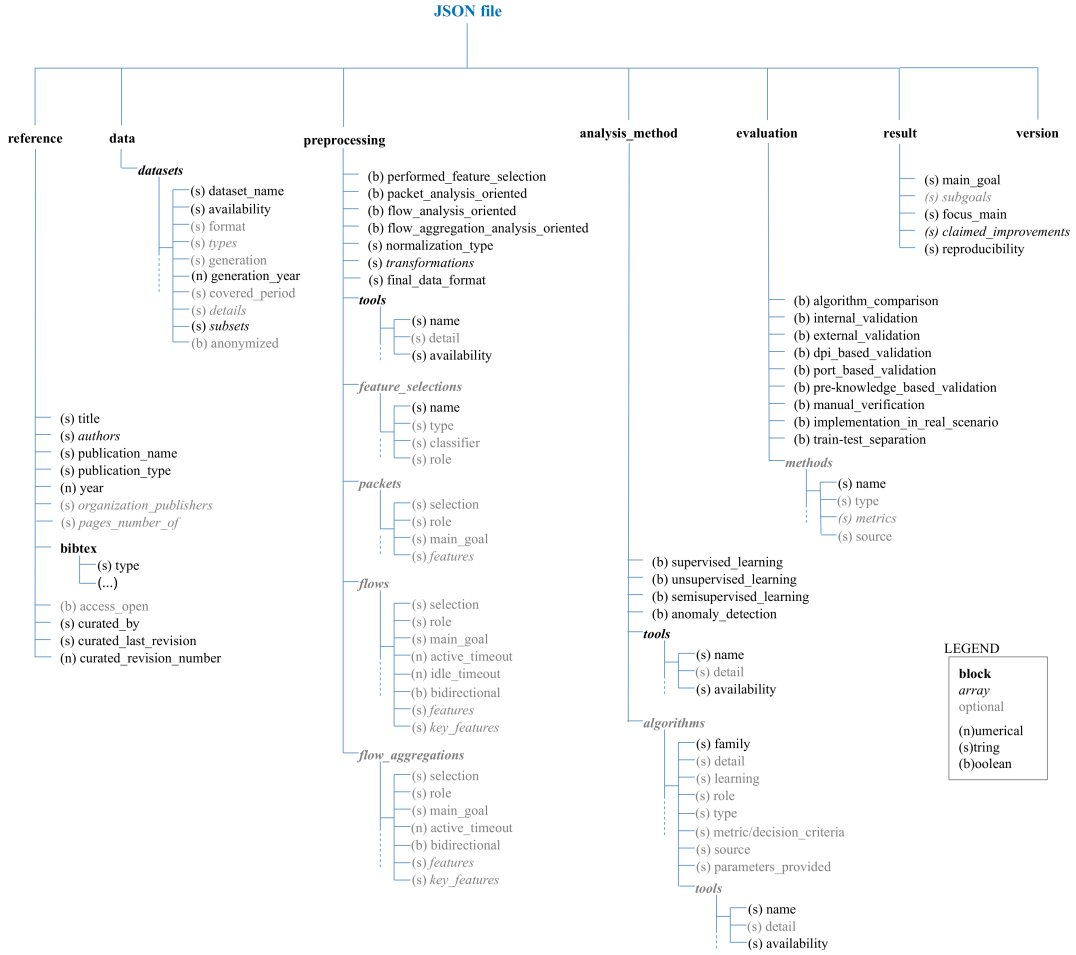


Figure 2. Tree-like scheme of an NTARC object. The major subtrees are: `reference`, which contains the data for identifying the scientific work and metadata describing the details of the NTARC object curation; `data`, which describes the datasets used; `preprocessing`, which contains selected features and feature transformations; `analysis method`, describing the analysis algorithms used; `evaluation`, depicting the evaluation metrics; `result`, summarizing the claimed results, improvements, and reproducibility of the scientific work; `version`, which specifies the NTARC version used for this object.

NTARC uses the JSON format [33], which can be easily parsed and written by computers while still being human readable. Creating, sharing, and distributing NTARC data are simple and straightforward since JSON files are text-based, and each file addresses only one scientific publication. A first, minimal prototype was used for the research conducted in [34].

The readers will notice that fields defined in the NTARC structure are exhaustive. For the sake of flexibility and time optimization, some fields in the structure are mandatory, and other fields are optional. Furthermore, contributors are free to define their own fields that might be added to future NTARC versions or will simply remain as notes.

The NTARC model consists of six main blocks: `reference`, `data`, `preprocessing`, `analysis_method`, `evaluation`, and `results`. Additionally, a `version` field stores the NTARC version that corresponds to the JSON object. The `version` field helps automated tools during parsing processes and makes the format future-proof. The main blocks are described as follows:

- The `reference` block:
  This block collects information that identifies the scientific work, the publication media, and the curation process itself.

- The `data` block:
  This block stores information about the network traffic data used. It is not intended to refer to the original dataset version, but the version retrieved by the paper authors, which might have been modified. Here, we find one of the anchors that facilitates comparative analysis, since the scope of the data is always NTA and is provided in the shape of either packet captures, flow records, or preprocessed data derived from packet captures or flow records.
  The `data` block consists of one or several `datasets`. By definition, two `datasets` must be reported separately when they clearly come from different setups, projects, or sensor environments; otherwise, they must be defined as `subsets`.
- The `preprocessing` block:
  This block summarizes all transformation and modification processes that datasets underwent previous to the main analysis (e.g., normalization, dimensionality reduction, feature extraction, filtering). The stored information is limited to the preprocessing specifically mentioned in the paper as a part of the presented methodology. This block also captures the set of network traffic features and/or flow keys that were used to represent traffic during subsequent analysis.
  Most fields in this block are binary and mandatory, allowing a fast curation of relevant preprocessing aspects. Subsequent blocks (e.g., `feature_selections`, `packets`, `flows`, and `flow_aggregations`) are optional, being suitable for cases where a more detailed, fine-grained definition is desired. Specifically, `packets`, `flows`, and `flow_aggregations` are blocks that indicate the type of traffic objects to analyze during experiments. The habitual trend is focusing on only one of these traffic objects.
- The `analysis_method` block:
  This block depicts the analysis. It captures relevant details of the analysis methodology and identifies the algorithms used. Note that here, `tools` are repeated at two context levels: general for the analysis method and specific for algorithms.
- The `evaluation` block:
  This block gathers information to understand how analysis outcomes were validated, evaluated, and interpreted. It basically registers the metrics used and the perspectives that the authors found relevant to assess the analysis success or failure.
- The `result` block:
  In this block, goals, sub-goals, and improvements claimed by the authors are collected. It also defines the focus of the paper and if the published work meets reproducibility standards [35].

4. Tools Developed for NTARC

NTARC formats, structures, and libraries are openly available [36]. In addition to format specification, we provide a broad, complete documentation with examples, editing rules, and explanations. Thus, contributors are guided in the process of curating data and creating NTARC objects, and users are guided in the process of exploiting NTARC datasets. Being built on top of a standard format like JSON, NTARC benefits from all existing tools already developed by third parties. In addition, we developed several tools to facilitate the curation process and the interaction with the NTARC database. We mention here some of these tools, which are openly available in [37].

4.1. JSON Schema

The JSON Schema is a format that allows the formal specification of what constitutes a valid JSON file for a particular application [38]. We maintain a JSON Schema that formalizes our description of NTARC files [36]. This schema helps verification tools to validate NTARC files and paves the way for the development of additional tooling. The periodic revision of mismatches between dataset objects and the JSON Schema enables the detection of errors, ambiguities, new trends, or missing values, as well as the updating of the Schema itself.

4.2. NTARC Editor

NTARC objects can be directly edited with common text-editors. However, the NTARC structure includes many attributes and is based on JSON, which is a language that makes use of extensive punctuation; therefore, errors can easily occur during direct manual editions. A custom editor was developed to lighten the generation of NTARC files. Rather than having all fields of the specification implemented, the editor creates the user interface from the JSON Schema described above.

The editor incorporates the specification and includes pointers to the documentation at the appropriate places. Additionally, fields not conforming to the specification are marked, and error messages are displayed. To further ease editing NTARC objects, the editor implements an interface for specifying network traffic features in a more formula-based language. Finally, to accommodate heterogeneous computing environments and boost the adoption of the NTARC format, the NTARC editor was built with cross-platform capabilities in mind. This was achieved by using the Electron framework [39], which allows building cross-platform applications using HTML, JavaScript, and CSS. The source code and pre-built binaries for Linux, Windows, and MacOS are freely available from [37].

4.3. Verification Tool

To minimize human errors during curation processes, we developed a verification tool that automatically checks NTARC objects when submitted to databases. This tool is freely available in [40]. The verification tool uses the JSON Schema as a first step to assert that the submitted file is compatible with the NTARC specification; afterwards, a second analysis is performed by parsing the file with the NTARC Extraction Library (Section 4.4). The first step checks grammar and NTARC structural consistency; the second steps works deeper and checks syntax and semantic aspects (for instance, a defined division operation must necessarily come with two terms: a numerator and a denominator).

4.4. NTARC Extraction Library

The NTARC Extraction Library is a Python library that enables information extraction from NTARC objects and datasets. This library implements methods and classes linked to the multiple defined blocks and fields. Therefore, it is possible to perform deep searches and queries in databases (i.e., metadata analysis) by using keys with any combination of fields and values. Additionally, as previously mentioned, the library allows the verification of NTARC files. The library also supports calls to external APIs with the capability of augmenting the information available in the NTARC files. For example, by querying the Microsoft Academic Services API [41], the library can collect additional relevant information that does not appear in the paper and store it in a local cache, e.g., number of citations and authors' affiliations. The extraction library is freely available from [42].

4.5. Content Validation

The tools presented above are used to ensure that new NTARC objects are compliant with the specifications. Therefore, all files included in the dataset are previously verified and consistent in terms of grammar, syntax, and structure, meaning that they can be used for analysis and information extraction. However, the curation of paper information is manual in essence and requires experts reading papers and abstracting contents according to the NTARC structure. Therefore, errors and subjectiveness are possible and happen often. In spite of the efforts for creating supporting tools, documentation, and reducing ambiguities in the file format, some issues are impossible to control automatically; for instance: data curators' misinterpretations, uncommon terminology, fundamental methodology aspects that are missing or unclear in the paper, etc. In such cases, field values might be compliant with the NTARC specification, but wrong with regard to the research under curation. In this respect, the `reference` block includes a `curated_revision_number` field that shows the number of times an NTARC object has been reviewed by curators, enabling the control of data revisions.

5. Dissemination Initiatives

Initiatives to make NTARC fully accessible for the scientific community include the publication of the NTARC database in generalist repositories, the open availability of NTARC documentation and tools for creating, accessing, analyzing, and updating content, endorsing the NTARC grown in academic centers, and directly contacting authors and encouraging the inclusion of NTARC data objects for participation in related conferences and workshops.

Regular citable releases of NTARC databases are provided via Zenodo in [43]. Additionally, the whole project, including databases, tools, documentation, and specification, is fully accessible through GitHub: [8,40,42]. Therefore, external curators and users have a complete environment to submit contributions, download research resources, and obtain feedback.

As a part of the NTARC project, Masters' and Bachelors' degree university students, in addition to using the NTARC database for their respective research, also review papers as part of their academic portfolio. This initiative helps the NTARC database to grow, as well as promote the critical reading of scientific publications among students, who get familiar with the field state-of-the-art and additionally are trained in methodologies of scientific experimentation and dissemination. Students are also encouraged to contact the original authors during the paper curation process, so extending the NTARC network and creating links between researchers and students.

Finally, an ongoing plan is to require NTARC objects as additional publishable material for papers accepted in related conferences, workshops, and journals. This initiative pursues the increase of potential contributors and users of NTARC and, at the same time, raises awareness within scientists of the importance of data sharing, reproducibility, and the consolidation of best practices.

6. The NTARC Database

The NTARC database is released with 80 NTARCv.3 objects corresponding to research papers published between 2002 and 2020. The database is constantly growing and being updated by NTARC developers, research authors, and any external contributors after a proper evaluation of the submitted NTARC object. It is accessible both from GitHub [8] —which allows accepting outside contributions, automatic tests, as well as keeping a history of the changes—and Zenodo—for providing regular citable snapshots [43]. The database is made available under a Creative Commons Attribution 4.0 license. This license is suitable for databases, allows everyone to use, adapt, and share the data, and requires an indication of changes made.

Dataset objects are text-files in JSON format that follow the NTARC.v3 structure. Broad documentation, examples, and tools are openly available at [36]. The initial selection of works was performed by using Google Scholar and searching with the keywords: "traffic classification", "network traffic analysis", "traffic monitoring", "anomaly detection in communication networks", and "forensic analysis of traffic". The criteria to prioritize papers were:

- Papers matching the structure in Figure 1.
- Citations. Highly referenced papers were the priority.
- Year of publication. Recent papers were the priority.
- Publication medium. Papers published in top peer-reviewed journals and conferences were the priority (based on high scientometric indices, e.g., the impact factor).

An overview of the number of papers per year in the current release of the NTARC dataset is shown in Figure 3. References are included in the bibliography: [44–115]

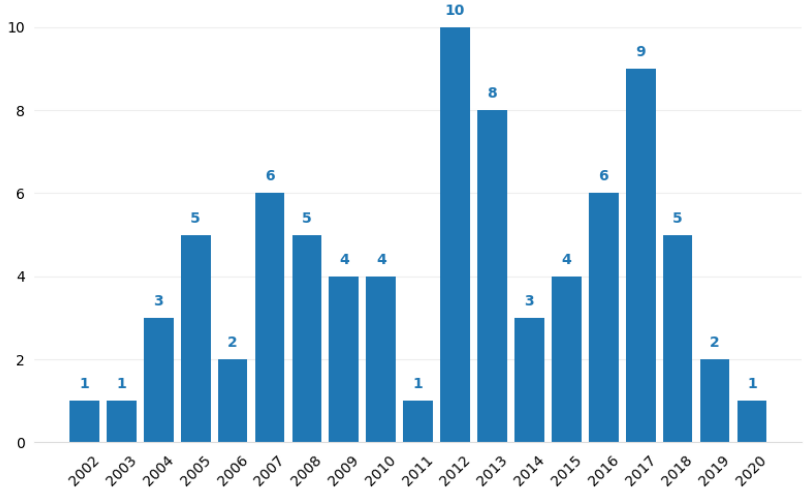


Figure 3. Number of papers per publication year.

7. Analysis of NTA Research (NTARC Database)

In this section, we use the NTARC tools and database (Sections 4 and 6) to perform a critical exploration of the top NTA research conducted during the last two decades. The reviewed NTARC database release is [43]. We explored goals, foci, datasets, features, predominant algorithms, claimed contributions, and reproducibility aspects.

7.1. Research Goals

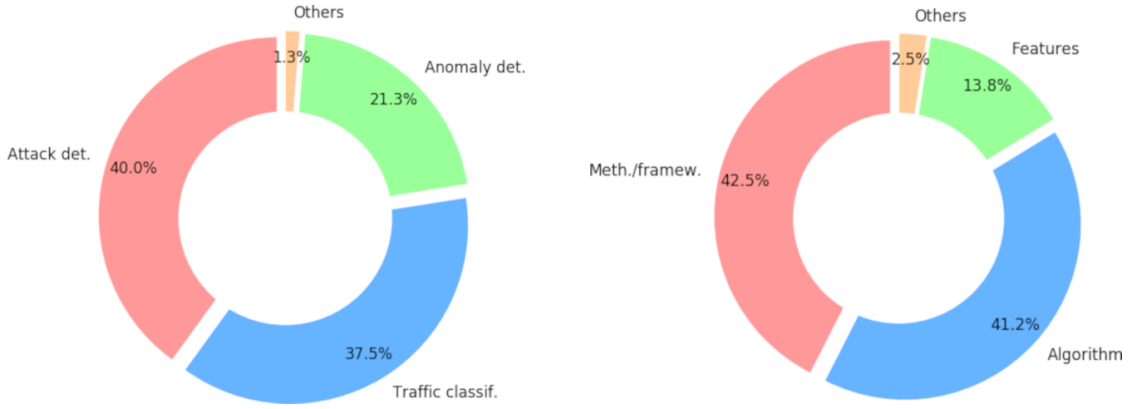
In the explored works, we found three main goals pursued by NTA research (Figure 4a):

- Attack detection. This is whenever the applicability of the proposal detects attacks in network traffic, meaning the identification of traffic associated with malicious behavior. An example is the work by Potluri and Diedrich in [116].
- Anomaly detection. This is if the paper aims to detect anomalies in network traffic, meaning traffic that is abnormal, breaks expected patterns, or cannot be defined as following normal behaviors. Such anomalies do not have to be necessarily malicious, and the authors do not address a priori any particular attack, scheme, or traffic class. An example is the work by Bhuyan et al. in [117].
- Traffic classification. The is whenever the methods in the paper identify specific classes in network traffic. Such papers are not focused on the identification of attacks; otherwise, attack detection would be the appropriate goal. An example is the work by Wright et al. in [118].

7.1.1. Attack Detection

Attack detection is the main goal and covers approximately 40% in the NTARC database. It usually deals with methods that are: (a) binary or dichotomous (i.e., the goal is bisecting traffic into attack-related and non-attack-related instances) or (b) multiclass. The common trend in multiclass classification is a learning scheme that reserves several classes for different types of attacks and one class for normal, legitimate, or non-attack-related traffic. Both (a) and (b) lead to problem spaces with specific idiosyncrasies, yet describing the peculiarities of such spaces does not receive a proper attention in the research. The number of works depicting and visualizing network traffic spaces is small, even in spite of the importance of visualization for the successful application of any kind of data mining, machine learning, or statistical analysis method [119].

Furthermore, regardless of the type of analysis (binary or multiclass), traffic classes are commonly retrieved from Intrusion Detection System (IDS) datasets without proper discussion about the class meaning, the label assignment, the nature of attacks, and the attack deployment within the tested data.



(**a**) Top-level research goals. (**b**) Main research foci.

Figure 4. NTA top-level research goals and main research foci (NTARC database).

To give an example, in binary classification (attack/non-attack), it is commonly not clear which binary label should correspond to backscatter traffic (for a detailed description of backscatter traffic, see [120]). Backscatter traffic is not formed by attack packets, but indirectly caused by malicious activities, which provoke that vulnerable servers to generate such spurious traffic. Furthermore, in general, low attention is dedicated to the high variability and network dependence of the "normal" class, which is a key factor to draw the underlying traffic picture in which attack classes are superimposed.

A third relevant insufficiency is the frequently obviated post-analysis, i.e., checking the reasons and sources of misclassifications. This means, for instance, checking if classification errors are due class overlap among attack types or if, instead, attacks are mixed up with specific "normal" traffic shapes.

7.1.2. Anomaly Detection

In anomaly detection research, we found two main trends: (a) analysis of multi-dimensional data points, in which the studied objects are commonly communication flows; and (b) analysis of time series, in which time series usually represent aggregated network data or network properties.

A peculiarity often observed in anomaly detection research is that the definition of "anomaly" is usually preconceived, therefore resulting in a blurred construct that mixes notions of novel attack, disruption, abnormality, and outlierness. Given that real traffic contains massive volumes of known, irrelevant traffic (i.e., novel attacks are negligible in comparison), real applications require the strong prevention of false positives, which might become as relevant as minimizing false negatives. This issue was already mentioned by Axelsson [121] in the past century, who claimed that IDS must face the base-rate fallacy challenge. Therefore, anomaly detection proposals must sooner or later cope with some evidence, namely:

- Harmless, legitimate traffic is often also anomalous and a source of deviations as well.
- Neither attacks, nor anomalies have to appear isolated, but can occur as bursty events or small clusters.
- The feature space and the underlying distributions of real traffic have a strong, decisive impact on the performance of unsupervised analysis methods.

Neglecting these aspects leads to solutions that might be deemed as irrelevant and unpractical. For this reason, some suggestions for best practices are:

- Clear, unambiguous definitions of the type of anomalies.
- The study of the selected NTA features and the problem spaces drawn by feature sets.

- The use of datasets with distributions that are representative of real-life scenarios. For instance, testing unsupervised detection frameworks with synthetically crafted IDS datasets is not recommended.

7.1.3. Traffic Classification

The last main top-level goal in NTA is traffic classification (about 40% of the studied papers). Whereas attack detection tends to use supervised machine learning and anomaly detection usually resorts to unsupervised algorithms, more varied options are developed for traffic classification. Here, the use of heuristics, schemes designed ad-hoc, and pre-knowledge not extracted by algorithms (i.e., known rules) are common. The frequent methodological weaknesses mentioned for the research on the previous top-level goals also appear in traffic classification, yet this type of research is more heterogeneous, and therefore, finding common peculiarities is also more difficult.

7.2. Research Foci

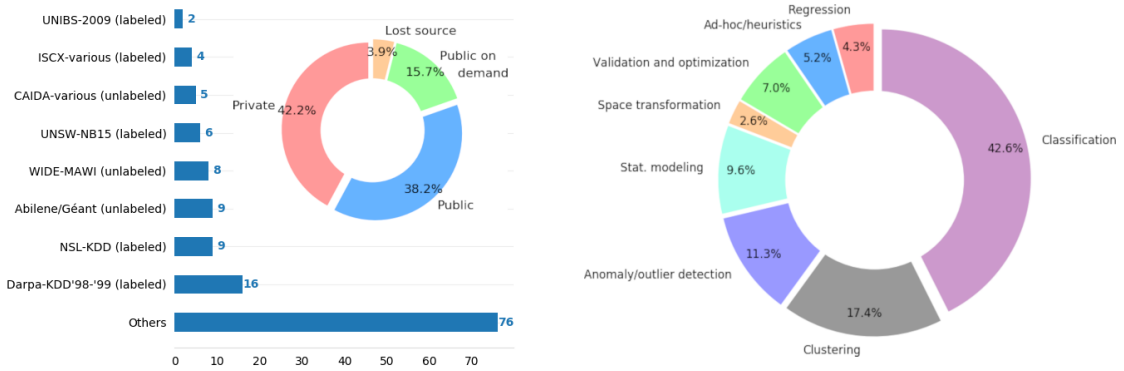
By querying metadata to explore where authors establish the focus of their proposals (Figure 4b), we found three main pillars: (a) the algorithm, (b) the methodology/framework, and (c) the features. The distribution in Figure 4b shows that the research effort mainly focuses on making NTA more accurate by the application of novel methodologies and algorithms (83.4%), and comparatively less attention is given to understanding network traffic phenomena or abstracting knowledge from collected data. The study of features is still relevant (13.8%), but the analysis of patterns, the study of outliers, or enhanced data descriptions are not core aspects of the research. These figures do not imply that papers do not tackle such low-rated tasks; instead, they emphasize that, when some kind of knowledge discovery is undertaken and disclosed, this is not claimed as the principal novelty of the scientific work. A closer look at what authors claim as their principal contributions corroborates this point (Section 7.6)

Actually, discovering knowledge in data should be easier to face than designing effective detection systems, a task that implies some constraints that are difficult for scientists to overcome (see Section 8). Here, it is worth remarking that the high complexity of network traffic data justifies its use for testing novel analysis algorithms, frequently without taking into account the feasibility and transportability of the techniques used for real-life environments.

7.3. Used Datasets

The available data for experimentation play a determinant role in the scientific research. Even in spite of the fact that the datasets used come from very different sources (Figure 5a), DARPA-KDD datasets (traffic collected in 1998 and 1999) are very popular and still being used today (note that all studied papers in the NTARC database have been published after 2000, 2012 being the median year of publication). However, such datasets have not been representative of network traffic for a long time, both in terms of attack classes and legitimate traffic forms. Such insight reveals a common lack of reliable, high-level network data for science, a problem that has been reported several times [122,123]. Issues related to data privacy, enterprise security, and governmental interests minimize the possibilities for data sharing. CAIDA [124] and MAWI [125] are organizations that lead initiatives for publishing network data for research purposes; however, even in spite of these efforts, available data are preprocessed and shared without payloads and with anonymized IP addresses, considerably reducing the value for investigation. In this sense, the Canadian Institute for Cybersecurity [126] is also worth mentioning, perhaps the most active group in the creation and publication of high-quality IDS datasets (e.g., the ISCX and CIC dataset families).

Figure 5a also shows a disturbing ratio of datasets that are not publicly available ("private" or "lost sources"), a fact that hinders reproducibility and honest research.



(**a**) NTA datasets and their availability.

(**b**) Types of analysis (algorithm families).

Figure 5. NTA datasets and types of analysis (algorithm families). The algorithms shown include the authors' main proposals, as well as other approaches used as benchmarks.

7.4. Features

Lim et al. [127] emphasized feature selection as one of the main challenges of machine learning-based pattern recognition and classification in NTA. Literally, the authors remarked: "(i) key feature selection, (ii) finding the best algorithm(s) for traffic classification, and (iii) obtaining representative data sets for training and testing machine learning algorithms".

The selection of network features that are relevant for NTA is an open question that has been addressed several times from analytic perspectives. Since there is no global agreement in this regard, feature selection is an expected step and part of best practices recommendations (mainly for supervised analysis). Thirty-five percent of the studied works conducted some type of feature selection. In parallel, Lim et al. also emphasized the importance of studying the discriminative power of features [127].

An example of a study focused on features is [128], where stability selection and diverse filters and wrappers were applied for feature selection, concluding a set of the 16 most relevant features. This set is hardly generalizable since it is inevitably linked to the NSL-KDD dataset. This dataset and other datasets belonging to the popular DARPA-KDD family are published pre-processed and account for an original set of 41 features. As mentioned in Section 7.3, normal traffic and attack vectors in these datasets were captured and generated before the year 2000. The facts that network communications, applications, and protocols are so variable and evolve so quickly make it difficult for analytical approaches to be representative, and they soon become outdated. Furthermore, the possibilities for extracting features from network data are immense. Nevertheless, the study showed how researchers highly disagree about the relevance of features even when targeting the same questions with the same datasets and the same initial set of 41 features.

A meta-analysis with an early version of NTARC (NTARC.v1) was used to explore the problem of feature selection in NTA in [34]. This work explored what the scientific community recommends by achieving a consensus based on the main, most cited works published from 2002 to 2017. Results showed a set of 12 features that clearly stood out. These features are:

(1) `octetTotalCount` ,
(2) `ipTotalLength`,
(3) `destinationTransportPort`,
(4) `sourceTransportPort`,
(5) `flowDurationMilliseconds`,
(6) `packetTotalCount`,
(7) `destinationIPv4Address`,
(8) `sourceIPv4Address`,
(9) `protocolIdentifier`,
(10) `server_to_client`,
(11) `client_to_server`,
(12) `interPacketTime`.

The meaning of the features can be consulted in the IANA-IPFIX documentation [129], except for Features (10) and (11), which simply mark flow direction. This feature set was used in a comparative study of lightweight NTA feature vectors [130], achieving the best commitment between accuracy and processing costs together with the feature set introduced in [54], which consisted of 30 features. Even in spite of the fact that the 12 feature set obtained in [34] might be insufficient for complete, high-accuracy detectors, it might be taken as a benchmark or a coarse-granularity phase for NTA frameworks.

7.5. Predominant Algorithms

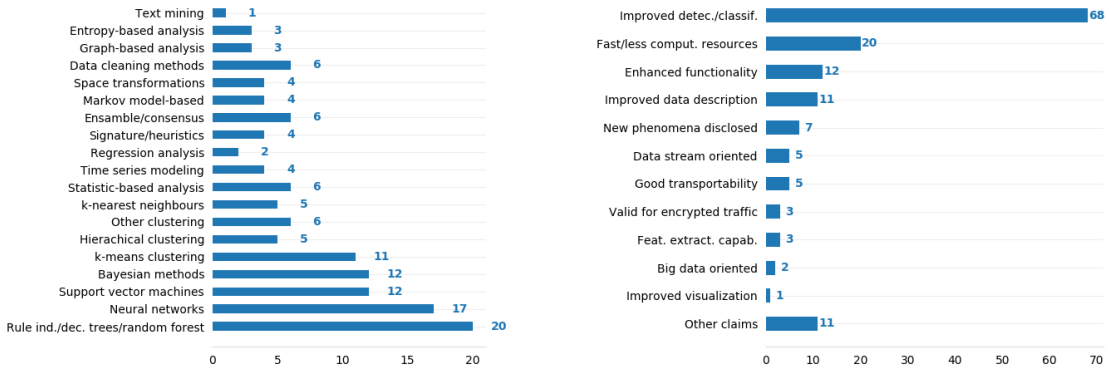
A considerable part of the studied analysis approaches belongs to supervised methods, specifically in works performing attack detection. Figure 5b shows the overall share of types of analysis techniques. Exclusive unsupervised methods (e.g., clustering, outlier detection) are also common. The application of both supervised and unsupervised approaches has its grounds. The tendency to use unsupervised methods can be explained by the higher availability of unlabeled TCP/IP data in big volumes; however, such proposals face strong difficulties when validating models and results, which can hardly be exhaustive or unambiguous. For this reason, supervised methods are often preferred, although a main problem in this case is the alarming scarcity of labeled data. Labeled datasets are extremely difficult to create, become obsolete soon, can hardly claim to be representative, or include novel forms of attacks, and the generation mechanisms are often questioned.

As shown in Figure 5a, data availability and quality are key aspects that determine how research is conducted. Actually, criteria for selecting analysis approaches often seem to be more related to convenience than utility, causing methodological trends that might be even incongruous or unpractical. For instance, works addressing traffic analysis from global perspectives with single, traditional, one-step methods are very common; however, some experts consider that the high complexity of traffic analysis stands out for multi-step, multi-phase solutions [131]. Furthermore, semi-supervised approaches seem to be suitable according to the problem specification (i.e., using pre-knowledge to identify old attacks and being able to detect the evolution of old attacks, as well as new attack schemes), but they only covered 9.7% of the consulted papers.

Figure 6a shows the algorithm families of the methods proposed by authors as the most suitable ones. Outstanding options are:

- Rule induction, decision trees, and random forests:
  Using decision trees and similar approaches has strong grounds given the peculiarities of network traffic data. Such methods are robust, easy to adjust, not affected by irrelevant features, capable of working with mixed datasets (i.e., numerical and categorical data together), and provide interpretable solutions that help create knowledge and understand the contribution of network features. On the other hand, a main drawback is that class imbalance severely affects such methods. This situation is typical in network traffic datasets, in which differences in class representations can account for several orders of magnitude.
- Neural networks and support vector machines:
  These two algorithm families share some common drawbacks. They behave as black boxes (i.e., knowledge extraction is hard, if not unfeasible), involve high computational costs, require complex parameterizations, are prone to suffer instability, and commonly demand feature transformations and painful increments of dimensionality. Nevertheless, ideally, both options are able to obtain highly accurate results. The last years' growth of data availability in big volumes and the increase of computational power have entailed a considerable technology push favoring them. Both support vector machines and neural networks are suitable for network traffic data spaces, which are normally high-dimensional, and the class shapes are not necessarily globular. Validation, implications of noise, local-minima problems, and lack of robustness—especially when considering adversarial settings—are issues often obviated, but key in real applications, which must endure with minimal degradation. Noteworthy is the fact that, as of 2016, deep learning has captured the majority of supervised learning proposals.

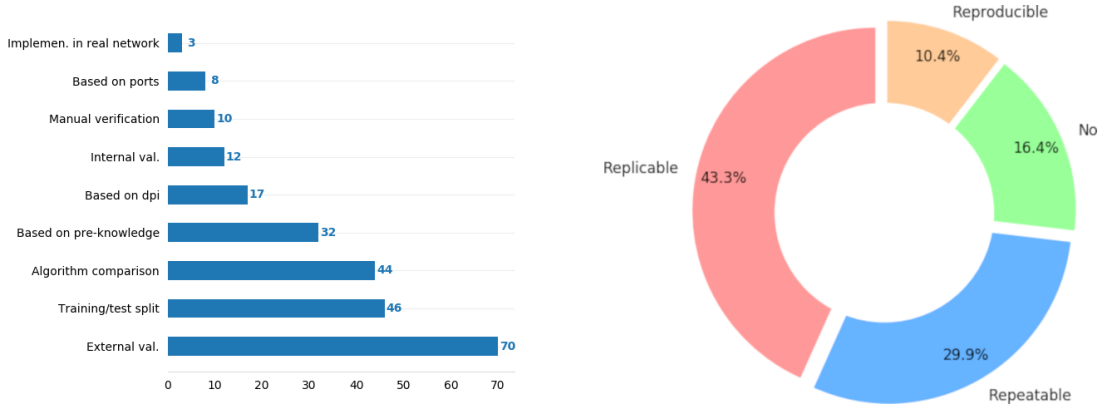
- Probabilistic and Bayesian methods:
  These methods are often used due to their simplicity, high speed, suitability for high-dimensional data, and the fact that probabilistic decision-making is appropriate for feature vectors that contain very different types of properties. However, naive solutions assume that features are independent, and more complex Bayesian methods require the modeling of such dependencies. In this regard, network traffic features are prone to show high correlation [128], a fact that in principle advises against Bayes-based approaches.
- Clustering:
  NTA unsupervised methods consist mainly of clustering. Unsupervised methods are more often applied as parts of the analysis frameworks than supervised methods. Clustering is frequently used for data reduction or space simplifications, also after other space transformations (e.g., PCA, graph representation, SOM). Therefore, their suitability must be assessed within the corresponding framework. Two main trends are observed here: prioritizing fastness (K-means and K-means variants) or accuracy (e.g., hierarchical clustering, DBSCAN). K-means is a simple, fast algorithm, but unstable and prone to generate suboptimal results. Hence, internal validation is almost mandatory, though it is actually not often incorporated in detection frameworks (Figure 7a). On the other hand, hierarchical clustering and other popular clustering options like DBSCAN or OPTICS are more accurate and robust, but computationally costly and less flexible for stream data and evolving scenarios. Their incorporation into real, stand-alone detection systems is therefore difficult.



(**a**) Most used algorithm families.

(**b**) Claimed paper contribution.

Figure 6. Algorithm families used by research authors as the main options for NTA analysis and the contributions claimed by authors in the conclusions.



(**a**) Evaluation methods. (**b**) Reproducibility.

Figure 7. Methods for evaluating NTA proposals and the reproducibility of experiments and setups.

7.6. Claimed Contribution

Checking algorithm families in the backdrop of authors' proposals is relevant since the improvement of detection ratios is by far the main claimed contribution (Figure 6b). Beyond such achievement, other repeated claims are generally related to enhancing method feasibility, namely: reducing complexity, allowing fast processing, transportability, or addressing big data. However, low attention is given to two important requirements that NTA applications currently demand and prioritize, which are: analyzing encrypted communications and analyzing stream data. Data encryption is a strong limitation for modern NTA and makes most previous proposals almost unpractical. Encryption is progressively gaining attention in recent works, e.g., [102]. Furthermore, considering that research is mainly focused on the actual detection and not that much on knowledge discovery, obviating the temporal implications of real setups and facing analysis only from static perspectives are disturbing. Online detection and prompt reactions are principal demands in the NTA application, whereas forensic analysis plays a secondary role. Related research often proposes methods that do not necessarily distinguish between online and offline applications, and therefore, the derived limitations and peculiarities are omitted. Furthermore, as Figure 7a shows, implementations of the detection proposals in real systems are seldom undertaken.

7.7. Reproducibility

In [35], ACMpresents a terminology to define to what degree a research work can be repeated by using the information provided by the authors in the paper and in the linked resources. These categories are:

- Reproducible: Experiments are fully reproducible by a different team based on the information given in the paper. The setup, parameters, tools, and datasets are described and/or provided (references to valid links) in a clear and open way. Results are expected to be the same or very similar.
- Replicable: The experiment can be replicated by a different team, but with a different setup. The methodology is clearly explained, at least at a theoretical level. Not all parameters or tools are provided, but readers obtain enough know-how from the paper and references to develop their own setups based on the provided descriptions.
- Repeatable: The methodologies and setups are clearly described with scientific rigor; however, experiments can only be repeated by the authors given that some resources are not publicly available (e.g., they use datasets that are not openly available).

- No: Important information about the part of the methodology is missing in a way that the experiment cannot be repeated in comparable conditions. Papers show findings or results, but it is not clear how they were obtained (information is hidden, omitted, or simply missing).

Figure 7b shows that 43.3% of the papers were replicable, but only 10.4% were reproducible; while 29.9% were only repeatable, and the remaining 16.4% did not meet the minimum repeatability standards. Considering that the paper selection prioritizes most cited papers and papers published in reputed journals and conferences, there is much room for improving the reproducibility culture in the field. Even when considering completely reproducible works, experiments that can be repeated out-of-the-box are very scarce. One of the most noticeable impediments for reproducibility is the dataset availability (Figure 5a), which is not openly available in almost half of the cases. A second relevant reproducibility hindrance is the use of software, tools, and scripts for preprocessing and analysis that are not publicly available or even not specified in the paper (57.5% of the studied papers did not provide information about the tools used for preprocessing, while 53.8% did not provide information about the analysis tools). Not providing algorithm parameters and descriptions of analysis setups is also common.

8. Conclusions

NTARC is a data model for storing relevant information related to network traffic research. We widely described NTARC structures and introduced the tools developed for its creation, validation, sharing, and deployment. NTARC databases are expected to grow with the curation of new and old published papers, whereas NTARC structures are expected to be progressively refined with usage. Overall, NTARC is devised to improve how science is done in the field, and this is achieved by enhancing how research material and information is reused.

By using the "NTARC Database"—a release of NTARC objects containing the last years' principal field investigations—we reviewed the trends and characteristics of NTA research from a critical and systematic perspective. NTA is particularly focused on attack detection, anomaly detection, and traffic classification, and the standard profile for a research paper is the proposal of a method based on machine learning that claims to improve detection accuracy. However, as posed by Sommer and Paxon in [132], machine learning has been widely used for security and NTA research for the last few decades, but its presence in commercial and real-world solutions has been almost non-existent. This conclusion draws an incongruous picture in which research and application seem to live in distant worlds.

We also identified some undesired trends to avoid. Summarizing: (a) a lack of accurate descriptions of NTA problem spaces, (b) an insufficient discussion about the traffic classes aimed at, (c) obviating post-analysis, (d) inaccurate, vague, or undefined descriptions of aimed anomalies, (e) inappropriate, unrealistic data setups for unsupervised analysis, (e) the use of obsolete, irrelevant datasets, (f) monolithic approaches for too complex problems, (g) neglecting encryption and streaming characteristics, and (f) non-replicable experiments or non-public experimental resources.

Such undesired traits are partially caused by the limited access to valuable network data by researchers (especially to labeled data), also due to a lack of realistic test environments and methods for evaluating new proposals. NTA is therefore tackled under laboratory conditions that do not properly consider the constraints, peculiarities, and limitations of final implementations and might not cover practical requirements in many cases. As a consequence, the relevance and efficacy of expert research is severely reduced.

Author Contributions: Conceptualization, F.I. and D.C.F.; data curation, F.I., D.C.F., G.V., and M.B.; formal analysis, F.I.; methodology, F.I.; project administration, T.Z.; software, D.C.F. and G.V.; supervision, T.Z.; validation, D.C.F. and G.V.; writing, original draft, F.I.; writing, review and editing, D.C.F., T.Z., G.V., and M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the project MALware cOmmunication in cRitical Infrastructures (MALORI), funded by the Austrian security research program KIRAS of the Federal Ministry for Transport,

Innovation and Technology (BMVIT). The authors also acknowledge TU Wien University Library for financial support through its Open Access Funding Programme.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NTA	Network Traffic Analysis
NTARC	Network Traffic Analysis Research Curation
JSON	JavaScript Object Notation
XML	Extensible Markup Language
QoS	Quality of Service
RFC	Request for Comments
CMR	NASA's Common Metadata Repository
SOLE	Science Object Linking and Embedding
DCC	Digital Curation Center
NDSA	National Digital Stewardship Alliance
RDA	Research Data Alliance
CODATA	Committee on Data for Science and Technology
LODD	Linking Open Drug Data
W3C	World Wide Web Consortium
HCLSIG	Health Care and Life Sciences Interest Group
CAIDA	Center for Applied Internet Data Analysis
IMDC	Internet Measurement Data Catalog
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
API	Application Programming Interface
IDS	Intrusion Detection System
KDD	Knowledge Discovery in Databases
CIC	Canadian Institute for Cybersecurity
ISCX	Installation Support Center of Expertise
MAWI	Measurement and Analysis on the WIDE Internet
IANA	Internet Assigned Numbers Authority
TCP	Transmission Control Protocol
IP	Internet Protocol
IPFIX	IP Flow Information Export
PCA	Principal Component Analysis
SOM	Self-Organizing Maps
DBSCAN	Density-Based Spatial Clustering
OPTICS	Ordering Points to Identify the Clustering Structure

References

1. Li, B.; Springer, J.; Bebis, G.; Gunes, M.H. A survey of network flow applications. *J. Netw. Comput. Appl.* **2013**, *36*, 567–581.
2. Quittek, J.; Zseby, T.; Claise, B.; Zander, S. *Requirements for IP Flow Information Export (IPFIX)*; RFC 3917; IETF Network Working Group, The Internet Society, Reston, VA, USA, 2004.
3. Kim, H.; Claffy, K.; Fomenkov, M.; Barman, D.; Faloutsos, M.; Lee, K. Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices. In Proceedings of the 2008 ACM CoNEXT Conference, New York, NY, USA, 10–12 December 2008; pp. 11:1–11:12.
4. Ahmed, M.; Naser Mahmood, A.; Hu, J. A Survey of Network Anomaly Detection Techniques. *J. Netw. Comput. Appl.* **2016**, *60*, 19–31.
5. Callado, A.; Kamienski, C.; Szabo, G.; Gero, B.P.; Kelner, J.; Fernandes, S.; Sadok, D. A Survey on Internet Traffic Identification. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 37–52.

6. Bhuyan, M.H.; Bhattacharyya, D.K.; Kalita, J.K. Network anomaly detection: Methods, systems and tools. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 303–336.
7. Munafò, M.R.; Nosek, B.A.; Bishop, D.V.M.; Button, K.S.; Chambers, C.D.; Percie du Sert, N.; Simonsohn, U.; Wagenmakers, E.J.; Ware, J.J.; Ioannidis, J.P.A. A manifesto for reproducible science. *Nat. Hum. Behav.* **2017**, *1*, 1–9.
8. Ferreira, D.C. NTARC Database (GitHub). 2018. Available online: https://github.com/CN-TU/nta-meta-analysis (accessed on 25 April 2020) .
9. Ardestani, S.B.; Håkansson, C.J.; Laure, E.; Livenson, I.; Stranák, P.; Dima, E.; Blommesteijn, D.; van de Sanden, M. B2SHARE: An Open eScience Data Sharing Platform. In Proceedings of the 2015 IEEE 11th International Conference on e-Science, Munich, Germany, 31 August–4 September 2015; pp. 448–453.
10. Singh, J. FigShare. *J. Pharmacol. Pharmacother.* **2011**, *2*, 138–139.
11. Chard, K.; Pruyne, J.; Blaiszik, B.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. Globus Data Publication as a Service: Lowering Barriers to Reproducible Science. In Proceedings of the IEEE 11thInternational Conference on e-Science, Munich, Germany, 31 August–4 Septembe 2015; pp. 401–410.
12. TU Delft Library. 4TU.Centre for Research Data. 2017. Available online: https://data.4tu.nl/ (accessed on 25 April 2020).
13. CERN Data Centre and Invenio. Zenodo, 2013. Last Updated: July 2017. Available online: https://zenodo.org/ (accessed on 25 April 2020).
14. The Dataverse Network: An Open-source Application for Sharing, Discovering and Preserving Data. *D-Lib Mag.* **2011**, *17*, 2.
15. Greenberg, J.; White, H.C.; Carrier, S.; Scherle, R. A Metadata Best Practice for a Scientific Data Repository. *J. Libr. Metadata* **2009**, *9*, 194–212.
16. Assante, M.; Candela, L.; Castelli, D.; Tani, A. Are scientific data repositories coping with research data publishing? *Data Sci. J.* **2016**, *15*, 6.
17. EarthData-NASA. Common Metadata Repository (CMR), Earth Science Data & Information System Project (ESDIS), 2017. Last Updated: June 2017. Available online: https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository (accessed on 25 April 2020).
18. Devarakonda, R.; Palanisamy, G.; Green, J.M. Digitizing scientific data and data retrieval techniques. *arXiv* **2010**, arXiv:1010.3983v2.
19. Harrison, K.A.; Wright, D.G.; Trembath, P. Implementation of a workflow for publishing citeable environmental data: successes, challenges and opportunities from a data centre perspective. *Int. J. Digit. Libr.* **2017**, *18*, 133–143.
20. Bardi, A.; Manghi, P. Enhanced Publications: Data Models and Information Systems. *LIBER Q.* **2014**, *23*, 240–273.
21. Candela, L.; Castelli, D.; Manghi, P.; Tani, A. Data journals: A survey. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1747–1762.
22. Bardi, A.; Manghi, P. Enhanced Publication Management Systems: A Systemic Approach Towards Modern Scientific Communication. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May, 2015; pp. 1051–1052.
23. Pham, Q.; Malik, T.; Foster, I.; Di Lauro, R.; Montella, R., SOLE: Linking Research Papers with Science Objects. In *Provenance and Annotation of Data and Processes: 4th Int. Provenance and Annotation Workshop, IPAW*; Groth, P., Frew, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 203–208.
24. Group, N.P. About the *Scientific Data* journal, 2014. Available online: https://researchdata.springernature.com/users/69239-scientific-data (accessed on 25 April 2020).
25. Cragin, M.H.; Palmer, C.L.; Carlson, J.R.; Witt, M. Data sharing, small science and institutional repositories. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2010**, *368*, 4023–4038.
26. Council, N.R. *Preparing the Workforce for Digital Curation*; The National Academies Press: Washington, DC, USA, 2015.
27. Samwald, M.; Jentzsch, A.; Bouton, C.; Kallesøe, C.S.; Willighagen, E.; Hajagos, J.; Marshall, M.S.; Prud'hommeaux, E.; Hassanzadeh, O.; Pichler, E.; Stephens, S. Linked open drug data for pharmaceutical research and development. *J. Cheminform.* **2011**, *3*, 19.
28. Shannon, C.; Moore, D.; Keys, K.; Fomenkov, M.; Huffaker, B.; claffy, k. The Internet Measurement Data Catalog. *SIGCOMM Comput. Commun. Rev.* **2005**, *35*, 97–100.

29. CAIDA (Center for Applied Internet Data Analysis). Internet Traffic Classification, 2015. Last Updated: May, 2015. Available online: http://www.caida.org/research/traffic-analysis/classification-overview/ (accessed on 20 February 2020).
30. IMPACT. Information Marketplace for Policy and Analysis of Cyber-risk & Trust, 2017. Available online: https://www.impactcybertrust.org/ (accessed on 25 April 2020).
31. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. *Introduction to Meta-Analysis*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA,2009; pp. 409–414.
32. Haidich, A.B. Meta-analysis in medical research. *Hippokratia* **2010**, *14*, 29.
33. Bray, T. *RFC 7159: The JavaScript Object Notation (JSON) Data Interchange Format*; Technical Report; Internet Engineering Task Force (IETF): Fremont, CA, USA, 2014.
34. Ferreira, D.C.; Iglesias, F.; Vormayr, G.; Bachl, M.; Zseby, T. A Meta-Analysis Approach for Feature Selection in Network Traffic Research. In Proceedings of the Reproducibility Workshop, Los Angeles, CA, USA, 21–25 August 2020. ACM: New York, NY, USA, 2017; pp. 17–20.
35. Association for Computing Machinery (ACM). ACM Result and Artifact Review and Badging Publication Policy. 2017. Available online: https://www.acm.org/publications/policies/artifact-review-badging (accessed on 25 April 2020).
36. Ferreira, D.C.; Bachl, M.; Vormayr, G.; Iglesias, F.; Zseby, T. NTARC Specification (Version v3.0.0), 12 November 2018. doi:10.5281/ zenodo.1484190. Available online: http://doi.org/10.5281/zenodo.1484190 (accessed on 25 April 2020).
37. Vormayr, G. Editor for the NTARC data format (Version v3.1.6), 28 November 2018. doi:10.5281/zenodo.1243058. Available online: http://doi.org/10.5281/zenodo.1625380 (accessed on 25 April 2020).
38. Wright, A.; Andrews, H. JSON Schema: A Media Type for Describing JSON Documents. Internet Engineering Task Force, IETF Secretariat (Internet Draft), 19 March 2018. Available online: https://json-schema.org/draft-07/json-schema-core.html (accessed on 25 April 2020).
39. Electron. Available online: https://electronjs.org (accessed on 2 February 2018).
40. Ferreira, D.C. NTARC Verification Tool (Github), 2018. Available online: https://github.com/CN-TU/nta-meta-analysis-verification (accessed on 25 April 2020).
41. Sinha, A.; Shen, Z.; Song, Y.; Ma, H.; Eide, D.; Hsu, B.J.P.; Wang, K. An Overview of Microsoft Academic Service (MAS) and Applications. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 243–246.
42. Ferreira, D.C. NTARC Extractor Library (Github), 2018. Available online: https://github.com/CN-TU/nta-meta-analysis-library (accessed on 25 April 2020).
43. Ferreira, D.C.; Bachl, M.; Vormayr, G.; Iglesias, F.; Zseby, T. Curated Research on Network Traffic Analysis (Version 2020.2) [Data set], 10 February 2020. doi:10.5281/zenodo.3661423. Available online: http://doi.org/10.5281/zenodo.3661423 (accessed on 25 April 2020).
44. Barford, P.; Kline, J.; Plonka, D.; Ron, A. A Signal Analysis of Network Traffic Anomalies. In Proceedings of the ACM SIGCOMM Workshop on Internet Measurement, Marseille, France, 6–8 November 2002; pp. 71–82.
45. Mahoney, M.V.; Chan, P.K. Learning rules for anomaly detection of hostile network traffic. In Proceedings of the 3rd IEEE International Conference on Data Mining, Melbourne, FL, USA, 22 November 2003; pp. 601–604.
46. Lakhina, A.; Crovella, M.; Diot, C. Characterization of Network-Wide Anomalies in Traffic Flows. In Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, Taormina, Sicily, Italy, 25–27 October 2004; pp. 201–206.
47. Lakhina, A.; Crovella, M.; Diot, C. Diagnosing Network-Wide Traffic Anomalies. In Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '04), Portland, Oregon, USA, August 30 – September 3 2004; pp. 219–230.
48. Wang, K.; Stolfo, S.J. Anomalous Payload-based Network Intrusion Detection. In Proceedings of the International Workshop on Recent Advances in Intrusion Detection, French Riviera, France, 15–17 September 2004.
49. Gu, Y.; McCallum, A.; Towsley, D. Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation. In Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, Berkeley, CA, USA, 19–21 October 2005; pp. 345–350.

50. Karagiannis, T.; Papagiannaki, K.; Faloutsos, M. BLINC: Multilevel Traffic Classification in the Dark. In Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Philadelphia, PA, USA, 22–26 August 2005; pp. 229–240.
51. Lakhina, A.; Crovella, M.; Diot, C. Mining Anomalies Using Traffic Feature Distributions. In Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Philadelphia, PA, USA, 22–26 August 2005; Volume 35, pp. 217–228.
52. Moore, A.W.; Zuev, D. Internet Traffic Classification Using Bayesian Analysis Techniques. In Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Banff, AB, Canada, 6–10 June 2005; Volume 33, pp. 50–60.
53. Thottan, M.; Ji, C. Anomaly Detection in IP Networks. *IEEE Trans. Signal Process.* **2005**, *51*, 2191–2204.
54. Williams, N.; Zander, S.; Armitage, G. A preliminary performance comparison of five machine learning algorithms for practical ip traffic flow classification. *ACM SIGCOMM Comput. Commun. Rev.* **2006**, *36*, 5–16.
55. wright, c.; monrose, f.; masson, g. on inferring application protocol behaviors in encrypted network traffic. *J. Mach. Learn. Res.* **2006**, *7*, 2745–2769.
56. Auld, T.; Moore, A.W.; Gull, S.F. Bayesian Neural Networks for Internet Traffic Classification. *IEEE Trans. Neural Netw.* **2007**, *18*, 223–239.
57. Crotti, M.; Dusi, M.; Gringoli, F.; Salgarelli, L. Traffic Classification through Simple Statistical Fingerprinting. *ACM SIGCOMM Comput. Commun. Rev.* **2007**, *37*, 7–16.
58. Erman, J.; Mahanti, A.; Arlitt, M.; Williamson, C. Identifying and Discriminating Between Web and Peer-to-Peer Traffic in the Network Core. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; Volume 16, pp. 883–892.
59. Erman, J.; Mahanti, A.; Arlitt, M.; Cohen, I.; Williamson, C. Offline/realtime traffic classification using semi-supervised learning. *Perform. Eval.* **2007**. *64*, 1194–1213.
60. Liu, Y.; Li, W.; Li, Y. Network Traffic Classification Using K-means Clustering. In Proceedings of the Second International Multisymposium on Computer and Computational Sciences, Iowa City, IA, USA, 13–15 August 2007; Volume 1, pp. 360–365.
61. Ringberg, H.; Soule, A.; Rexford, J.; Diot, C. Sensitivity of PCA for traffic anomaly detection. In Proceedings of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, San Diego, CA, USA, 12–17 June 2007; Volume 35, pp. 109–120.
62. Dainotti, A.; De Donato, W.; Pescape, A.; Rossi, P.S. Classification of Network Traffic via Packet-Level Hidden Markov Models. In Proceedings of the IEEE GLOBECOM—Global Telecommunications Conference, New Orleans, LA, USA, 30 November–4 December 2008; pp. 1–5.
63. Gu, G.; Perdisci, R.; Zhang, J.; Lee, W. BotMiner: clustering analysis of network traffic for protocol- and structure-independent botnet detection. In Proceedings of the 17th Conference on Security Symposium USENIX, San Jose, CA, USA, 28 July–1 August 2008; pp. 139–154.
64. Nychis, G.; Sekar, V.; Andersen, D.G.; Kim, H.; Zhang, H. An empirical evaluation of entropy-based traffic anomaly detection. In Proceedings of the ACM SIGCOMM Conference on Internet Measurement, Vouliagmeni, Greece 20–22 October 2008.
65. Yang, A.m.; Jiang, S.y.; Deng, H. A P2P Network Traffic Classification Method Using SVM. In Proceedings of the International Conference for Young Computer Scientists, Hunan, China, 18–21 November 2008; pp. 398–403.
66. Zhao, J.; Huang, X.; Sun, Q.; Ma, Y. Real-time feature selection in traffic classification. *J. China Univ. Posts Telecomm.* **2008**, *15*, 68–72.
67. Alshammari, R.; Zincir-Heywood, A.N. Machine Learning Based Encrypted Traffic Classification: Identifying SSH and Skype. In Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), Ottawa, Canada, 8–10 July 2009; pp. 1–8.
68. Este, A.; Gringoli, F.; Salgarelli, L. Support vector Machines for TCP traffic classification. *Comput. Netw.* **2009**, *53*, 2476–2490.
69. Kind, A.; Stoecklin, M.P.; Dimitropoulos, X. Histogram-based traffic anomaly detection. *IEEE Trans. Netw. Serv. Manag.* **2009**, *6*, 110–121.
70. Zhani, M.F.; Elbiaze, H.; Kamoun, F. Analysis and Prediction of Real Network Traffic. *JNW* **2009**, *4*, 855–865.
71. Dewaele, G.; Himura, Y.; Borgnat, P.; Fukuda, K.; Abry, P.; Michel, O.; Fontugne, R.; Cho, K.; Esaki, H. Unsupervised host behavior classification from connection patterns. *Int. J. Netw. Manag.* **2010**, *20*, 317–337.

72. Lim, Y.; Kim, H.; Jeong, J.; Kim, C.; Kwon, T.T.; Choi, Y. Internet Traffic Classification Demystified: On the Sources of the Discriminative Power. In Proceedings of the 6th International Conferenceon Co-NEXT, Philadelphia, PA, USA, 30 November–3 December 2010.
73. Shrivastav, A.; Tiwari, A. Network Traffic Classification using Semi-Supervised Approach. In Proceedings of the IEEE International Conference on Machine Learning and Computing (ICMLC), Qingdao, China, 11–14 July 2010; pp. 345–349.
74. Zeidanloo, H.R.; Manaf, A.B.; Vahdani, P.; Tabatabaei, F.; Zamani, M. Botnet Detection Based on Traffic Monitoring. In Proceedings of the International Conference on Networking and Information Technology, Bradford, UK, 29 June–1 July 2010; pp. 97–101.
75. Amiri, F.; Yousefi, M.R.; Lucas, C.; Shakery, A.; Yazdani, N. Mutual information-based feature selection for intrusion detection systems. *J. Netw. Comput. Appl.* **2011**, *34*, 1184–1199.
76. Agarwal, B.; Mittal, N. Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques. *Procedia Technol.* **2012**, *6*, 996–1003.
77. Bujlow, T.; Riaz, T.; Pedersen, J.M. A method for classification of network traffic based on C5.0 Machine Learning Algorithm. International Conference on Computing, Networking and Communications (ICNC), Maui, HI, USA, 30 January–2 February 2012; pp. 237–241.
78. Catania, C.A.; Bromberg, F.; Garino, C.G. An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection. *Expert Syst. Appl.* **2012**, *39*, 1822–1829.
79. Grimaudo, L.; Mellia, M.; Baralis, E. Hierarchical learning for fine grained internet traffic classification. In Proceedings of the 8th International Wireless Communications and Mobile Computing Conference (IWCMC), Limassol, Cyprus, 27–31 August 2012; pp. 463–468.
80. Jin, Y.; Duffield, N.; Jeffrey, E.; Haffner, P.; Sen, S.; Zhang, Z.L. A Modular Machine Learning System for Flow-Level Traffic Classification in Large Networks. *ACM Trans. Knowl. Discov. Data* **2012**, *6*, 1–34.
81. Nguyen, T.; Armitage, G.; Branch, P.; Zander, S. Timely and Continuous Machine-Learning-Based Classification for Interactive IP Traffic. *IEEE/ACM Trans. Netw. (TON)* **2012**, *20*, 1880–1894.
82. Yin, C.; Li, S.; Li, Q. Network traffic classification via HMM under the guidance of syntactic structure. *Comput. Netw.* **2012**, *56*, 1814–1825.
83. Zargari, S.; Voorhis, D. Feature Selection in the Corrected KDD-dataset. In Proceedings of the 3rd International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), Bucharest, Romania, 19–21 September 2012; pp. 174–180.
84. Zhang, H.; Lu, G.; Qassrawi, M.T.; Zhang, Y.; Yu, X. Feature selection for optimizing traffic classification. *Comput. Commun.* **2012**, *35*, 1457–1471.
85. Zhang, J.; Xiang, Y.; Zhou, W.; Wang, Y. Unsupervised traffic classification using flow statistical properties and IP packet payload. *J. Comput. Syst. Sci.* **2012**, *79*, 573–585.
86. Comar, P.M.; Liu, L.; Saha, S.; Tan, P.N.; Nucci, A. Combining supervised and unsupervised learning for zero-day malware detection. In Proceedings of the IEEE INFOCOM, Turin, Italy, 14–19 April 2013; pp. 2022–2030.
87. Fiore, U.; Palmieri, F.; Castiglione, A.; De Santis, A. Network anomaly detection with the restricted Boltzmann machine. *Neurocomputing* **2013**, *122*, 13–23.
88. Huang, S.Y.; Huang, Y. Network forensic analysis using growing hierarchical SOM. In Proceedings of the International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 536–543.
89. Jadidi, Z.; Sheikhan, M. Flow-Based Anomaly Detection Using Neural Network Optimized with GSA Algorithm. In Proceedings of the 33rd International Conference on Distributed Computing Systems Workshops, Philadelphia, PA, USA, 8–11 July 2013; pp. 76–81.
90. Zhang, F.; Wang, D. An effective feature selection approach for network intrusion detection. In Proceedings of the IEEE 8th International Conference on Networking, Architecture and Storage (NAS), Shaanxi, China, 17–19 July 2013; pp. 307–311.
91. Zhang, J.; Chen, C.; Xiang, Y.; Zhou, W.; Vasilakos, A.V. An Effective Network Traffic Classification Method with Unknown Flow Detection. *IEEE Trans. Netw. Serv. Manag.* **2013**, *10*, 133–147.
92. Zhang, J.; Xiang, Y.; Wang, Y.; Zhou, W.; Xiang, Y.; Guan, Y. Network Traffic Classification Using Correlation Information. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 104–117.

93. Zhang, J.; Chen, C.; Xiang, Y.; Zhou, W. Robust network traffic identification with unknown applications. In Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security, Berlin, Germany, 4–8 November 2013; pp. 405–414.
94. Jun, J.H.; Ahn, C.W.; Kim, S.H. DDoS attack detection by using packet sampling and flow features. In Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Korea, 24–28 March 2014; pp. 711–712.
95. Ma, X.; Chen, Y. DDoS Detection Method Based on Chaos Analysis of Network Traffic Entropy. *IEEE Comm. Lett.* **2014**, *18*, 114–117.
96. Singh, K.; Guntuku, S.C.; Thakur, A.; Hota, C. Big data analytics framework for peer-to-peer botnet detection using random forests. *Inf. Sci.* **2014**, *278*, 488–497.
97. Qin, X.; Xu, T.; Wang, C. DDoS attack detection using flow entropy and clustering technique. In Proceedings of the 11th International Conference on Computational Intelligence and Security (CIS), Angkor Wat, Cambodia, 15–17 July 2015; pp. 412–415.
98. Singh, R.; Kumar, H.; Singla, R. An intrusion detection system using network traffic profiling and online sequential extreme learning machine. *Expert Syst. Appl.* **2015**, *42*, 8609–8624.
99. van der Toorn, O.; Hofstede, R.; Jonker, M.; Sperotto, A. A first look at HTTP(S) intrusion detection using NetFlow/IPFIX. In Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM), Otawa, Canada, 11–15 May 2015; pp. 862–865.
100. Zhang, J.; Chen, X.; Xiang, Y.; Zhou, W.; Wu, J. Robust network traffic classification. *IEEE/ACM Trans. Netw. (TON)* **2015**, *23*, 1257–1270.
101. Ambusaidi, M.A.; He, X.; Nanda, P.; Tan, Z. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. *IEEE Trans. Comput.* **2016**, *65*, 2986–2998.
102. Anderson, B.; McGrew, D. identifying encrypted malware traffic with contextual flow data. In Proceedings of the ACM Workshop on Artificial Intelligence and Security, Vienna, Austria, 28 October 2016; pp. 35–46.
103. Gharaee, H.; Hosseinvand, H. A new feature selection IDS based on genetic algorithm and SVM. In Proceedings of the 8th International Symposium on Telecomm (IST), Tehran, Iran, 27–29 September 2016; pp. 139–144.
104. Iglesias, F.; Zseby, T. Time-activity footprints in IP traffic. *Comput. Netw.* **2016**, *107, Pt 1*, 64–75.
105. Javaid, A.; Niyaz, Q.; Sun, W.; Alam, M. A Deep Learning Approach for Network Intrusion Detection System. In Proceedings of the 9th EAI Int. Conf. on Bio-inspired Information and Communications Technologies ICST, New York, NY, USA, 3–5 December 2016; pp. 21–26.
106. Mishra, P.; Pilli, E.S.; Varadharajant, V.; Tupakula, U. NvCloudIDS: A security architecture to detect intrusions at network and virtualization layer in cloud environment. In Proceedings of the Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI), Cebu, PA, USA, 27–29 May 2016; pp. 56–62.
107. Al-Zewairi, M.; Almajali, S.; Awajan, A. Experimental Evaluation of a Multi-layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System. In Proceedings of the Int. Conf. on New Trends in Computing Sciences (ICTCS), Amman, Jordan, 11–13 October 2017; pp. 167–172.
108. Anderson, B.; McGrew, D. Machine learning for encrypted malware traffic classification: accounting for noisy labels and non-stationarity. In Proceedings of the ACM SIGKDD Int Conf. on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1723–1732.
109. Ashfaq, w.R.A.R.; Wang, X.Z.; Huang, J.Z.; Abbas, H.; He, Y.L. Fuzziness based semi-supervised learning approach for intrusion detection system. *Inf. Sci.* **2017**, *378*, 484–497.
110. Baig, M.M.; Awais, M.M.; El-Alfy, E.S.M. A multiclass cascade of artificial neural network for network intrusion detection. *J. Intell. Fuzzy Syst.* **2017**, *32*, 2875–2883.
111. Bamakan, S.M.H.; Wang, H.; Shi, Y. Ramp loss K-Support Vector Classification-Regression; a robust and sparse multi-class approach to the intrusion detection problem. *Knowl.-Based Syst.* **2017**, *126*, 113–126.
112. Iglesias, F.; Zseby, T. Pattern Discovery in Internet Background Radiation. *IEEE Trans. Big Data* **2019**, *5*, 467–480.
113. Taylor, V.F.; Spolaor, R.; Conti, M.; Martinovic, I. Robust Smartphone App Identification Via Encrypted Network Traffic Analysis. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 63–78.
114. Vlăduţu, A.; Comăneci, D.; Dobre, C. Internet traffic classification based on flows' statistical properties with machine learning. *Int. J. Netw. Manag.* **2017**, *27*.

115. Mirsky, Y.; Doitshman, T.; Elocivi, Y.; Shabtai, A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. In Proceedings of the Network and Distributed System Security Symposium–NDSS 2018, San Diego, CA, USA, 18–21 February 2018.
116. Potluri, S.; Diedrich, C. Accelerated deep neural networks for enhanced Intrusion Detection System. In Proceedings of the IEEE 21st Int. Conf. on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6–9 September 2016; pp. 1–8.
117. Bhuyan, M.H.; Bhattacharyya, D.; Kalita, J. A multi-step outlier-based anomaly detection approach to network-wide traffic. *Inf. Sci.* **2016**, *348*, 243 – 271.
118. Wright, C.; Monrose, F.; Masson, G.M. HMM Profiles for Network Traffic Classification. In Proceedings of the ACM Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC), Washington, DC, USA, 29 October 2004; pp. 9–15.
119. Vellido, A.; Martín-Guerrero, J.D.; Rossi, F.; Lisboa, P.J.G. Seeing is believing: The importance of visualization in real-world machine learning applications. In Proceedings of the ESANN 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, 27–29 April 2011.
120. Pang, R.; Yegneswaran, V.; Barford, P.; Paxson, V.; Peterson, L. Characteristics of Internet Background Radiation. In Proceedings of the 4th ACM SIGCOMM Conf. on Internet Measurement, Taormina, Sicily, Italy, 25–27 October 2004; p. 27–40.
121. Axelsson, S. The Base-Rate Fallacy and Its Implications for the Difficulty of Intrusion Detection. In Proceedings of the 6th ACM Conference on Computer and Communications Security, Singapore, 2–4 November 1999; 1–7.
122. Claffy, K. The Inevitable Conflict between Data Privacy and Science, 2009. Visited on April 2020. Available online: https://blog.caida.org/best_available_data/2009/01/04/the-inevitable-conflict-between-data-privacy-and-data-utility-revisited/ (accessed on 25 April 2020).
123. Kenneally, E.; Claffy, K. Dialing privacy and utility: a proposed data sharing framework to advance Internet research . *IEEE Secur. Priv.* **2010**, *8*, 31–39.
124. CAIDA. Data—Overview of Datasets, Monitors, and Reports, 2020. Available online: https://www.caida.org/data/overview/ (accessed on 25 April 2020).
125. MAWI Working Group. Packet Traces from WIDE backbone, 2020. Available online: http://mawi.wide.ad.jp/mawi/ (accessed on 25 April 2020).
126. Canadian Institute for Cybersecurity. Datasets, 2020. Available online: https://www.unb.ca/cic/datasets/index.html (accessed on 25 April 2020).
127. Lim, Y.s.; Kim, H.c.; Jeong, J.; Kim, C.k.; Kwon, T.T.; Choi, Y. Internet Traffic Classification Demystified: On the Sources of the Discriminative Power. In Proceedings of the 6th International COnference, Philadelphia, USA, 30 November–3 December 2010; ACM: New York, NY, USA, 2010; Co-NEXT, pp. 9:1–9:12.
128. Iglesias, F.; Zseby, T. Analysis of network traffic features for anomaly detection. *Mach. Learn.* **2015**, *101*, 59–84.
129. Claise, B.; Trammell, B. *RFC 7012: Information Model for IP Flow Information Export (IPFIX)*; Technical Report; Internet Engineering Task Force (IETF), Fremont, CA, USA, 2013.
130. Meghdouri, F.; Zseby, T.; Iglesias, F. Analysis of Lightweight Feature Vectors for Attack Detection in Network Traffic. *Appl. Sci.* **2018**, *8*, 2196.
131. Dainotti, A.; Pescape, A.; Claffy, K.C. Issues and future directions in traffic classification. *IEEE Netw.* **2012**, *26*, 35–40.
132. Sommer, R.; Paxson, V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In Proceedings of the IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 22–25 May 2010; pp. 305–316.