COLLEGES BRETAGNE LOIRE | SCIENCES POUR L'INGENIEUR ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE
PAYS DE LA LOIRE – IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le Numérique
Spécialité : *Mathématiques et Sciences et Technologies de l'Information et de la Communication*

Par
Léo LAVAUR

L'Apprentissage Fédéré comme Outil pour la Détection Collaborative d'Intrusions

Thèse présentée et soutenue à Rennes, le XX septembre 2024
Unité de recherche : IRISA (UMR 6074), SOTERN

Rapporteurs avant soutenance :

Anne-Marie Kermarrec	Professeure à l'Université Polytechnique Fédérales de Lausanne (EPFL)
Éric Totel	Professeur à Télécom SudParis

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer quelle est conforme et devra être répercutée sur la couverture de thèse

Président :	*À compléter une fois élu.*	
Examinateurs :	Sonia Ben Mokhtar	Directrice de Recherche CNRS au laboratoire LIRIS
	Pierre-François Gimenez	Maître de Conférence à CentraleSupelec
	Vincent Nicomette	Professeur à l'INSA de Toulouse
	Fabien AUTREL	Ingénieur de Recherche à IMT Atlantique
	Marc-Oliver PAHL	Directeur d'Étude à IMT Atlantique
Dir. de thèse :	Yann BUSNEL	Directeur de la Recherche et de l'Innovation (DRI) à IMT Nord Europe

Invité(s) :

Prénom NOM	Fonction et établissement d'exercice

Abstracts

Résumé

La collaboration entre les différents acteurs de la cybersécurité est essentielle pour lutter contre des attaques de plus en plus sophistiquées et nombreuses. Pourtant, les organisations sont souvent réticentes à partager leurs données, par peur de compromettre leur confidentialité, et ce même si cela pourrait d'améliorer leurs modèles de détection d'intrusions. L'apprentissage fédéré est un paradigme récent en apprentissage automatique qui permet à des clients distribués d'entraîner un modèle commun sans partager leurs données. Ces propriétés de collaboration et de confidentialité en font un candidat idéal pour des applications sensibles comme la détection d'intrusions. Si un certain nombre d'applications ont montré qu'il est, en effet, possible d'entraîner un modèle unique sur des données distribuées de détection d'intrusions, peu se sont intéressées à l'aspect collaboratif de ce paradigme. En plus de l'aspect collaboratif, d'autres problématiques apparaissent dans ce contexte, telles que l'hétérogénéité des données des différents participants ou la gestion de participants non fiables. Dans ce manuscrit, nous explorons l'utilisation de l'apprentissage fédéré pour construire des systèmes collaboratifs de détection d'intrusions. En particulier, nous explorons l'impact de la qualité des données dans des contextes hétérogènes, certains types d'attaques par empoisonnement, et proposons des outils et des méthodologies pour améliorer l'évaluation de ce type d'algorithmes distribués.

Abstract

Collaboration between different cybersecurity actors is essential to fight against increasingly sophisticated and numerous attacks. However, stakeholders are often reluctant to share their data, fearing confidentiality and privacy issues, although it would improve their intrusion detection models. Federated learning is a recent paradigm in machine learning that allows distributed clients to train a common model without sharing their data. These properties of collaboration and confidentiality make it an ideal candidate for sensitive applications such as intrusion detection. While several applications have shown that it is indeed possible to train a single model on distributed intrusion detection data, few have focused on the collaborative aspect of this paradigm. In addition to the collaborative aspect, other challenges arise in this context, such as the heterogeneity of the data between different participants or the management of untrusted contributions. In this manuscript, we explore the use of federated learning to build collaborative intrusion detection systems. In particular, we explore the impact of data quality in heterogeneous contexts, some types of poisoning attacks, and propose tools and methodologies to improve the evaluation of these types of distributed algorithms.

ACKNOWLEDGEMENTS

Table of Contents

Chapter 1

Introduction

Contents

1.1 Context and Motivation

Modern information security is made difficult by the scale, complexity, and heterogeneity of information systems. Because security by design in these conditions is an impossible task, security agencies also recommend complementary measures. For instance, the NIST Cybersecurity Framework [Nat24] suggests a five-stage lifecycle for managing risks in information systems: identify, protect, detect, respond, and recover.

Detection and response immensely benefit from the recent advances in Artificial Intelligence (AI) and Machine Learning (ML), enabling the analysis of more complex behaviors. Yet, because organizations usually face similar threats, including large-scale campaigns such as Mirai in 2016 or NotPetya in 2017, they would greatly benefit from sharing insights on the intrusions they have encountered, or any knowledge that might help others to identify the incident before the damages are too important. Collaboration is further encouraged by regulation, for instance with the NIS [16] and NIS2 [22] European directives. Sharing data is made even more important for training ML and Deep Learning (DL) models, which require large amounts of data to be effective. Yet, stakeholders are often reluctant to involve their organization in data-sharing practices, fearing confidentiality and privacy breaches, reputation loss, or regulation non-compliance.

Federated Learning (FL) [McM+17] has emerged as a promising paradigm for collaborative ML, enabling model training across distributed data sources while preserving privacy. Deployed in intrusion detection contexts, FL can help organizations to virtually extend the size of their training sets, thus producing more accurate models. This architecture could also be used to disseminate information about esoteric attacks or devices behavior owned locally, that would benefit to other organizations. FL also promises to solve other drawbacks of ML-based Intrusion Detection Systems (IDSs), such as the need for continuous retraining[?], the lack of adaptability to new threats[?], or the risk of local

biases due to a lack of heterogeneity in the training set[**?**].

Consequently, applying FL to IDS seems like a promising approach to collaboratively improve the local detection of cyber threats. This is supported by the amount of recent literature on the topic, which has grown exponentially since 2018 [Lav+22b; Ism+24]. Yet, novel challenges arise in this context, such as how to handle the heterogeneity of data sources or how to deal with untrusted participants. But more importantly, *what makes applying FL to IDS different from other applications? And is FL even a suitable framework for collaborative IDS?*

This dissertation aims to investigate the potential of Federated Learning as a collaborative framework for Intrusion Detection System, which we will refer to as Federated Intrusion Detection System (FIDS). The remaining of this manuscript will discuss the state of the art in FL and IDS, some of the challenges that arise in this context, and the potential solutions to address them.

1.1.1 Use case boundaries

While applying FL to IDS can already be considered as a restricted scope, the IDS literature contains a wide variety of use cases, each coming with its own set of specificities and constraints. For instance, IDSs can be deployed at the network level, the host level, or the application level. Likewise, objectives and constraints may vary depending on the context and the type of devices involved: Internet of Things (IoT), Industrial Control System (ICS), or traditional information systems. Among the most common combinations, Network-based Intrusion Detection System (NIDS) on Information Technology (IT) network data stands out, notably in terms of implemented algorithms and available datasets. This is particularly important for evaluation purposes, as it makes it easier to compare the performance of different approaches.

Consequently, this dissertation will focus on the use case of building collaborative NIDSs by leveraging FL on IT network data. In this context, FL clients are assumed to be organizations that own or oversee an information system, and that are interested in improving their local IDS by sharing information with other organizations. However, the results of these works could theoretically be extended to other use cases. Figure 1.1 illustrates this use case.

1.1.2 Research Objectives

Based on the context and motivation laid out in the previous section, we formulize the general objectives of this dissertation as a set of research questions. The questions stated hereafter are intended to be completed and extended in the following chapters, some of which introduce their own research questions. Overall, this work aims to answer

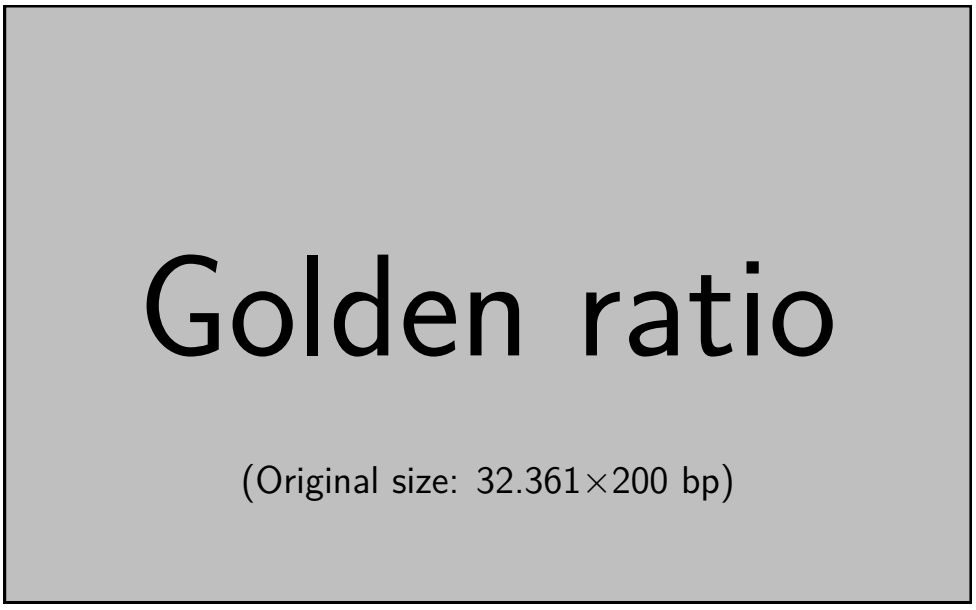


Figure 1.1 – Illustration of FL in a Collaborative IDS (CIDS) use case.

the following: *Can FL serve as a trustable knowledge-sharing framework for collaboratively improving intrusion detection mechanisms?*

Specifically, we focus on the following research questions:

RQ1. *What makes applying FL to IDSs specific?*

RQ2. *Can FL be used to federate IDSs across heterogeneous data sources?*

RQ3. *How does FL handle malicious contributions in a federated IDS?*

RQ4. *How can we address the main challenges in applying FL to IDS?*

1.2 Contributions

We summarize the contributions of this dissertation as follows:

1. The first Systematic Literature Review (SLR) on applying FL to IDS, with quantitative and qualitative analyses of the existing works, as well as the proposal of a reference architecture and a taxonomy for structuring the domain.
2. A demonstration highlighting the challenges of heterogeneity and malicious contributions in FIDS.
3. An extensible evaluation framework for FIDSs called Eiffel, leveraging popular open source libraries like Flower [Beu+20] and Hydra [Yad19], and a set of malicious clients simulators.
4. A systematic analysis of the impact of label-flipping attacks on an FL-based collaborative IDS, leveraging the aforementioned evaluation framework.

5. The first FL architecture for collaborative IDS that handles malicious contributions in heterogeneous environments, leveraging a cross-evaluation mechanism and a reputation system.
6. A methodology allowing to generate network topologies with heterogeneity constraints, and laying down the foundations toward a more realistic evaluation of FIDS and distributed networking telemetry experiments in general.

1.3 Outline

Outside of the introduction and conclusion, the manuscript is organized in three parts: defining FIDSs, quantifying their limitations, and providing solutions to address them.

Part I: The first part delves into the application of FL to IDS. After layout out the necessary background in Chapter 2, we present the state of the art in FIDS in Chapter 3. This chapter notably presents the results of our SLR on the topic, and focus on the related challenges and research opportunities. Chapter 4 then closes this first part by highlighting the main challenges in FIDS using toy examples.

Part II: The second part presents our contributions to quantifying the limitations of FIDS. Chapter 5 introduces a practical method to generate network topologies based on the composition of sub-topologies, and lays down the foundations for further studies on distributed networking analyses. Finally, Chapter 6 introduces our evaluation framework, and systematically analyses the impact of label-flipping attacks on FIDS.

Part III: The last part focuses on providing solutions to the challenges studied in Part II. Notably, Chapter 7 introduces a novel FL architecture for FIDS that handles malicious contributions in heterogeneous environments. Chapter 8 then makes a statement on the future of FIDS, with discussions about open issues and potential research directions.

1.4 Publications

Journal articles

[Lav+22b] Léo Lavaur, Marc-Oliver Pahl, *et al.*, « The Evolution of Federated Learning-based Intrusion Detection and Mitigation: A Survey », *in*: *IEEE Transactions on Network and Service Management*, Special Issue on Network Security Management (June 2022).

International conference papers

[Lav+ew] Léo Lavaur, Pierre-Marie Lechevalier, Yann Busnel, Romaric Ludinard, *et al.*, « RADAR: Model Quality Assessment for Reputation-aware Collaborative Federated Learning », *in*: Under Review.

[LBA24a] Léo Lavaur, Yann Busnel, and Fabien Autrel, « Demo: Highlighting the Limits of Federated Learning in Intrusion Detection », *in*: *Proceedings of the 44th International Conference on Distributed Computing Systems (ICDCS)*, July 2024.

[LBA24b] Léo Lavaur, Yann Busnel, and Fabien Autrel, « Systematic Analysis of Label-flipping Attacks against Federated Learning in Collaborative Intrusion Detection Systems », *in*: *Proceedings of the 19th International Conference on Availability, Reliability and Security (ARES), Workshop on Behavioral Authentication for System Security*, Aug. 2024.

National conference papers

[Lav+21] Leo Lavaur, Marc-Oliver Pahl, *et al.*, « Federated Security Approaches for IT and OT », *in*: *Journée thématique du GT sur la Sécurité des Systèmes, Logiciels et Réseaux (GT-SSLR)*, May 2021.

[Lav+22a] Leo Lavaur, Benjamin Coste, *et al.*, « Federated Learning as Enabler for Collaborative Security between Not Fully-Trusting Distributed Parties », *in*: *Proceedings of the 29th Computer & Electronics Security Application Rendezvous (C&ESAR), Ensuring Trust in a Decentralized World*, Oct. 2022.

[Lav+23] Léo Lavaur, Pierre-Marie Lechevalier, Yann Busnel, Marc-Oliver Pahl, *et al.*, « Metrics and Strategies for Adversarial Mitigation in Federated Learning-based Intrusion Detection », *in*: *Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information (RESSI)*, May 2023.

Tutorials

[BL23] Yann Busnel and Léo Lavaur, « Federated Learning × Security for Network Management », 15th International Conference on Network of the Future (NoF), Sept. 2023.

[BL24] Yann Busnel and Léo Lavaur, « Tutorial: Federated Learning × Security for Network Monitoring », *in*: *Proceedings of the 44th International Conference on Distributed Computing Systems (ICDCS)*, July 2024.

Part I

Federated Learning to build CIDSs

Chapter 2

Preliminaries

Contents

This chapter provides the necessary background on Machine Learning (ML) for intrusion detection, the inner of Federated Learning (FL), and the related threats.

2.1 Machine Learning for Intrusion Detection

2.2 Fundamentals of Federated Learning

2.3 Threats against Federated Learning

Chapter 3

State of the Art

Contents

3.1 Introduction and Motivation

In the previous chapter, we introduced the concepts of Intrusion Detection System (IDS) and Machine Learning (ML), the challenges of deploying Collaborative IDSs (CIDSs), and why Federated Learning (FL) is a promising solution to these challenges. This chapter's prime objective is to provide a comprehensive review of how FL can be leveraged for intrusion detection purposes, and shed light on the gaps in the literature that are discussed in this thesis.

A recent topic without identity Because of the novelty of FL in the field of IDS, the literature on the topic is still scarce. Only a handful of reviews [Ala+21; Agr+22; Cam+22] had been published on the topic when we stopped our data collection for this study in late 2021. While these papers provide a good overview of the existing works, they fail to provide synthesis and extract the core characteristics of the field. Notably, *what makes FL for IDS different from FL for other applications, and what challenges are specific to the field of intrusion detection?*

A systematic approach We aim to address this gap as thoroughly and transparently as possible, and leverage the Systematic Literature Review (SLR) methodology to that end. This methodology [KC07] relies on a structured process to identify, select, and analyze the relevant literature on a given topic. With explicitly defined research questions and inclusion/exclusion criteria, the SLR methodology ensures that the review is reproducible

and unbiased. Therefore, we intend to provide a comprehensive overview of the existing literature, and reproducible, evidence-based conclusions on the specificities of FL for IDS.

Content The content of this chapter is based on our survey published in TNSM in May 2022 [Lav+22b] and its accompanying extension at the C&ESAR conference in November 2022 [Lav+22a]. Because the initial paper was submitted in November 2021, the quantitative analysis has been updated during the writing of this manuscript to include the latest publications on the topic. The qualitative analysis has also completed to a lesser extent.

Contributions of this chapter

- The first SLR on the use of FL for IDS, including qualitative and quantitative analyses of the literature.
- A generalization of the selected works as a reference architecture for FIDSs, providing a starting point for future works on the topic.
- A taxonomy synthesizing the state of the art of FIDS, providing a framework to analyze and compare existing and upcoming literature.
- The identification of the main challenges and opportunities in the field, and a set of research directions to address them.

3.2 Methodology

This section details the methodology applied to review the state of the art of FIDSs. The original article follows the SLR methodology introduced to the engineering field by Kitchenham *et al.* [KC07]. SLR uses analytical methods to answer research questions about the literature on a specific topic. The update to the original article is less structured and more focused on the evolution of the field, so the methodology is adapted accordingly.

3.2.1 Research Questions

The SLR methodology recommends defining explicit research questions to structure the review and the selection of papers. This survey aims at evaluating FIDS and their maturity, as well as their core components, and relevant variations. Therefore, using related and selected works, we identify the following Research Questions (RQs) that cover the topic of FIDSs. The questions complete and extend **??** RQ1 which was introduced in Chapter 1.

RQ1-1. *What are FIDS?*

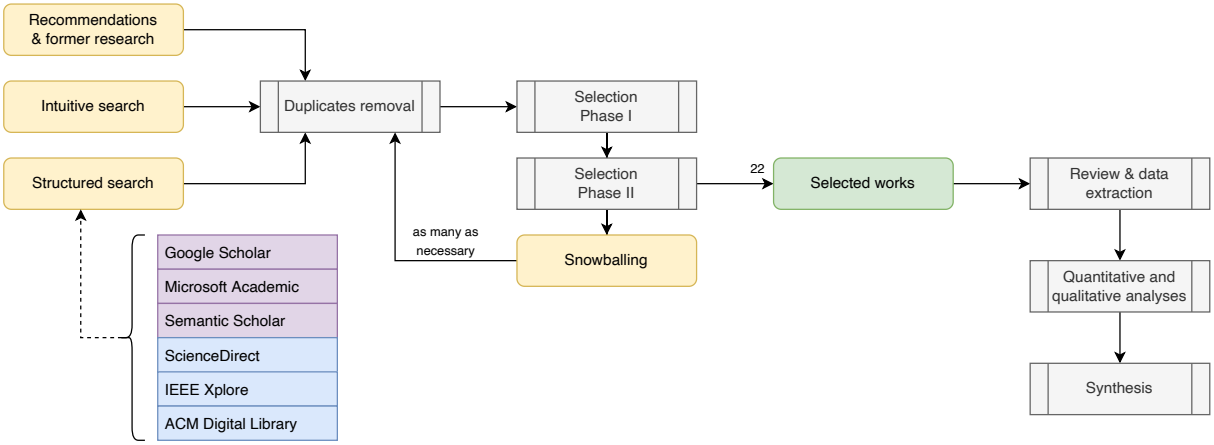


Figure 3.1 – Search and selection processes

1.a. What challenges do FIDS help to cope with?

1.b. Which techniques exist to federate ML–based detection and mitigation mechanisms?

RQ1-2. *What are the differences between FIDS?*

2.a. What are the key components of FIDS? How do they influence the system's performance?

2.b. Which metrics are used to measure and compare FIDS?

RQ1-3. *What is the state of the art of FIDS?*

3.a. What are the topics covered by the academic literature since 2016?

3.b. Where was the literature published? Which research groups and communities are active in this area?

3.c. What are open questions according to existing works?

3.2.2 Search and Selection Process

Figure 3.1 presents the methodology and its search, selection, and synthesis processes. In yellow, we identify the sources of papers, in green the final selection, and in gray the processing steps of the methodology. The tools used in the *Structured search* are presented with search engines in purple, and online databases in blue.

The searching of relevant literature involves four sources: recommendations, intuitive search, structured search, and snowballing.

(1) *Recommendations* were given by supervisors and coworkers throughout the realization of this work. This initial set of relevant papers is also used as a source of snowballing for further searching. Moreover, we included references from an aborted survey on *Collaborative security approaches*, which already yielded a substantial amount of literature by using the same methods.

(2) *Intuitive search* has been performed at the beginning of the survey to get a first grasp on the topic, and to learn about the functioning of FIDSs. At first, mostly Google Scholar has been used.

(3) *Structured search* has been adopted afterward, following the principles of SLR [KC07]. Different search engines and online databases are used for the sake of completeness, as illustrated in Figure 3.1. Databases can provide different results depending on their ownership. Search engine results differ according to the way requests are parsed, and the papers they have indexed. Thus, multiple sources provide more exhaustive results. (a) application of FL to IDSs, and (b) literature addressing the topic of FIDS with unusual keywords.

   (a) `("federated learning" OR "fl" OR "federated")`
   `AND ("intrusion detection systems" OR "ids")`

   (b) `("federated" OR "collaborative")`
   `AND ("detection" OR "defense" OR "mitigation")`

(4) *Snowballing* identifies relevant works that would have been missed otherwise, such as publications cited by articles of our selected corpus, or papers that refer to them. The related surveys identified in this work (Section 3.6) contain a lot of references to technical articles, making them relevant for snowballing. Furthermore, as this survey proceeds with quantitative analysis of the venues and groups (Section 3.4), it provides extended snowballing opportunities by looking at other publications in the most represented venues or research groups in the selected corpus.

Approximately two hundred papers have been identified. Duplicate removal is performed with Zotero which allows identifying and merging redundant items. The selection then happens in two phases. Firstly, the title and abstract are used to discriminate *out-of-scope* papers in Phase I, along with their number of citations given the search engines, and age. However, a paper with few citations, but interesting abstract, probably only lacks visibility. Thus, it is moved to Phase II, which consists of a more thorough analysis of the selected works, using the *three-pass* approach defined by Keshav [Kes07].

After the two selection phases, 22 papers were selected, excluding the 18 initial surveys seen in Section 3.6. All present technical solution for FIDS. The challenges identified in Chapter 2 were also used to either search or select papers, mostly through the *intuitive search* part.

3.2.3 Data Extraction and Analysis

The quantitative section of the original paper was solely based on the 22 selected papers. However, a significant amount of literature has been published since the initial survey. Therefore, we updated the quantitative analysis to include the latest publications on the topic. The qualitative analysis has also been completed to a lesser extent, just

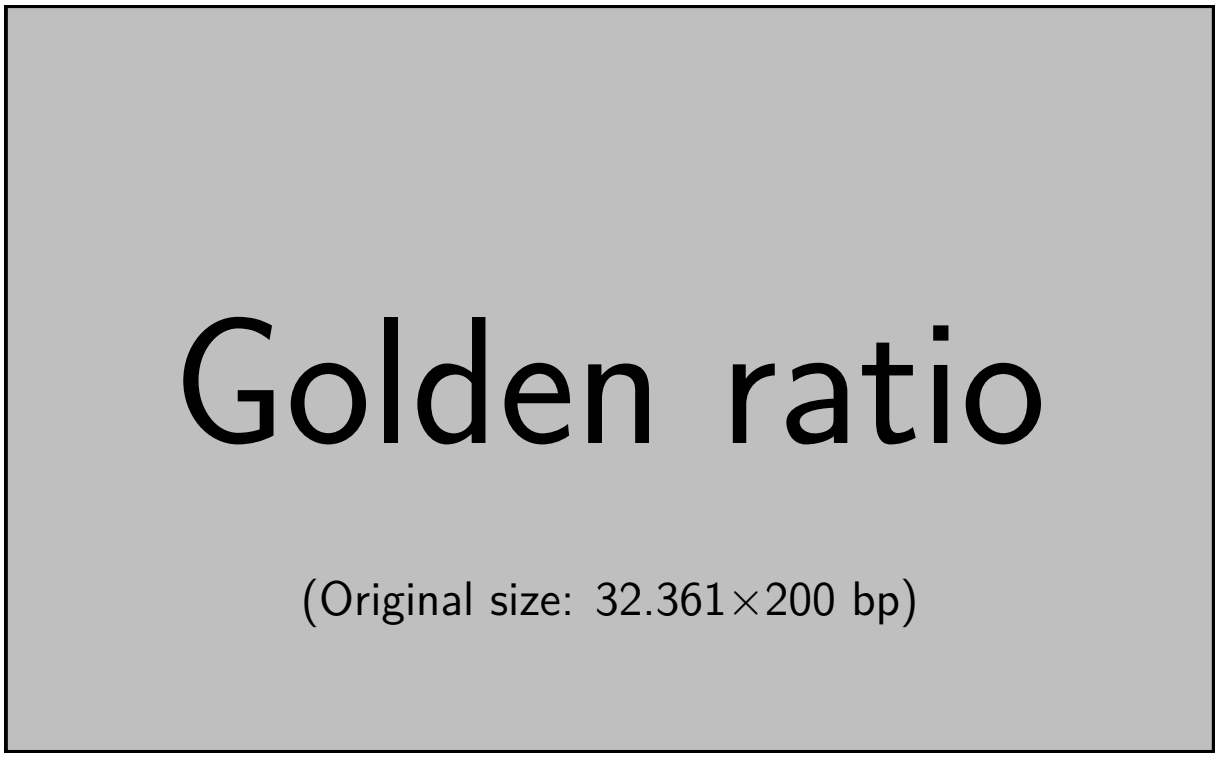


Figure 3.2 – Updated selection process.

enough to provide a general overview of the field's evolution. **??** presents the methodology applied to update the presented results.

We set up an automated collection system on Google Scholar, composed of an alert based on the search queries defined in Section 3.2.2 and automated recommendations. The system was set up in 2021 and ran until the end of the writing of this manuscript. It brought 423 emails containing 2490 links after duplicate removal. A first selection was performed on the title and abstract, yielding 238 papers. After manual filtering, we select 158 relevant papers, which amount to 136 new publications since the original survey.

To process this new corpus, we use Litstudy [Hel+22], a Python library providing tools to extract and analyze bibliographic data. On the 158 selected papers, 153 only were available on Scopus, the database available in Litstudy that provides the most complete data. The list of papers is available as appendix of this manuscript.

Literature distribution The distribution of the literature is analyzed in terms of publication year, venue, research group, and authors. We use the properties of the document set generated by Litstudy to filter and group the papers, and use the provided Pandas adn Matplotlib bindings to analyze and plot the data.

Topic modeling Litstudy also provides tools to perform topic modeling on the text data of the papers, mainly title and abstract. We first preprocess the text data by removing stop words, punctuation, and numbers, and then associate each word with its frequency in the

corpus. Then, we test the two main approaches of available in the literature, namely Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF), to identify the main topics in the corpus. The presented results have been obtained using the NMF algorithm on 20 topics and after 2000 iterations, as it provided the most interpretable results.

3.3 Qualitative Analysis

This section contains the results of our literature review. First, it synthesizes the analyses into a reference architecture and a taxonomy for FIDSs, which help structure the field. Then, it goes over a comparison of the selected works to answer Question RQ1-2.a on the components of FIDSs and their impact on performance.

3.3.1 Structuring the Literature

The qualitative (Section 3.3) and quantitative (Section 3.4) analyzes provide results that we synthesize hereafter in a reference architecture and a taxonomy. The reference architecture presents the components of FIDSs and their interactions, while the taxonomy provides comparison criteria for the selected works.

We build the taxonomy upon different existing ones related to CIDS [VKF15; ZLK10], ML–based intrusion detection [dCos+19], and FL [Ale+20; LYY20; Mot+21b]. First, we extract classes relevant to the domain of FIDS, before filtering out irrelevant ones by validating the taxonomy against the reference architecture (Figure 3.3). The latter displays both the operation and the design of the system. By confronting the taxonomy and the architecture, we ensure that each item of the taxonomy is related to a component of the architecture, and *vice versa*. Then, we add any commonalities between the selected works that are not already represented in the previous taxonomies. This identifies new criteria on which to compare the selected works.

Reference Architecture

This section presents the reference architecture synthesized from the selected works, as depicted in Figure 3.3. It can be divided in three parts:

— The *Managed system* represents the monitored system, *e.g.*, IT network, industrial devices, or health-monitoring wearables. As noticed in Section 3.3.2, collected data can either concern system or environment behavior. The former relates to information generated by the systems, *e.g.*, network traces or resource consumption. The latter refers to what the monitored system operates on, *e.g.*, health metrics for medical devices of temperature and atmospheric pressure for building management systems.

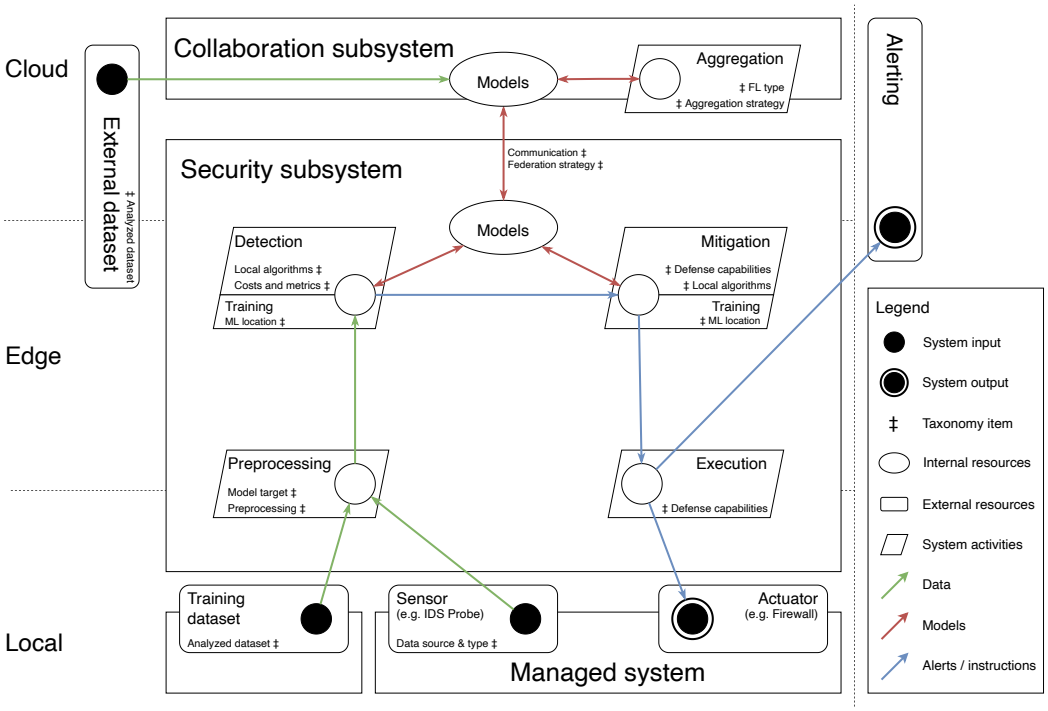


Figure 3.3 – The proposed reference architecture for FIDSs—Figure from Lavaur, Pahl, *et al.* [Lav+22b] © IEEE 2022.

— The *Security subsystem* is the core of the architecture. It contains all the system's activities, from model training to detection and counter-measures deployment. Depending on the objectives and constraints, this subsystem can either be run locally like [PA18] or [Hei+20], on a dedicated edge-device as in [Li+20a]. In the case of centralized learning, this entire subsystem runs in the cloud. The subsystem is assumed to run a device that embeds enough computing power to perform real-time anomaly detection against ML models. It is also capable of training its own model based on collected data.

— The *Collaboration subsystem* provides the *sharing* feature of the system, essentially model aggregation (Section 3.3.2). It also provides optional training from other sources, like online datasets.

This architecture has similarities with the principles of autonomic systems, as defined by IBM in 2001 [KC03], referred to as Monitor-Analyze-Plan-Execute plus Knowledge (MAPE-K). Classic autonomic systems are local, and therefore use a database to provide *knowledge*. In FIDS, FL fills this role in the reference architecture, as the knowledge is being shared among all agents through model aggregation.

Taxonomy for FIDS

The taxonomy depicted in Figure 3.4 summaries the core components and specificities of FIDSs, as extracted from the selected works and existing related taxonomies. Correlations between the taxonomy items and the system's components can be seen in the reference architecture (Figure 3.3). It also serves as a framework for the comparisons of the selected works. Each class represents a building block, for which multiple approaches exist depending on use case and constraints.

The proposed taxonomy contains 12 classes describing the selected works that span over five main aspects:

— Two classes cover the topic of **Data**: *Data Source and Distribution* and *Preprocessing*. It defines the type of data considered and how it is distributed among clients, how it is collected, and the preprocessing strategies that are used.

— **Local operation** is represented by 3 classes: *Algorithm location*, *Local Algorithm*, and *Defense Mechanism*. It describes the detection and mitigation strategies, how models are built and trained, and where the computing resources are located.

— The **Federation** aspect is covered by 2 classes: *Federation Strategy* and *Communication*. They refer to the communication between the agents and the server, and how data sharing is organized.

— **Aggregation** is also covered by 3 classes: *FL Type*, *Aggregation Strategy*, and *Model Target*. It describes the type of FL used, how the models are merged, in accordance with the objectives of the system.

— Finally, 2 classes address the **Experimentation** topic: *Analyzed Dataset* and *Costs and Metrics*. This meta-category does not relate to the proposed solution, but to how the experiments are performed.

3.3.2 Federated Learning for Intrusion Detection

This section reviews the selected literature. Using the taxonomy as a reference, it details and compares the selected works. Table 3.1 summarizes the information and helps identify differences between the works. It gives partial answers to research questions about the components of FIDSs and how to measure their impact on performance (Questions RQ1-2.a and RQ1-2.b), while Section 3.3.2 replies to Question RQ1-1.b about federation techniques.

Data Source and Distribution

The selected works highlight two main characteristics of the training data that impact the design of FIDSs: the origin of the data and its distribution among clients. The type of data used in the selected works is diverse, ranging from network traffic [CZY20; RWP19] to

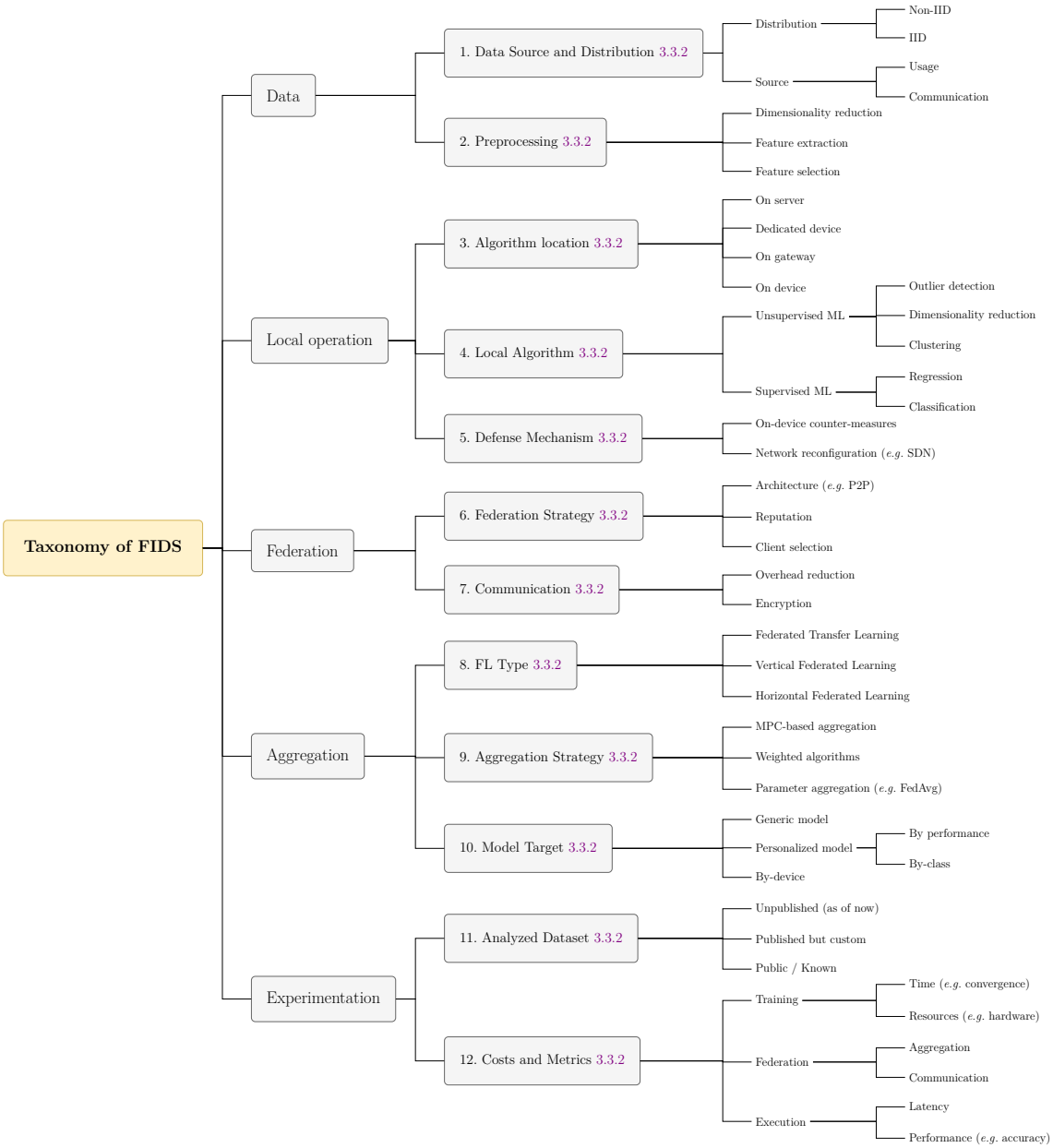


Figure 3.4 – Proposed taxonomy for FIDS—Figure from Lavaur, Pahl, *et al.* [Lav+22b] © IEEE 2022.

Table 3.1 – Comparative overview of selected works in the original study—approach and objectives (1/2)

	Ref	Internet of Things	Information Technologies	Cyber Physical Systems	Autonomous Vehicles	Satellite-terrestrial networks	Horizontal FL	Vertical FL	Federated Transfer Learning	Federated MTL	Federated Mimic Learning	Online learning	Semi-supervised	Supervised	Unsupervised	Personalized models	Network-based	Usage-based	Training location	Data type	Strengths
2018	Pahl *et al.* [PA18]	●	○	○	○	○	●	○	○	○	○	○	●	○	○	●	●	●	○	Device	Abstracted network traffic (middleware)	relatively lightweight, online, no labels
2019	Rathore *et al.* [RWP19]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Edge-controller (SDN)	Network traffic (SDN)	offers mitigation, decentralized
2019	Schneble *et al.* [ST19]	●	○	○	○	○	●	○	○	○	○	○	●	○	○	●	●	●	○	Gateway	IoT network traffic (TCPdump)	online, offers per-class models, no labels
2019	Nguyen, Marchal, *et al.* [Ngu+19]	○	●	○	○	○	○	○	○	●	○	○	○	●	○	○	●	●	○	Gateway	Encrypted network traffic (CICFlowMeter)	versatile (multi-task)
2019	Zhao *et al.* [Zha+19]	○	○	●	○	○	●	○	○	○	○	○	●	○	○	●	○	○	●	Gateway	Healthcare sensor values	high adaptability, no labels
2019	Cetin *et al.* [Cet+19]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Gateway	Network traffic (WIFI)	–
2020	Li, Wu, *et al.* [Li+20a]	●	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	●	Gateway	Air conditioner sensor values	offers traceability (blockchain)
2020	Chen, Zhang, *et al.* [CZY20]	○	○	●	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Gateway	MODBUS traffic	confidentiality (encryption)
2020	Zhang *et al.* [Zha+20]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Device	IoT network traffic (TCPdump)	–
2020	Fan *et al.* [Fan+20]	○	●	○	○	○	●	○	○	○	○	○	○	○	○	●	○	●	○	Gateway	IoT network traffic (TCPdump)	no labels
2020	Rahman *et al.* [Rah+20]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Gateway	Network traffic (PCAP)	segmented (performance-based models)
2020	Sun, Ochiai, *et al.* [SOE20]	●	○	○	○	○	○	○	●	○	○	○	○	●	○	○	●	●	○	Gateway (MEC)	IoT network traffic (TCPdump, CICFlowMeter)	knowledge transfer between public and private datasets
2020	Al-Athba Al-Marri *et al.* [ACA20]	○	●	○	○	○	○	○	○	○	●	○	○	●	○	○	○	●	○	Gateway	Network traffic (TCPdump)	enhanced privacy (mimic learning)
2020	Kim, Cai, *et al.* [Kim+20]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Gateway	Network traffic (TCPdump)	–
2020	Qin, Poularakis, *et al.* [Qin+20]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Gateway (SDN)	Network traffic (SDN)	very lightweight, line-speed classification, P4 language compatible
2020	Chen, Lv, *et al.* [Che+20]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Gateway	Network traffic (CICFlowMeter)	robust to poisoning, scalable
2020	Hei *et al.* [Hei+20]	○	●	○	○	○	●	○	○	○	○	○	●	○	●	○	○	●	○	Device	Network traffic (TCPdump)	online, offers traceability (blockchain)
2020	Li, Zhou, *et al.* [Li+20b]	○	●	○	○	●	●	○	○	○	○	○	○	●	○	○	○	●	○	Gateway	Network traffic (PCAP, CICFlowMeter, Argus)	relatively lightweight, confidentiality (encryption)
2021	Liu *et al.* [Liu+21]	●	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Gateway	IoT network traffic (TCPdump, Argus)	zero-days detection
2021	Popoola, Gui, *et al.* [Pop+21b]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Device	Network traffic (TCPdump)	relatively lightweight
2021	Qin and Kondo [QK21]	○	○	○	●	○	●	○	○	○	○	○	○	●	○	○	○	●	○	Device	Network traffic (TCPdump)	decentralized
2021	Sun, Esaki, *et al.* [SEO21]	○	●	○	○	○	●	○	○	○	○	○	○	●	○	○	●	●	○	Gateway	Network traffic (PCAP)	segmented (performance-based models)
		Use case					FL type						Training				Approach					

sensor values [Zha+20; ST19]. The former is significantly more represented, probably due to the availability of public datasets like CICIDS2017 [SHG18] and UNSW-NB15 [MS15] (see Section 3.3.2).

Most papers [CZY20; RWP19; Ngu+19; Li+20a; Rah+20; SOE20; Pop+21a; Hei+20] use similar network features, such as source and destination, local and remote ports, TCP flags, protocol, and packet length. The authors of [Qin+20] also target network features but at packet-level, all translated to 1D vectors: IP addresses, layer-4 protocol, ports, and IP packet length as a 120-bit input vector. Li, Wu, *et al.* [Li+20a] also explore network-related features in their use case of satellite communications. These values can be completed with preprocessing (see Section 3.3.2) to extract other features from the raw data. For instance, both Pahl *et al.* [PA18] and Nguyen, Marchal, *et al.* [Ngu+19] analyze the periodicity of packets, which is notably useful for volumetric attack detection. By using a middleware to classify the data, Pahl *et al.* [PA18] can train per-class models. Such models are more specialized and thus more accurate, but most communication layers do not provide such metadata. Training per-class models usually requires then a prior classification step, like in [Ngu+19]. The use of specialized models is further discussed in Section 3.3.2.

On the other hand, Zhang *et al.* [Zha+20] and Schneble *et al.* [ST19] use sensor values, such as hearth rate and oxygen saturation. In this case, one does not seek to detect intrusions per se, but rather anomalies in the data that could indicate a malfunction or an attack. The observed data can be seen as a side-channel, leaking information about the actions of potential attackers. More recently, FL has been applied to Host-based Intrusion Detection Systems (HIDSs) [Guo+23], were similar considerations apply, particularly in terms of data distribution.

Finally, even when considering the same data type, use cases introduce significant differences in the available features. For instance, two systems targeting the communication between devices may encounter different protocols, services, and even communication support. In the literature, the most common use cases are (sorted by representation): Information Technology (IT), Internet of Things (IoT), Cyber-Physical System (CPS), and Autonomous Vehicles (AV). While it is unlikely that a system would target multiple use cases, discrepancies in the data distribution can exist within a single use case. Chen, Zhang, *et al.* [CZY20], and partly Hei *et al.* [Hei+20], address the topic of skewed data distribution. A non-Independent and Identically Distributed (IID) data distribution can negatively impact training performance [Yan+19]. However, most real-world scenarios generate non-IID data, which is a major drawback of the selected works, as most of them do not address this issue.

Preprocessing

In addition to the type of data considered, the preprocessing pipeline has a significant impact on the performance of the system. Preprocessing implies the transformation of raw data into a format that can be better leveraged by ML models, either by extracting new features or by reducing the dimensionality of the data. Three main non-exclusive approaches are distinguishable in the selected works: feature extraction, feature embedding, and feature selection:

— *Feature extraction* refers to the computation of numerical characteristics after the data collection; *e.g.* Inter-Arrival Time (IAT) or number of packets per device in the context of traffic monitoring. For instance, both Nguyen, Marchal, *et al.* [Ngu+19] and Pahl *et al.* [PA18] extract periodicity features from the data. Because they only process binary features, Qin, Poularakis, *et al.* [Qin+20] extract numerical features, and convert them to 1D vectors.

— *Feature embedding* or *dimensionality reduction* is used for algorithms that do not deal efficiently with high-dimensional vectors. We mostly use the term *embedding* when the authors use DL techniques, as it implies that the model learns the best representation of the data, such as with autoencoders [CZY20]. Other *dimensionality reduction* techniques include Principal Component Analysis (PCA), used for example by Kim, Cai, *et al.* [Kim+20].

— *Feature selection* relates to the automated selection of relevant features, before learning. The authors of [QK21] use a greedy feature selection algorithm based on accuracy. Logistic regression-based selection [ACA20] can also be used to eliminate features with a recursive algorithm.

The other works [Zha+20; ST19; Li+20a; RWP19] do not emphasize on their feature selection strategy. Moreover, some papers [Li+20a; ST19; Zha+19] use datasets that contains computed features (3.3.2). For experiments on live prototypes, feature computation is required.

Depending on the use case, additional features after *feature selection* or *extraction* may vary. Network analysis often relies on basic features, such as addresses and ports for source and destination, protocol, data type, packet length, and timestamp. However, these characteristics can also vary regarding their provenance: network capture [Sig99; Tav+09] or abstracted communications [PA18]. Extracted features are very common, such as inter-packet time, bytes sent per host, or bytes per packets [BG16; Cha+19]. For instance, both Nguyen, Marchal, *et al.* [Ngu+19] and Pahl *et al.* [PA18] target IoT devices, which have a sporadic, but periodic and thus more predicable traffic. In this context, anomaly in the packet-sequence, or in the inter-arrival time might indicate an attack.

Usage-based analysis, on the other hand, is entirely dependent on the monitored device. Schneble *et al.* [ST19] monitor health-related features, like arterial blood pressure

or the raw ECG signals. The authors of [Zha+20] focus on air conditioners, and therefore measure related information such as water or air temperature.

Algorithm location

The proposed taxonomy (3.4) considers three types of locations: on-device, on-gateway, and on-server. However, a large majority of the literature concerns either on-device training, or uses a dedicated device acting as a gateway. Most selected works use a dedicated device to perform the analysis, while the others assume the devices can support their own processing. The on-server processing is not represented here, since it does not suit the definition of FL. Some hybrid approaches are also represented, with multi-stage aggregation [Liu+21].

The device types and architectural choices are mostly dicted by the chosen use case. For instance, Zhang *et al.* [Zha+20] focus on a medical use case where the analyzed data is composed solely of sensor outputs (Section 3.3.2). Connected sensors are typically lightweight devices unable to process data. Thus, they require a gateway to be usable. Most works [Li+20a; CZY20; ST19; Zha+19; ACA20; Kim+20; CZY20; Pop+21a] rely on gateways because they are more suitable for traffic analysis. It allows to capture all communications, even if the devices are connected with different supports (*e.g.*, IEEE 802.3 *vs.* IEEE 802.11). Gateway-based processing can also be motivated by the architecture of the monitored system. For instance, the authors of [Fan+20] reuse the existing infrastructure of 5G by exploiting Mobile Edge-Computing (MEC) gateways to capture traffic and perform analysis for a 5G IoT use case.

Table 3.2 – Comparative overview of selected works in the original study—algorithms and performance (2/2)

	Ref	Local Algorithm	Federation Algorithm	Accuracy	Precision	Recall	Fall-out	F-Score	K [a]	Dataset
2018	Pahl *et al.* [PA18]	BIRCH K-means	Parameter addition	0.9900	–	0.9600	0.0020	–	7	Generated
2019	Rathore *et al.* [RWP19]	ANN	Vector concatenation	‡ 0.9100	‡ 0.9100	‡ 0.9100	–	‡ 0.9100	15	NSL-KDD [Tav+09]
2019	Schneble *et al.* [ST19]	MLP	Weight and biases average	0.9930	–	–	–	–	64	MIMIC [Joh+16]
2019	Nguyen, Marchal, *et al.* [Ngu+19]	GRU	FedAvg	–	–	0.9543	0	–	15	Generated
2019	Zhao *et al.* [Zha+19]	FC (shared layers) → FC	Weight and biases average	∗ 0.9797	∗ 0.9634	∗ 0.9681	–	–	–	CICIDS2017 [SHG18] ISCXVPN2016 [Dra+16] ISCXTor2016 [Hab+17]
2019	Cetin *et al.* [Cet+19]	SAE	FedAvg	–	–	–	–	–	933	AWID [Kol+16]
2020	Li, Wu, *et al.* [Li+20a]	CNN-GRU → MLP	Homomorphic parameter addition	0.9920	0.9885	0.9745	–	0.9813	7	CPS dataset [MG14]
2020	Chen, Zhang, *et al.* [CZY20]	DAGMM	Parameter addition	–	0.7447	0.9803	–	‡ 0.8700	2 [b]	KDD 99 [Sig99]
2020	Zhang *et al.* [Zha+20]	ANN	CDW_FedAvg	∗‡ 0.8900	∗‡ 0.8600	∗‡ 0.9450	–	∗‡ 0.8500	4	Generated
2020	Fan *et al.* [Fan+20]	CNN	Parameter aggregation	∗ 0.9100	–	∗‡ 0.9350	∗‡ 0.0020	–	4	CICIDS2017 [SHG18] NSL-KDD [Tav+09] Generated
2020	Rahman *et al.* [Rah+20]	ANN	FedAvg	∗ 0.7731	–	–	–	–	4	NSL-KDD [Tav+09]
2020	Sun, Ochiai, *et al.* [SOE20]	CNN	Parameter aggregation	∗ 0.8710	–	–	–	–	20	LAN-Security Monitoring Project [Hid18]
2020	Al-Athba Al-Marri *et al.* [ACA20]	ANN	FedAvg	0.9812	∗ 0.9900	∗ 0.9900	∗ 0.1320	∗ 0.9900	10	NSL-KDD [Tav+09]
2020	Kim, Cai, *et al.* [Kim+20]	MLP	FedAvg	0.9712	–	–	–	–	4	NSL-KDD [Tav+09]
2020	Qin, Poularakis, *et al.* [Qin+20]	BNN	SignSGD	∗ 0.9640	∗ 0.9555	∗ 0.8645	–	∗ 0.9055	8	CICIDS2017 [SHG18] ISCX Botnet 2014 [Big+14]
2020	Chen, Lv, *et al.* [Che+20]	GRU-SVM	FedAGRU	∗ 0.9905	–	–	∗ 0.0108	∗ 0.9762	20	CICIDS2017 [SHG18] KDD 99 [Sig99] WSN-DS [AAA16]
2020	Hei *et al.* [Hei+20]	MLP	FedAvg	∗‡ 0.8950	∗‡ 0.9750	∗‡ 0.8775	–	∗‡ 0.9225	3	DARPA 1999 [Hai+01]
2020	Li, Zhou, *et al.* [Li+20b]	CNN	Homomorphic parameter addition	∗ 0.8100			∗ 0.1900		4	Generated
2021	Liu *et al.* [Liu+21]	MLP	Parameter aggregation	‡ 0.9600	0.9400	0.9500	–	–	6	KDD 99 [Sig99]
2021	Popoola, Gui, *et al.* [Pop+21b]	ANN	FedAvg	∗ 0.9939	∗ 0.9819	∗ 0.9676	–	∗ 0.9728	5	Bot-IoT [Kor+19] N-BaIoT [Mei+18]
2021	Qin and Kondo [QK21]	ONLAD [TKM20] (ELM + AE)	FedAvg	0.7040	–	–	–	–	8	NSL-KDD [Tav+09]
2021	Sun, Esaki, *et al.* [SEO21]	CNN	Parameter aggregation	–	–	–	–	∗ 0.8930	20	LAN-Security Monitoring Project [Hid18]
						Metrics				

∗ Value is an average of those provided by the authors.
‡ Value is read from a graph in the article, and may vary a few from the exact value.

[a] K is the highest number of client considered in the experiments.
[b] Chen, Zhang, *et al.* [CZY20] measure how one client performs, by training one other.

Local Algorithm

Defense Mechanism

Federation Strategy

Communication

FL Type

Aggregation Strategy

Model Target

Analyzed Dataset

Costs and Metrics

3.4 Quantitative Analysis

This section provides indicators of the representation of FIDSs in the scientific literature: the evolution of the publications, the relevant venues, the active groups, and the topics of interest. Notably, the identification of the most active groups and most relevant venues provides insights on how to keep track of the most recent advances in the field.

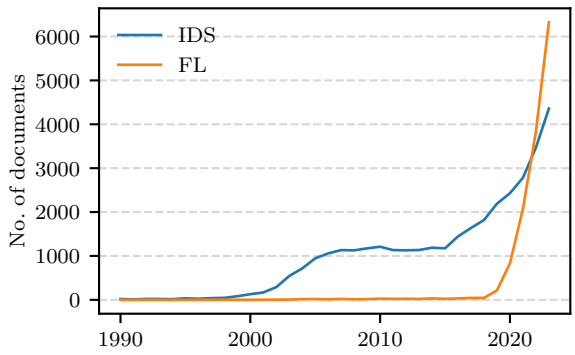
3.4.1 Evolution of the Topic

The topic of IDS started to gain traction in the late 1990', as depicted in Figure 3.5a. After a stagnation period, the topic regained interest around 2015, with an increase of the research on IoT and Industrial Internet of Things (IIoT) [DAF18; Cha+19], alongside other specific use cases. With the introduction of FL by McMahan *et al.* [McM+17], the community started to explore the application of FL to IDS around the years 2018–2019. Figure 3.5a has been generated using the analytics offered by Scopus and the following queries:

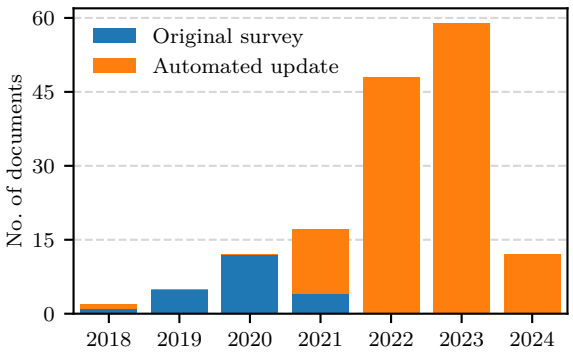
(a) `intrusion AND detection AND system`;

(b) `federated AND learning`.

Recent works on FL focus on its security and privacy-preserving aspects [Ngu+20; LYY20; Mot+21b]. Techniques like homomorphic encryption were introduced as early as 2017 [Har+17], and have been extensively reviewed since. More recently, other privacy-preserving techniques have been applied to FL, such as Multi-Party Computation (MPC) in FLGUARD [Ngu+21] or differential privacy in [KGS21]. FIDSs present a similar tendency with more research towards algorithm security and privacy-preserving techniques. For instance, Li, Wu, *et al.* [Li+20a] use homomorphic encryption to provide a secure and privacy-preserving aggregation of models. Aside from security, variations of Horizontal



(a) Evolution of the topics using Queries (a) and (b) according to Scopus, up to 2023.



(b) Evolution of the number of publications on FIDSs.

Figure 3.5 – Evolution of the topics and number of publications.

Federated Learning (HFL) started to appear in 2021, such as segmented FL in [Sun+20], as standard HFL has significantly been studied for FIDS.

Finally, the numerous literature reviews published since 2021 [Agr+22; Ala+21; Cam+22; Lav+22b; FNS22; GR22; Ism+24] show the continuous interest of the community for the study of FIDSs. These also show the need for synthesis and structuring of research in this area.

3.4.2 Relevant Venues

The initial study published in 2022 [Lav+22b] observed very few recurring venues for the publication of FIDS research. Indeed, only three venues had more than one publication on the topic: the *IEEE Internet of Things Journal* [Pop+21b; Zha+20], *IEEE Access* [Che+20; Li+20b], and the *IEEE BigData* conference [Cet+19; Fan+20]. The original distribution in terms of venue type (11 conferences, 10 journals and 1 book chapter) has significantly changed, since journals represent two thirds of the publications. This is a probable sign of the field gaining maturity, as publishing in conferences first and journals afterward is a common strategy. Figure 3.6 shows the distribution of the publications in the most recurring venues.

Another observation of the initial study was the diversity of the venues, spanning a wide range of topics, from IoT to ICS, including transportation systems and extra-terrestrial networks. This diversity is still present in the most recent publications, although a few generic venues now host significantly more publications: the *IEEE Internet of Things Journal*, *IEEE Access* and *Computers & Security*. The latter is the first security-specific venue to appear in the list. Its place in the top venues is a sign of the increasing interest of the security community for FL and FIDS, as most contributions were previously published in more use-case specific venues. Lastly, while relevant venues have been accepting FL literature since its introduction, they start to host specific tracks or special issues, such as

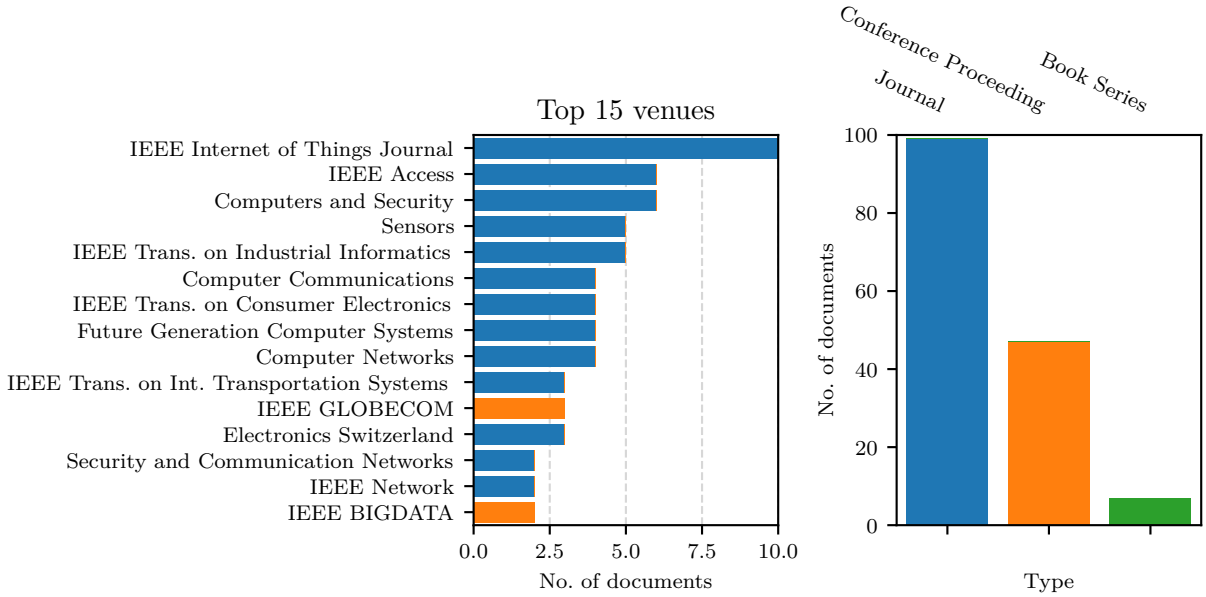


Figure 3.6 – Distribution of the publications in the most recurring venues.

ICDCS's track on "Federated Learning, Analytics, and Deployment", or IEEE BigData's "Special Session on Federated Learning on Big Data".

3.4.3 Active Groups

Since they introduced the topic of FL in 2016, the team at Google Research has been a big influence for the research community [Kon+16b; Kon+16a; McM+17; Bon+17; Bon+19]. They mostly work on the primitives behind FL, such as model aggregation with the `FedAvg` algorithm [McM+17]. The team of TU Darmstadt (Germany) has also been very active in the field, with a focus on IDS with `DÏoT` [Mar+19; Ngu+19] and FL security [Ngu+20]. The two collaborated, bringing `FLGUARD` [Ngu+21] and `FLAME` [Ngu+22], two algorithms focusing on limiting the impact of poisoning attacks in FL. These series of works makes them one of the most impactful groups in the field.

Other noteworthy groups include the Aalto University (Finland) [Ngu+21] and the University of Tokyo (Japan) [Sun+20; SEO21; QK21]. The most active country remains China, with a dozen institutions now amounting to a third of the publications in the field, as illustrated by Figures 3.7 and 3.8b.

Investigating the major authors tells another story, as the most active authors are not necessarily affiliated with the groups mentioned above. In particular, Popoola, Gui, *et al.* co-authored several publications on FIDS [Pop+21b; Pop+21a; Pop+22; Pop+23] as a collaboration between the Nanjing University of Posts and Telecommunications and multiple British universities. Likewise, Duy *et al.*, from the University of Information Technology (VNU, Vietnam), are also quite represented in terms of publications [Duy+21; Vy+21; Thi+22; Quy+22].

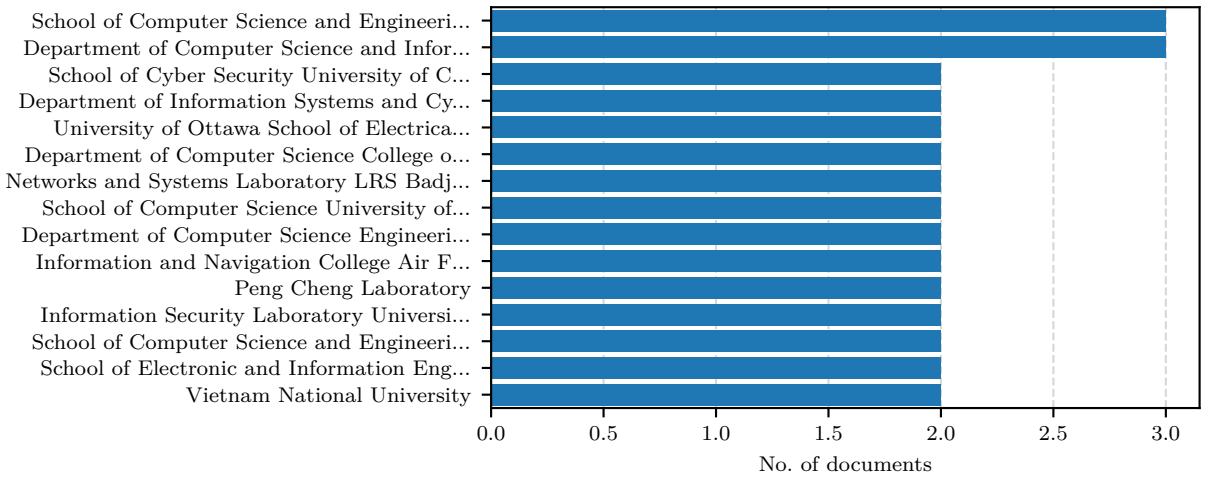
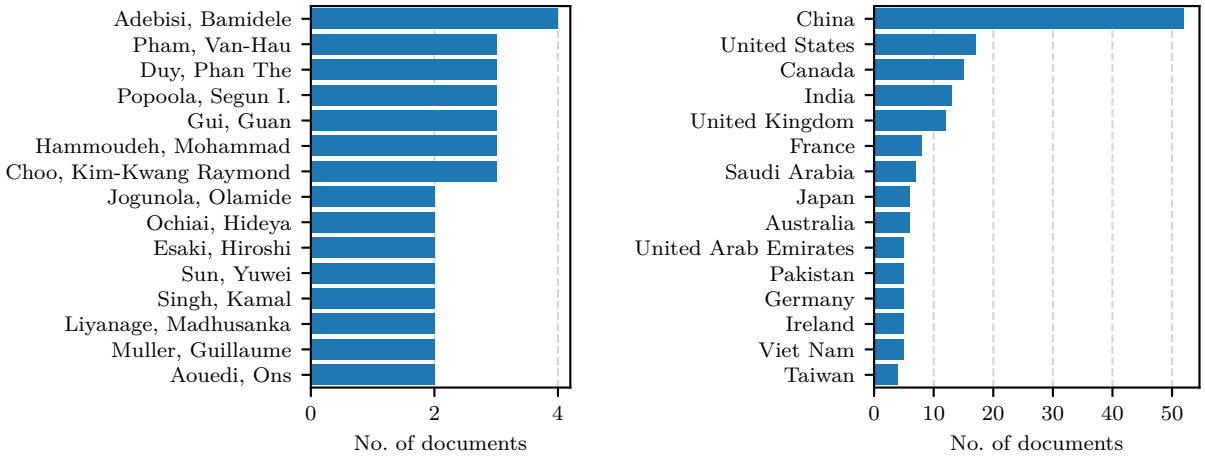


Figure 3.7 – Distribution of the publications by affiliation.



(a) Publications by author.

(b) Publications by country.

Figure 3.8 – Distribution of the publications by author and country.

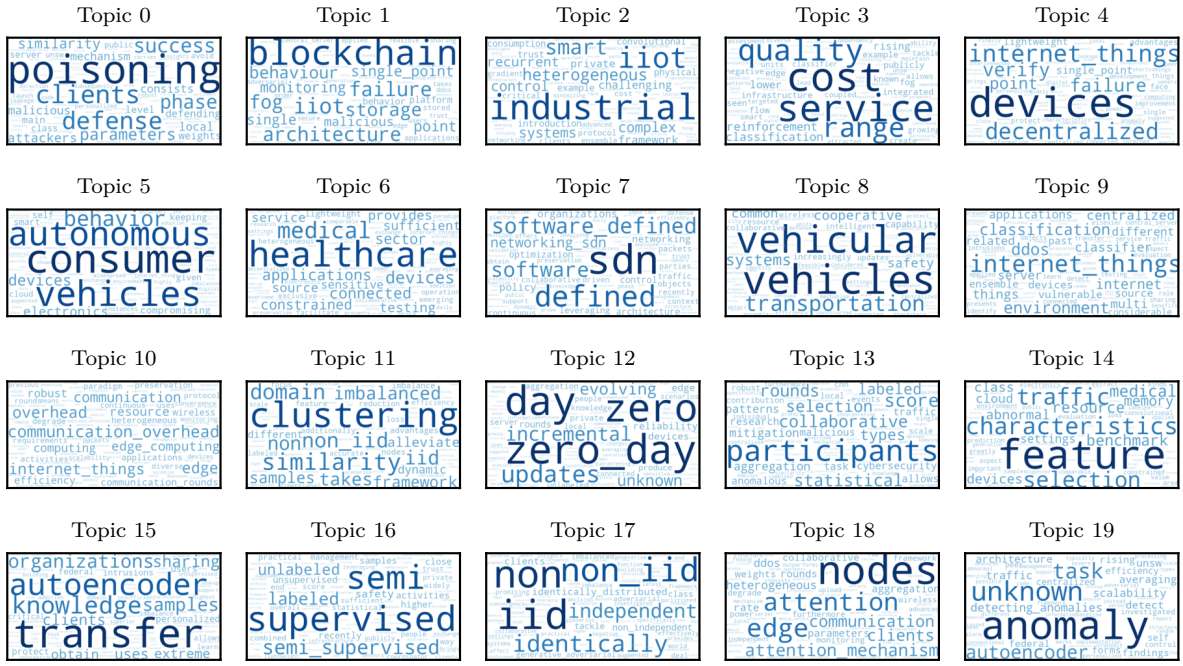


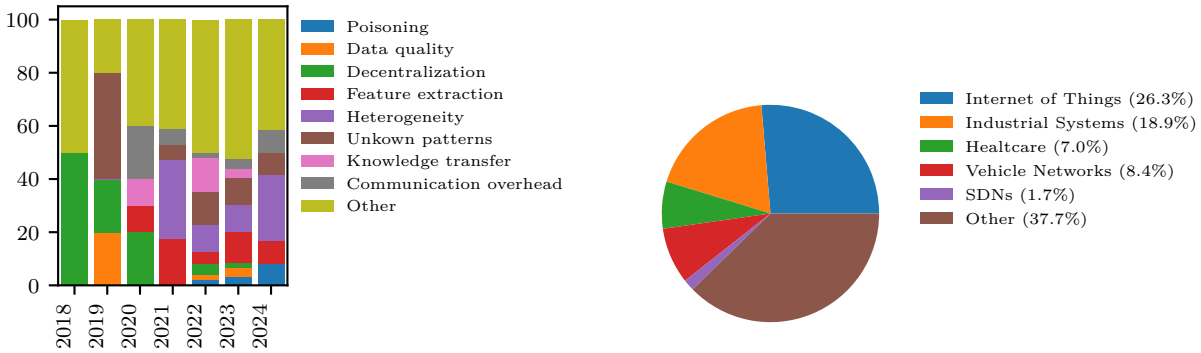
Figure 3.9 – Topics of interest in the field of FIDSs.

3.4.4 Topics of Interest

Using topic modeling, we extract the main topics of interest from the 153 publications on FIDSs identified in the updated selection. By construction, the model is unable to differentiate between application domains (such as IoT or ICS) the techniques used (*e.g.*, blockchains) or the addressed challenges in a paper. However, it provides a good overview of the main topics of interest in the field, especially for the consequent amount of literature published since the initial study. Figure 3.9 present the topics identified by the model, with the most recurring keywords for each topic.

First, this analysis highlights the application domains of FIDSs, where the topic of IoT (*i.e.*, `internet_things`, `edge`, `things`) is one of the most recurring (Topics 4, 9, and 10). Other applications stand out, such as ICS (`industrial`, `iiot`), Internet of Medical Things (IoMT) (`medical`, `healthcare`), Vehicle-to-Everything (V2X) (`vehicle`, `vehicular`, `transportation`), and Software-Defined Networking (SDN) (`software`, `defined`, `sdn`). These applications also correlate with the venues identified in Section 3.4.2, as the *IEEE Internet of Things Journal* or the *IEEE Trans. on Industrial Informatics* do focus on IoT and ICS, respectively. Figure 3.10b depicts the distribution of the publications by domain overall.

Likewise, some topics are directly associated with the challenges identified in Section 3.5.3. For instance, Topic 0 (`poisoning`, `defense`, `malicious`) represents works focusing on adversarial attacks against FIDSs and their mitigation. Some techniques can also be extracted from these results. For instance, Topic 0 also contains `similarity` as



(a) Distribution of the addressed challenges over time.

(b) Distribution of the publications by domain.

Figure 3.10 – Exploiting the topics of interest.

a keyword, which is likely to refer to the use of similarity metrics to detect poisoning attacks. This is indeed one of the most represented mitigation techniques in the literature on FIDS [Yan+23] or FL alike [FYB20; Ngu+22]. Figure 3.10a depicts the distribution of the addressed challenges over time. Unlike the distribution of the publications by domain, some challenges are addressed in the literature much later, such as handling the heterogeneity of the data (**??**) or resisting to adversarial attacks (**??**). Both are challenges that have been identified in the initial study as open issues in the field [Lav+22b; Lav+22a].

3.5 Discussion

3.5.1 Synthesizing the State of the Art

3.5.2 Limitations of this Study

3.5.3 Open Issues and Future Directions

3.6 Related Work

At the time of writing this literature review, the literature on FL for IDS was still scarce. Only a handful of reviews had been published on the topic [Ala+21; Agr+22; Cam+22]. Therefore, we extended our search of related works to related topics that were susceptible to share similar challenges or conclusions. This extended selection can be divided into three main categories: (a) security information sharing, (b) intrusion detection, and (c) collaborative ML. Table 3.3 provides a summary of this selection, grouped by topic and sorted by publication date. In addition to the initial selection, we also included more recent surveys on the topic [FNS22; GR22; Ism+24], whose number highlights the massive interest in the community.

Table 3.3 – *Related literature reviews, their topics, contributions, and number of citations according to Google Scholar (Apr. 2024).* Works marked ∗ were originally available as preprints, and were only published afterward. Works marked ‡ are added for the sake of completeness, but were not included in the initial selection.

Domain	Year	Authors	Contributions							Cited	Ref.
			Qualitative analysis	Quantitative analysis	Taxonomy	Reference architecture	Performance evaluation	Research directions	Systematic Literature Review		
Security information sharing	2016	Skopik *et al.*	●	○	○	○	○	●	○	291	[SSF16]
	2018	Tounsi *et al.*	●	●	○	○	○	●	○	448	[TR18]
	2019	Wagner *et al.*	●	●	○	○	○	●	○	240	[Wag+19]
	2019	Pala *et al.*	●	●	○	●	○	●	○	63	[PZ19]
ML for intrusion detection	2016	Buczak *et al.*	●	○	○	○	○	◐	○	3105	[BG16]
	2018	Meng *et al.*	●	○	○	○	○	●	○	562	[Men+18]
	2019	Chaabouni *et al.*	●	○	◐	○	○	●	○	790	[Cha+19]
	2019	da Costa *et al.*	●	○	○	○	○	●	○	492	[dCos+19]
Collaborative detection	2010	Zhou *et al.*	●	○	○	○	○	●	○	517	[ZLK10]
	2015	Vasilomanolakis *et al.*	●	○	●	○	○	●	○	379	[VKF15]
Federated learning	2020	Aledhari *et al.*	●	○	○	○	○	○	○	517	[Ale+20]
	2020	Lyu *et al.* ∗	●	○	○	○	○	●	○	436	[LYY20]
	2020	Shen *et al.*	●	○	○	○	○	●	○	69	[She+20]
	2021	Mothukuri *et al.*	●	○	●	○	○	●	○	376	[Mot+21a]
	2021	Lo *et al.*	●	●	○	○	○	●	●	158	[Lo+21]
FL for intrusion detection	2021	Agrawal *et al.* ∗	●	○	○	○	○	●	○	142	[Agr+22]
	2021	Alazab *et al.*	●	○	○	○	○	●	○	158	[Ala+21]
	2021	Campos *et al.* ∗	●	○	○	○	●	●	○	123	[Cam+22]
	2022	**Lavaur *et al.***	●	●	●	●	○	●	●	22	[Lav+22b]
	2022	Fedorchenko *et al.* ‡	◐	○	○	○	○	○	○	22	[FNS22]
	2022	Ghimire *et al.* ‡	●	○	○	○	○	●	○	208	[GR22]
	2024	Isma'ila *et al.* ‡	●	●	○	○	○	●	●	0	[Ism+24]

● covers topic; ◐ partly addresses topic; ○ does not cover topic.

Common issues of collaborative systems, such as the need for trust, privacy, and security, can also apply to FL-based collaboration systems. Therefore, we include four surveys [SSF16; TR18; Wag+19; PZ19] where the authors discuss the challenges and opportunities of sharing security-related information. They highlight the need for standardization, automation, and incentives, to achieve efficient and effective collaboration. The topic of trust is a clearly identified challenge in these works [Wag19; TR18]. The present study rather focuses on FL as a technical mean for collaboration, but such as trust or incentives are also relevant in this context.

Because ML-based IDS can be considered as a key component of FIDS, we review existing surveys on the topic [BG16; Men+18; Cha+19; dCos+19]. These work cover a wide range of solutions, from traditional ML (Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF), among others) to more recent approaches, such as deep learning, the latter being overrepresented in the literature of FIDSs. They also provide a good overview of the existing datasets and evaluation metrics, which can be useful for the evaluation of FL-based IDS. However, as noted in Section 3.5.3, typical IDS datasets present limitations that can hinder the evaluation of FL-based IDS.

FL is obviously another critical aspect of FIDSs. Consequently, related works include surveys on the collaborative aspects of ML (b) and FL [Ale+20; Lo+21]. They discuss FL approaches to work with distributed architectures. The security of FL is also heavily reviewed by [She+20; LYY20; Mot+21b]. They identify security threats like communication bottleneck, poisoning, and Distributed Denial of Service (DDoS) attacks, that could endanger FL-based systems. While the IDS use case can be seen as an application of FL, we argue that it raises specific concerns in terms of privacy, latency, and adaptability.

Zhou *et al.* [ZLK10] and Vasilomanolakis *et al.* [VKF15] survey the evolution of CIDS—at the merge of intrusion detection (b) and collaborative ML (c). Their works are however older and thus, cannot offer a comprehensive view of CIDS, as FL-based approaches did not exist at the time of their writing. Hence, the authors focus on collaboration in the sense of *detection+correlation*, whereas the analysis presented in this chapter (Section 3.3) surveys the use of FL in IDSs.

In addition to the above, recent work (*i.e.*, contemporary to the writing of the initial study) have reviewed the use of FL for intrusion detection [Agr+22; Cam+22; Ala+21]. Alazab *et al.* [Ala+21] address the wider topic of FL for cybersecurity, which only includes intrusion detection as an application. Their paper is explanatory and provides an overview of FL applications in information security. Like this work, Agrawal *et al.* [Agr+22] focus on FIDSs, but have different methodology. The authors list existing FIDSs and detail their approaches, and identify open issues. On the other hand, Campos *et al.* [Cam+22] review a subset of FIDSs by focusing on IoT use case, and the impact of non-IID (Independent and Identically Distributed) data on performance. While all identify challenges and research directions, this work also performs quantitative (Section 3.4) and qualitative

(Section 3.3) analyses of existing FIDSs, and extracts reference architecture and taxonomy. The existence of these papers emphasizes the importance and relevance of FIDSs for the research community.

The more recent works on the topic [FNS22; GR22; Ism+24] confirm these observations. The work of Fedorchenko *et al.* [FNS22] is of little interest, as it only lists and details existing works with close to no added value. Ghimire *et al.* [GR22] provide a more convincing study, closer to the method applied by Alazab *et al.* [Ala+21], but with a focus on the IoT. Finally, Isma'ila *et al.* [Ism+24] provide a comprehensive review, with up-to-date literature leveraging the SLR methodology, but still focuses on the IoT.

3.7 Conclusion

Part II

Quantifying the Limitations of FIDSs

Part III

Providing Solutions

BIBLIOGRAPHY

[16] *Directive (EU) 2016/1148 of 6 July 2016 Concerning Measures for a High Common Level of Security of Network and Information Systems across the Union*, 2016, URL: https://eur-lex.europa.eu/eli/dir/2016/1148/oj.

[22] *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on Measures for a High Common Level of Cybersecurity across the Union, Amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and Repealing Directive (EU) 2016/1148 (NIS 2 Directive)*, Dec. 14, 2022, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555 (visited on 05/26/2024).

[AAA16] Iman Almomani, Bassam Al-Kasasbeh, and Mousa AL-Akhras, « WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks », *in*: *Journal of Sensors* 2016 (2016), pp. 1–16, ISSN: 1687-725X, 1687-7268, DOI: 10.1155/2016/4731953, URL: https://www.hindawi.com/journals/js/2016/4731953/ (visited on 10/25/2021).

[ACA20] Noor Ali Al-Athba Al-Marri, Bekir S. Ciftler, and Mohamed M. Abdallah, « Federated Mimic Learning for Privacy Preserving Intrusion Detection », *in*: *2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), May 2020, pp. 1–6, DOI: 10.1109/BlackSeaCom48709.2020.9234959, URL: https://ieeexplore.ieee.org/document/9234959 (visited on 04/12/2024).

[Agr+22] Shaashwat Agrawal *et al.*, « Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions », *in*: *Computer Communications* 195 (Nov. 1, 2022), pp. 346–361, ISSN: 0140-3664, DOI: 10.1016/j.comcom.2022.09.012, URL: https://www.sciencedirect.com/science/article/pii/S0140366422003516 (visited on 06/09/2024).

[Ala+21] Mamoun Alazab *et al.*, « Federated Learning for Cybersecurity: Concepts, Challenges and Future Directions », *in*: *IEEE Transactions on Industrial Informatics* (2021), pp. 1–1, ISSN: 1551-3203, 1941-0050, DOI: 10/gnm4dj, URL: https://ieeexplore.ieee.org/document/9566732/ (visited on 12/02/2021).

[Ale+20] Mohammed Aledhari *et al.*, « Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications », *in*: *IEEE Access* 8 (2020), pp. 140699–140725, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2020.3013541, URL: https://ieeexplore.ieee.org/document/9153560/.

[Beu+20] Daniel J Beutel *et al.*, « Flower: A Friendly Federated Learning Research Framework », 2020, arXiv: 2007.14390.

[BG16] Anna L. Buczak and Erhan Guven, « A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection », *in*: *IEEE Communications Surveys Tutorials* 18.*2* (2016), pp. 1153–1176, ISSN: 1553-877X, DOI: 10.1109/COMST.2015.2494502.

[Big+14] Elaheh Biglar Beigi *et al.*, « Towards Effective Feature Selection in Machine Learning-Based Botnet Detection Approaches », *in*: *2014 IEEE Conference on Communications and Network Security*, 2014 IEEE Conference on Communications and Network Security (CNS), San Francisco, CA, USA: IEEE, Oct. 2014, pp. 247–255, ISBN: 978-1-4799-5890-0, DOI: 10.1109/CNS.2014.6997492, URL: https://ieeexplore.ieee.org/document/6997492 (visited on 10/25/2021).

[Bon+17] Keith Bonawitz, Vladimir Ivanov, *et al.*, « Practical Secure Aggregation for Privacy-Preserving Machine Learning », *in*: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA: ACM, Oct. 30, 2017, pp. 1175–1191, ISBN: 978-1-4503-4946-8, DOI: 10.1145/3133956.3133982, URL: https://dl.acm.org/doi/10.1145/3133956.3133982.

[Bon+19] Keith Bonawitz, Hubert Eichner, *et al.*, « Towards Federated Learning at Scale: System Design », *in*: *arXiv* (Feb. 4, 2019), URL: http://arxiv.org/abs/1902.01046.

[Cam+22] Enrique Mármol Campos *et al.*, « Evaluating Federated Learning for Intrusion Detection in Internet of Things: Review and Challenges », *in*: *Computer Networks* 203 (Feb. 11, 2022), p. 108661, ISSN: 1389-1286, DOI: 10.1016/j.comnet.2021.108661, URL: https://www.sciencedirect.com/science/article/pii/S1389128621005405 (visited on 04/25/2024).

[Cet+19] Burak Cetin *et al.*, « Federated Wireless Network Intrusion Detection », *in*: *2019 IEEE International Conference on Big Data (Big Data)*, 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 6004–6006, ISBN: 978-1-72810-858-2, DOI: 10.1109/BigData47090.2019.9005507, URL: https://ieeexplore.ieee.org/document/9005507/ (visited on 10/25/2021).

[Cha+19] Nadia Chaabouni *et al.*, « Network Intrusion Detection for IoT Security Based on Learning Techniques », *in*: *IEEE Communications Surveys & Tutorials* 21.*3* (2019), pp. 2671–2701, ISSN: 1553-877X, DOI: 10.1109/COMST.2019.2896380, URL: https://ieeexplore.ieee.org/document/8629941/.

[Che+20] Zhuo Chen, Na Lv, *et al.*, « Intrusion Detection for Wireless Edge Networks Based on Federated Learning », *in*: *IEEE Access* 8 (2020), pp. 217463–217472, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2020.3041793, URL: https://ieeexplore.ieee.org/document/9274294/ (visited on 10/25/2021).

[CZY20] Yang Chen, Junzhe Zhang, and Chai Kiat Yeo, « Network Anomaly Detection Using Federated Deep Autoencoding Gaussian Mixture Model », *in*: *Machine Learning for Networking*, ed. by Selma Boumerdassi, Éric Renault, and Paul Mühlethaler, Cham: Springer International Publishing, 2020, pp. 1–14, ISBN: 978-3-030-45778-5.

[DAF18] Rohan Doshi, Noah Apthorpe, and Nick Feamster, « Machine Learning DDoS Detection for Consumer Internet of Things Devices », *in*: *2018 IEEE Security and Privacy Workshops (SPW)* (Ml Apr. 11, 2018), pp. 29–35, DOI: 10.1109/SPW.2018.00013, URL: https://ieeexplore.ieee.org/document/8424629/.

[dCos+19] Kelton A.P. da Costa *et al.*, « Internet of Things: A Survey on Machine Learning-Based Intrusion Detection Approaches », *in*: *Computer Networks* 151 (Mar. 2019), pp. 147–157, ISSN: 13891286, DOI: 10.1016/j.comnet.2019.01.023, URL: https://doi.org/10.1016/j.comnet.2019.01.023.

[Dra+16] Gerard Draper-Gil *et al.*, « Characterization of Encrypted and VPN Traffic Using Time-related Features: » *in*: *Proceedings of the 2nd International Conference on Information Systems Security and Privacy*, 2nd International Conference on Information Systems Security and Privacy, Rome, Italy: SCITEPRESS - Science and and Technology Publications, 2016, pp. 407–414, ISBN: 978-989-758-167-0, DOI: 10.5220/0005740704070414, URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005740704070414 (visited on 10/15/2021).

[Duy+21] Phan The Duy *et al.*, « Federated Learning-Based Intrusion Detection in SDN-enabled IIoT Networks », *in*: *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, 2021 8th NAFOSTED Conference on Information and Computer Science (NICS), Dec. 2021, pp. 424–429, DOI: 10.1109/NICS54270.2021.9701525, URL: https://ieeexplore.ieee.org/abstract/document/9701525 (visited on 04/12/2024).

[Fan+20] Yulin Fan *et al.*, « IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT », *in*: *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE)*, 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE), Guangzhou, China: IEEE, Dec. 2020, pp. 88–95, ISBN: 978-1-66540-396-2, DOI: 10.1109/BigDataSE50710.2020.00020, URL: https://ieeexplore.ieee.org/document/9343358/ (visited on 10/04/2021).

[FNS22] Elena Fedorchenko, Evgenia Novikova, and Anton Shulepov, « Comparative Review of the Intrusion Detection Systems Based on Federated Learning: Advantages and Open Challenges », *in*: *Algorithms* 15.7 (7 July 2022), p. 247, ISSN: 1999-4893, DOI: 10.3390/a15070247, URL: https://www.mdpi.com/1999-4893/15/7/247 (visited on 04/24/2024).

[FYB20] Clement Fung, Chris J.M. M Yoon, and Ivan Beschastnikh, « The Limitations of Federated Learning in Sybil Settings », *in*: *23rd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2020)*, San Sebastian: {USENIX} Association, Oct. 2020, pp. 301–316, ISBN: 978-1-939133-18-2, URL: https://www.usenix.org/conference/raid2020/presentation/fung.

[GR22] Bimal Ghimire and Danda B. Rawat, « Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things », *in*: *IEEE Internet of Things Journal* 9.11 (June 2022), pp. 8229–8249, ISSN: 2327-4662, DOI: 10.1109/JIOT.2022.3150363, URL: https://ieeexplore.ieee.org/abstract/document/9709603 (visited on 04/12/2024).

[Guo+23] Wei Guo *et al.*, « A New Federated Learning Model for Host Intrusion Detection System Under Non-IID Data », *in*: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2023, pp. 494–500, DOI: 10.1109/SMC53992.2023.10393972, URL: https://ieeexplore.ieee.org/document/10393972 (visited on 06/11/2024).

[Hab+17] Arash Habibi Lashkari *et al.*, « Characterization of Tor Traffic Using Time Based Features: » *in*: *Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, 3rd International Conference on Information Systems Security and Privacy, Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 253–262, ISBN: 978-989-758-209-7, DOI: 10.5220/0006105602530262, URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006105602530262 (visited on 10/15/2021).

[Hai+01] J W Haines *et al.*, « 1999 DARPA Intrusion Detection Evaluation: Design and Procedures », *in*: (2001), p. 188.

[Har+17] Stephen Hardy *et al.*, « Private Federated Learning on Vertically Partitioned Data via Entity Resolution and Additively Homomorphic Encryption », Nov. 28, 2017, arXiv: 1711.10677 [cs], URL: http://arxiv.org/abs/1711.10677 (visited on 10/04/2021).

[Hei+20] Xinhong Hei *et al.*, « A Trusted Feature Aggregator Federated Learning for Distributed Malicious Attack Detection », *in*: *Computers & Security* 99 (Dec. 2020), p. 102033, ISSN: 01674048, DOI: 10.1016/j.cose.2020.102033, URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404820303060 (visited on 10/25/2021).

[Hel+22] Stijn Heldens *et al.*, « Litstudy: A Python Package for Literature Reviews », *in*: *SoftwareX* 20 (Dec. 1, 2022), p. 101207, ISSN: 2352-7110, DOI: 10.1016/j.softx.2022.101207, URL: https://www.sciencedirect.com/science/article/pii/S235271102200125X (visited on 06/06/2024).

[Hid18] Ochiai Hideya, *LAN-Security Monitoring Project*, Whitepaper, 2018, URL: https://lan-security.net/whitepaper.pdf (visited on 10/22/2021).

[Ism+24] Umar Audi Isma'ila *et al.*, « Review on Approaches of Federated Modeling in Anomaly-Based Intrusion Detection for IoT Devices », *in*: *IEEE Access* 12 (2024), pp. 30941–30961, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2024.3369915, URL: https://ieeexplore.ieee.org/document/10445150 (visited on 04/24/2024).

[Joh+16] Alistair E.W. Johnson *et al.*, « MIMIC-III, a Freely Accessible Critical Care Database », *in*: *Scientific Data* 3.*1* (May 24, 2016), p. 160035, ISSN: 2052-4463, DOI: 10.1038/sdata.2016.35, URL: https://doi.org/10.1038/sdata.2016.35.

[KC03] J.O. Kephart and D.M. Chess, « The Vision of Autonomic Computing », *in*: *Computer* 36.*1* (Jan. 2003), pp. 41–50, ISSN: 0018-9162, DOI: 10.1109/MC.2003.1160055, URL: http://ieeexplore.ieee.org/document/1160055/.

[KC07] B. Kitchenham and S Charters, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, EBSE-2007-01, 2007.

[Kes07] S. Keshav, « How to Read a Paper », *in*: *acm special interest group on data communication* 37.*3* (2007), pp. 83–84.

[KGS21] Muah Kim, Onur Gunlu, and Rafael F Schaefer, « Federated Learning with Local Differential Privacy: Trade-offs between Privacy, Utility, and Communication », *in*: (2021).

[Kim+20] Seongwoo Kim, He Cai, *et al.*, « Collaborative Anomaly Detection for Internet of Things Based on Federated Learning », *in*: *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, 2020 IEEE/CIC International Conference on Communications in China (ICCC), Chongqing, China: IEEE, Aug. 9, 2020, pp. 623–628, ISBN: 978-1-72817-327-6, DOI: 10.1109/ICCC49849.2020.9238913, URL: https://ieeexplore.ieee.org/document/9238913/ (visited on 10/25/2021).

[Kol+16] Constantinos Kolias *et al.*, « Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset », *in*: *IEEE Communications Surveys & Tutorials* 18.*1* (2016), pp. 184–208, ISSN: 1553-877X, DOI: 10.1109/COMST.2015.2402161, URL: http://ieeexplore.ieee.org/document/7041170/.

[Kon+16a] Jakub Konený, H. Brendan McMahan, Daniel Ramage, *et al.*, « Federated Optimization: Distributed Machine Learning for On-Device Intelligence », *in*: (Oct. 8, 2016), pp. 1–38, URL: http://arxiv.org/abs/1610.02527.

[Kon+16b] Jakub Konený, H. Brendan McMahan, Felix X. Yu, *et al.*, « Federated Learning: Strategies for Improving Communication Efficiency », *in*: (Oct. 18, 2016), pp. 1–10, URL: http://arxiv.org/abs/1610.05492.

[Kor+19] Nickolaos Koroniotis *et al.*, « Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset », *in*: *Future Generation Computer Systems* 100 (Nov. 2019), pp. 779–796, ISSN: 0167739X, DOI: 10.1016/j.future.2019.05.041, URL: https://linkinghub.elsevier.com/retrieve/pii/S0167739X18327687 (visited on 10/23/2021).

[Lav+22a] Leo Lavaur, Benjamin Coste, *et al.*, « Federated Learning as Enabler for Collaborative Security between Not Fully-Trusting Distributed Parties », *in*: *Proceedings of the 29th Computer & Electronics Security Application Rendezvous (C&ESAR): Ensuring Trust in a Decentralized World*, 2022, pp. 65–80, URL: http://ceur-ws.org/Vol-3329/paper-04.pdf.

[Lav+22b] Leo Lavaur, Marc-Oliver Pahl, *et al.*, « The Evolution of Federated Learning-based Intrusion Detection and Mitigation: A Survey », *in*: *IEEE Transactions on Network and Service Management*, Special Issue on Network Security Management (June 2022).

[Li+20a] Beibei Li, Yuhao Wu, *et al.*, « DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems », *in*: *IEEE Transactions on Industrial Informatics* 3203.*c* (2020), pp. 1–1, ISSN: 1551-3203,

DOI: 10.1109/TII.2020.3023430, URL: https://ieeexplore.ieee.org/document/9195012/.

[Li+20b] Kun Li, Huachun Zhou, *et al.*, « Distributed Network Intrusion Detection System in Satellite-Terrestrial Integrated Networks Using Federated Learning », *in*: *IEEE Access* 8 (2020), pp. 214852–214865, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2020.3041641, URL: https://ieeexplore.ieee.org/document/9274426/ (visited on 10/25/2021).

[Liu+21] Hong Liu *et al.*, « Blockchain and Federated Learning for Collaborative Intrusion Detection in Vehicular Edge Computing », *in*: *IEEE Transactions on Vehicular Technology* 70.*6* (June 2021), pp. 6073–6084, ISSN: 0018-9545, 1939-9359, DOI: 10.1109/TVT.2021.3076780, URL: https://ieeexplore.ieee.org/document/9420262/ (visited on 10/04/2021).

[Lo+21] Sin Kit Lo *et al.*, « A Systematic Literature Review on Federated Machine Learning: From A Software Engineering Perspective », *in*: *ACM Computing Surveys* 54.*5* (June 2021), pp. 1–39, ISSN: 0360-0300, 1557-7341, DOI: 10.1145/3450288, arXiv: 2007.11354, URL: http://arxiv.org/abs/2007.11354 (visited on 10/04/2021).

[LYY20] Lingjuan Lyu, Han Yu, and Qiang Yang, « Threats to Federated Learning: A Survey », *in*: *arXiv* (Mar. 4, 2020), URL: http://arxiv.org/abs/2003.02133.

[Mar+19] Samuel Marchal *et al.*, « AuDI: Toward Autonomous IoT Device-Type Identification Using Periodic Communication », *in*: *IEEE Journal on Selected Areas in Communications* 37.*6* (June 2019), pp. 1402–1412, ISSN: 0733-8716, 1558-0008, DOI: 10.1109/JSAC.2019.2904364, URL: https://ieeexplore.ieee.org/document/8664655/ (visited on 06/04/2021).

[McM+17] Brendan McMahan *et al.*, « Communication-Efficient Learning of Deep Networks from Decentralized Data », *in*: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ed. by Aarti Singh and Jerry Zhu, vol. 54, Proceedings of Machine Learning Research, PMLR, Apr. 20–22, 2017, pp. 1273–1282, URL: https://proceedings.mlr.press/v54/mcmahan17a.html.

[Mei+18] Yair Meidan *et al.*, « N-BaIoT: Network-based Detection of IoT Botnet Attacks Using Deep Autoencoders », *in*: *IEEE Pervasive Computing* 17.*3* (July 2018), pp. 12–22, ISSN: 1536-1268, 1558-2590, DOI: 10.1109/MPRV.2018.03367731, arXiv: 1805.03409, URL: http://arxiv.org/abs/1805.03409 (visited on 10/23/2021).

[Men+18] Weizhi Meng *et al.*, « When Intrusion Detection Meets Blockchain Technology: A Review », *in*: *IEEE Access* 6 (2018), pp. 10179–10188, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2018.2799854, URL: http://ieeexplore.ieee.org/document/8274922/.

[MG14] Thomas Morris and Wei Gao, « Industrial Control System Traffic Data Sets for Intrusion Detection Research », *in*: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ed. by Eduardo Bayro-Corrochano and Edwin Hancock, vol. 8827, Cham: Springer International Publishing, 2014, pp. 65–78, ISBN: 978-3-319-12567-1 978-3-319-12568-8, DOI: 10.1007/978-3-662-45355-1_5, URL: http://link.springer.com/10.1007/978-3-662-45355-1_5 (visited on 06/10/2021).

[Mot+21a] Viraaji Mothukuri, Prachi Khare, *et al.*, « Federated Learning-based Anomaly Detection for IoT Security Attacks », *in*: *IEEE Internet of Things Journal* (2021), pp. 1–1, ISSN: 2327-4662, DOI: 10/gmhhmw.

[Mot+21b] Viraaji Mothukuri, Reza M. Parizi, *et al.*, « A Survey on Security and Privacy of Federated Learning », *in*: *Future Generation Computer Systems* 115 (Feb. 2021), pp. 619–640, ISSN: 0167739X, DOI: 10.1016/j.future.2020.10.007, URL: https://doi.org/10.1016/j.future.2020.10.007.

[MS15] Nour Moustafa and Jill Slay, « UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set) », *in*: *2015 Military Communications and Information Systems Conference (MilCIS)*, 2015 Military Communications and Information Systems Conference (MilCIS), Nov. 2015, pp. 1–6, DOI: 10.1109/MilCIS.2015.7348942, URL: https://ieeexplore.ieee.org/abstract/document/7348942 (visited on 10/09/2023).

[Nat24] National Institute of Standards and Technology, *The NIST Cybersecurity Framework (CSF) 2.0*, NIST CSWP 29, Gaithersburg, MD: National Institute of Standards and Technology, Feb. 26, 2024, NIST CSWP 29, DOI: 10.6028/NIST.CSWP.29, URL: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf (visited on 05/23/2024).

[Ngu+19] Thien Duc Nguyen, Samuel Marchal, *et al.*, « DÏoT: A Federated Self-learning Anomaly Detection System for IoT », *in*: *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, vol. 2019-July, IEEE, July 2019, pp. 756–767, ISBN: 978-1-72812-519-0, DOI: 10.1109/ICDCS.2019.00080, URL: https://ieeexplore.ieee.org/document/8884802/.

[Ngu+20] Thien Duc Nguyen, Phillip Rieger, Markus Miettinen, *et al.*, « Poisoning Attacks on Federated Learning-based IoT Intrusion Detection System », *in*: *Proceedings 2020 Workshop on Decentralized IoT Systems and Security*, Workshop on Decentralized IoT Systems and Security, San Diego, CA: Internet Society, 2020, ISBN: 978-1-891562-64-8, DOI: 10.14722/diss.2020.23003, URL: https://www.ndss-symposium.org/wp-content/uploads/2020/04/diss2020-23003-paper.pdf (visited on 01/29/2024).

[Ngu+21] Thien Duc Nguyen, Phillip Rieger, Hossein Yalame, *et al.*, « FLGUARD: Secure and Private Federated Learning », Jan. 21, 2021, arXiv: 2101.02281 [cs], URL: http://arxiv.org/abs/2101.02281 (visited on 05/18/2021).

[Ngu+22] Thien Duc Nguyen, Phillip Rieger, Huili Chen, *et al.*, « FLAME: Taming Backdoors in Federated Learning », *in*: 31st USENIX Security Symposium (USENIX Security 22), 2022, pp. 1415–1432, ISBN: 978-1-939133-31-1, URL: https://www.usenix.org/conference/usenixsecurity22/presentation/nguyen (visited on 03/06/2024).

[PA18] Marc-Oliver Pahl and Francois Xavier Aubet, « All Eyes on You: Distributed Multi-Dimensional IoT Microservice Anomaly Detection », *in*: *14th International Conference on Network and Service Management, CNSM 2018, 1st Workshop on Segment Routing and Service Function Chaining* (2018), pp. 72–80.

[Pop+21a] Segun I. Popoola, Ruth Ande, *et al.*, « Federated Deep Learning for Zero-Day Botnet Attack Detection in IoT Edge Devices », *in*: *IEEE Internet of Things Journal* (2021), pp. 1–1, ISSN: 2327-4662, 2372-2541, DOI: 10.1109/JIOT.2021.3100755, URL: https://ieeexplore.ieee.org/document/9499122/ (visited on 10/01/2021).

[Pop+21b] Segun I. Popoola, Guan Gui, *et al.*, « Federated Deep Learning for Collaborative Intrusion Detection in Heterogeneous Networks », *in*: *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Sept. 2021, pp. 1–6, DOI: 10.1109/VTC2021-Fall52928.2021.9625505.

[Pop+22] Segun Popoola, Bamidele Adebisi, *et al.*, *Optimizing Deep Learning Model Hyperparameters for Botnet Attack Detection in IoT Networks*, preprint, Apr. 8, 2022, DOI: 10.36227/techrxiv.19501885.v1, URL: https://www.techrxiv.org/articles/preprint/Optimizing_Deep_Learning_Model_Hyperparameters_for_Botnet_Attack_Detection_in_IoT_Networks/19501885/1 (visited on 04/14/2022).

[Pop+23] Segun I. Popoola, Agbotiname L. Imoize, *et al.*, « Federated Deep Learning for Intrusion Detection in Consumer-Centric Internet of Things », *in*: *IEEE Transactions on Consumer Electronics* (2023), pp. 1–1, ISSN: 1558-4127, DOI: 10.1109/TCE.2023.3347170, URL: https://ieeexplore.ieee.org/abstract/document/10373897 (visited on 04/12/2024).

[PZ19] Ali Pala and Jun Zhuang, « Information Sharing in Cybersecurity: A Review », *in*: *Decision Analysis* 16.*3* (Sept. 2019), pp. 172–196, ISSN: 1545-8490, DOI: 10.1287/deca.2018.0387, URL: http://pubsonline.informs.org/doi/10.1287/deca.2018.0387.

[Qin+20] Qiaofeng Qin, Konstantinos Poularakis, *et al.*, « Line-Speed and Scalable Intrusion Detection at the Network Edge via Federated Learning », *in*: *2020 IFIP Networking Conference (Networking)*, 2020 IFIP Networking Conference (Networking), June 2020, pp. 352–360, URL: https://ieeexplore.ieee.org/document/9142704 (visited on 04/12/2024).

[QK21] Yang Qin and Masaaki Kondo, « Federated Learning-Based Network Intrusion Detection with a Feature Selection Approach », *in*: *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia: IEEE, June 12, 2021, pp. 1–6, ISBN: 978-1-66543-897-1, DOI: 10.1109/ICECCE52056.2021.9514222, URL: https://ieeexplore.ieee.org/document/9514222/ (visited on 10/04/2021).

[Quy+22] Nguyen Huu Quyen *et al.*, « Federated Intrusion Detection on Non-IID Data for IIoT Networks Using Generative Adversarial Networks and Reinforcement Learning », *in*: *Information Security Practice and Experience*, ed. by Chunhua Su, Dimitris Gritzalis, and Vincenzo Piuri, Cham: Springer International Publishing, 2022, pp. 364–381, ISBN: 978-3-031-21280-2, DOI: 10.1007/978-3-031-21280-2_20.

[Rah+20] Sawsan Abdul Rahman *et al.*, « Internet of Things Intrusion Detection: Centralized, On-Device, or Federated Learning? », *in*: *IEEE Network* 34.*6* (Nov. 2020), pp. 310–317, ISSN: 0890-8044, 1558-156X, DOI: 10.1109/MNET.011.2000286, URL: https://ieeexplore.ieee.org/document/9183799/ (visited on 06/01/2021).

[RWP19] Shailendra Rathore, Byung Wook Kwon, and Jong Hyuk Park, « BlockSecIoTNet: Blockchain-based Decentralized Security Architecture for IoT Network », *in*: *Journal of Network and Computer Applications* 143 (December 2018 Oct. 2019), pp. 167–177, ISSN: 10848045, DOI: 10.1016/j.jnca.2019.06.019, URL: https://doi.org/10.1016/j.jnca.2019.06.019.

[SEO21] Yuwei Sun, Hiroshi Esaki, and Hideya Ochiai, « Adaptive Intrusion Detection in the Networking of Large-Scale LANs With Segmented Federated Learning », *in*: *IEEE Open Journal of the Communications Society* 2 (2021), pp. 102–112, ISSN: 2644-125X, DOI: 10.1109/OJCOMS.2020.3044323, URL: https://ieeexplore.ieee.org/document/9296578/ (visited on 10/04/2021).

[She+20] Sheng Shen *et al.*, « From Distributed Machine Learning To Federated Learning: In The View Of Data Privacy And Security », *in*: *Concurrency and Computation: Practice and Experience* (Sept. 23, 2020), cpe.6002, ISSN: 1532-0626, 1532-0634, DOI: 10.1002/cpe.6002, arXiv: 2010.09258, URL: http://arxiv.org/abs/2010.09258 (visited on 10/04/2021).

[SHG18] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, « Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization », *in*: *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 4th International Conference on Information Systems Security and Privacy, Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 108–116, ISBN: 978-989-758-282-0, DOI: 10.5220/0006639801080116, URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006639801080116 (visited on 10/14/2021).

[Sig99] SigKDD, *KDD Cup 1999 Dataset*, 1999, URL: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (visited on 04/26/2021).

[SOE20] Yuwei Sun, Hideya Ochiai, and Hiroshi Esaki, « Intrusion Detection with Segmented Federated Learning for Large-Scale Multiple LANs », *in*: *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom: IEEE, July 2020, pp. 1–8, ISBN: 978-1-72816-926-2, DOI: 10.1109/IJCNN48605.2020.9207094, URL: https://ieeexplore.ieee.org/document/9207094/ (visited on 10/01/2021).

[SSF16] Florian Skopik, Giuseppe Settanni, and Roman Fiedler, « A Problem Shared Is a Problem Halved: A Survey on the Dimensions of Collective Cyber Defense through Security Information Sharing », *in*: *Computers & Security* 60 (2016), pp. 154–176, ISSN: 01674048, DOI: 10.1016/j.cose.2016.04.003.

[ST19] William Schneble and Geethapriya Thamilarasu, « Attack Detection Using Federated Learning in Medical Cyber-Physical Systems », *in*: *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, International Conference on Computer Communications and Networks, Aug. 2019.

[Sun+20] Wen Sun, Shiyu Lei, *et al.*, « Adaptive Federated Learning and Digital Twin for Industrial Internet of Things », Oct. 31, 2020, arXiv: 2010.13058 [cs], URL: http://arxiv.org/abs/2010.13058 (visited on 10/04/2021).

[Tav+09] Mahbod Tavallaee *et al.*, « A Detailed Analysis of the KDD CUP 99 Data Set », *in*: *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, IEEE, July 2009, pp. 1–6, ISBN: 978-1-4244-3763-4, DOI: 10.1109/CISDA.2009.5356528, URL: http://ieeexplore.ieee.org/document/5356528/.

[Thi+22] Huynh Thai Thi *et al.*, « Federated Learning-Based Cyber Threat Hunting for APT Attack Detection in SDN-Enabled Networks », *in*: *2022 21st International Symposium on Communications and Information Technologies (ISCIT)*, 2022 21st International Symposium on Communications and Information Technologies (ISCIT), Sept. 2022, pp. 1–6, DOI: 10.1109/ISCIT55906.2022.9931222, URL: https://ieeexplore.ieee.org/abstract/document/9931222 (visited on 04/12/2024).

[TKM20] Mineto Tsukada, Masaaki Kondo, and Hiroki Matsutani, « A Neural Network-Based On-device Learning Anomaly Detector for Edge Devices », *in*: *IEEE Transactions on Computers* (2020), pp. 1–1, ISSN: 0018-9340, 1557-9956, 2326-3814, DOI: 10.1109/TC.2020.2973631, URL: https://ieeexplore.ieee.org/document/9000710/ (visited on 10/23/2021).

[TR18] Wiem Tounsi and Helmi Rais, « A Survey on Technical Threat Intelligence in the Age of Sophisticated Cyber Attacks », *in*: *Computers & Security* 72 (Jan. 2018), pp. 212–233, ISSN: 01674048, DOI: 10.1016/j.cose.2017.09.001, URL: https://doi.org/10.1016/j.cose.2017.09.001.

[VKF15] Emmanouil Vasilomanolakis, Shankar Karuppayah, and Mathias Fischer, « Taxonomy and Survey of Collaborative Intrusion Detection », *in*: *ACM Computing Surveys* 47.4 (May 2015), p. 33, DOI: 10.1145/2716260.

[Vy+21] Nguyen Chi Vy *et al.*, « Federated Learning-Based Intrusion Detection in the Context of IIoT Networks: Poisoning Attack and Defense », *in*: *Network and System Security*, ed. by Min Yang, Chao Chen, and Yang Liu, vol. 13041, Cham: Springer International Publishing, 2021, pp. 131–147, ISBN: 978-3-030-92707-3 978-3-030-92708-0, DOI: 10.1007/978-3-030-92708-0_8, URL: https://link.springer.com/10.1007/978-3-030-92708-0_8 (visited on 03/05/2024).

[Wag+19] Thomas D. Wagner *et al.*, « Cyber Threat Intelligence Sharing: Survey and Research Directions », *in*: *Computers & Security* 87 (2019), p. 101589, ISSN: 01674048, DOI: 10.1016/j.cose.2019.101589.

[Wag19] Thomas D Wagner, « Cyber Threat Intelligence for “Things” », *in*: *2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, IEEE, June 2019, pp. 1–2, ISBN: 978-1-72810-232-0, DOI: 10.1109/CyberSA.2019.8899384, URL: https://ieeexplore.ieee.org/document/8899384/.

[Yad19] Omry Yadan, *Hydra - A Framework for Elegantly Configuring Complex Applications*, Github, 2019, URL: https://github.com/facebookresearch/hydra.

[Yan+19] Qiang Yang, Yang Liu, *et al.*, « Federated Machine Learning: Concept and Applications », *in*: *ACM Transactions on Intelligent Systems and Technology* 10.*2* (Feb. 28, 2019), pp. 1–19, ISSN: 2157-6904, DOI: 10.1145/3298981, URL: https://dl.acm.org/doi/10.1145/3298981.

[Yan+23] Run Yang, Hui He, *et al.*, « Dependable Federated Learning for IoT Intrusion Detection against Poisoning Attacks », *in*: *Computers & Security* 132 (Sept. 1, 2023), p. 103381, ISSN: 0167-4048, DOI: 10.1016/j.cose.2023.103381, URL: https://www.sciencedirect.com/science/article/pii/S0167404823002912 (visited on 03/04/2024).

[Zha+19] Ying Zhao *et al.*, « Multi-Task Network Anomaly Detection Using Federated Learning », *in*: *Proceedings of the Tenth International Symposium on Information and Communication Technology - SoICT 2019*, The Tenth International Symposium, Hanoi, Ha Long Bay, Viet Nam: ACM Press, 2019, pp. 273–279, ISBN: 978-1-4503-7245-9, DOI: 10.1145/3368926.3369705, URL: http://dl.acm.org/citation.cfm?doid=3368926.3369705 (visited on 06/07/2021).

[Zha+20] Weishan Zhang *et al.*, « Blockchain-Based Federated Learning for Device Failure Detection in Industrial IoT », *in*: *IEEE Internet of Things Journal* (Sept. 6, 2020), pp. 1–1, ISSN: 2327-4662, DOI: 10.1109/JIOT.2020.3032544, URL: https://ieeexplore.ieee.org/document/9233457/.

[ZLK10] Chenfeng Vincent Zhou, Christopher Leckie, and Shanika Karunasekera, « A Survey of Coordinated Attacks and Collaborative Intrusion Detection », *in*: *Computers & Security* 29.*1* (Feb. 2010), pp. 124–140, ISSN: 01674048, DOI: 10.1016/j.cose.2009.06.008, URL: https://linkinghub.elsevier.com/retrieve/pii/S016740480900073X (visited on 07/21/2021).

List of Figures

Appendices

A Additional figures

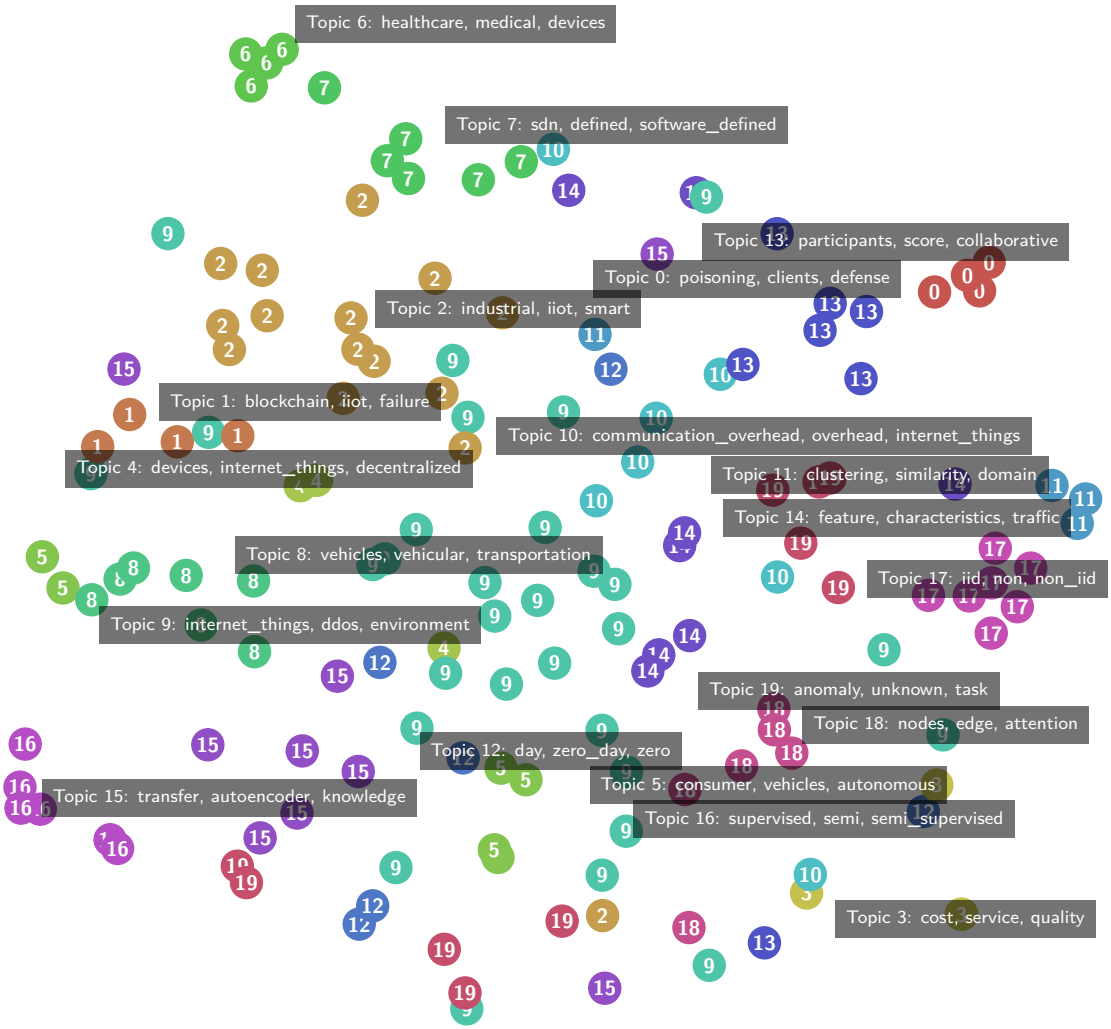


Figure 9.1 – Topic embedding of the FIDS literature using a NMF model with 20 topics. Each point represents a paper, and each are labelled with the topic they are the most associated with.

B Résumé en français de la thèse

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : L'Apprentissage Fédéré comme Outil pour la Détection Collaborative d'Intrusions

Mot clés : apprentissage automatique, apprentissage fédéré, détection d'intrusions, collaboration, confiance

Résumé : La collaboration entre les différents acteurs de la cybersécurité est essentielle pour lutter contre des attaques de plus en plus sophistiquées et nombreuses. Pourtant, les organisations sont souvent réticentes à partager leurs données, par peur de compromettre leur confidentialité, et ce même si cela pourrait d'améliorer leurs modèles de détection d'intrusions. L'apprentissage fédéré est un paradigme récent en apprentissage automatique qui permet à des clients distribués d'entraîner un modèle commun sans partager leurs données. Ces propriétés de collaboration et de confidentialité en font un candidat idéal pour des applications sensibles comme la détection d'intrusions. Si un certain nombre d'applications ont montré qu'il est, en effet, possible d'entraîner un modèle unique sur des données distribuées de détection d'intrusions, peu se sont intéressées à l'aspect collaboratif de ce paradigme. En plus de l'aspect collaboratif, d'autres problématiques apparaissent dans ce contexte, telles que l'hétérogénéité des données des différents participants ou la gestion de participants non fiables. Dans ce manuscrit, nous explorons l'utilisation de l'apprentissage fédéré pour construire des systèmes collaboratifs de détection d'intrusions. En particulier, nous explorons l'impact de la qualité des données dans des contextes hétérogènes, certains types d'attaques par empoisonnement, et proposons des outils et des méthodologies pour améliorer l'évaluation de ce type d'algorithmes distribués.

Title: On Federated Learning as a Framework for Collaborative Intrusion Detection

Keywords: machine learning, federated learning, intrusion detection, collaboration, trust

Abstract: Collaboration between different cybersecurity actors is essential to fight against increasingly sophisticated and numerous attacks. However, stakeholders are often reluctant to share their data, fearing confidentiality and privacy issues, although it would improve their intrusion detection models. Federated learning is a recent paradigm in machine learning that allows distributed clients to train a common model without sharing their data. These properties of collaboration and confidentiality make it an ideal candidate for sensitive applications such as intrusion detection. While several applications have shown that it is indeed possible to train a single model on distributed intrusion detection data, few have focused on the collaborative aspect of this paradigm. In addition to the collaborative aspect, other challenges arise in this context, such as the heterogeneity of the data between different participants or the management of untrusted contributions. In this manuscript, we explore the use of federated learning to build collaborative intrusion detection systems. In particular, we explore the impact of data quality in heterogeneous contexts, some types of poisoning attacks, and propose tools and methodologies to improve the evaluation of these types of distributed algorithms.