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

Apprentissage Fédéré 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 :

Prénom NOM Fonction et établissement d'exercice
Prénom NOM Fonction et établissement d'exercice
Prénom NOM Fonction et établissement d'exercice

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 :	Prénom NOM	Fonction et établissement d'exercice *(à préciser après la soutenance)*
Examinateurs :	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
	Prénom NOM	Fonction et établissement d'exercice
	Fabien AUTREL	Ingénieur de recherche à IMT Atlantique
	Marc-Oliver PAHL	Directeur détude 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

ACKNOWLEDGEMENTS

Table of Contents

LIST OF FIGURES

Part I

Introduction

Part II

Federated Learning to build CIDSs

Chapter 3

State of the Art

The Evolution of Federated-Learning–based Intrusion Detection and Mitigation

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 federated learning (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 intrusion detection system (IDS), the literature on the topic is still scarce. Only a handful of reviews [Ala+21; Agr+21; Cam+21] 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 systematic literature review (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.

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 (at the time of its publication) 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 federated intrusion detection systems (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 main challenges and opportunities in the field, and a set of research directions to address them.

3.2 Related Work

Literature related to FL for IDS can be divided into three main categories: (a) security information sharing, (b) intrusion detection, and (c) federated learning.

When the data collection for this study was completed in late 2021, the literature on FL for IDS was rather scarce. Only a handful of reviews had been published on the topic [Ala+21; Agr+21; Cam+21]. In addition, we extended our search to include works on security-related information sharing [SSF16; TR18; Wag+19; PZ19],

A consequent amount of literature exists on the topic of collaboration and intrusion detection [ZLK10; Vas+17], including the use of machine learning (ML) [BG16; Men+18; Cha+19; dCos+19] and security knowledge sharing [**<empty citation>**]. As FL was a rather novel concept at the time, Finally, the end of 2021 (when data collection for this initial study stopped) also saw the publication of a handful of reviews on FL for IDS specifically.

?? provides a summary of this selection, grouped by topic and sorted by publication date.

Part III

Quantifying the Limitations of FIDSs

Part IV

Providing Solutions

Part V

Conclusion

BIBLIOGRAPHY

[Agr+21] Shaashwat Agrawal et al., « Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions », June 16, 2021, arXiv: 2106.09527 [cs], URL: http://arxiv.org/abs/2106.09527 (visited on 12/02/2021).

[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).

[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.

[Cam+21] Enrique Mármol Campos et al., « Evaluating Federated Learning for Intrusion Detection in Internet of Things: Review and Challenges », Aug. 2, 2021, arXiv: 2108.00974 [cs], URL: http://arxiv.org/abs/2108.00974 (visited on 12/02/2021).

[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/.

[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.

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

[Lav+22a] Leo Lavaur 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*, Nov. 2022, pp. 65–80, URL: http://ceur-ws.org/Vol-3329/paper-04.pdf.

[Lav+22b] Leo Lavaur 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).

[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/.

[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.

[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.

[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.

[Vas+17] Emmanouil Vasilomanolakis et al., « Towards Trust-Aware Collaborative Intrusion Detection: Challenges and Solutions », *in*: *Trust Management XI*, ed. by Jan-Philipp Steghöfer and Babak Esfandiari, IFIP Advances in Information and Communication Technology, Cham: Springer International Publishing, 2017, pp. 94–109, ISBN: 978-3-319-59171-1, DOI: 10.1007/978-3-319-59171-1_8.

[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.

[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).

Appendices

A 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 : titre (en français).....

Mot clés : de 3 à 6 mots clefs

Résumé : Eius populus ab incunabulis primis ad usque pueritiae tempus extremum, quod annis circumcluditur fere trecentis, circummurana pertulit bella, deinde aetatem ingressus adultam post multiplices bellorum aerumnas Alpes transcendit et fretum, in iuvenem erectus et virum ex omni plaga quam orbis ambit inmensus, reportavit laureas et triumphos, iamque vergens in senium et nomine solo aliquotiens vincens ad tranquilliora vitae discessit. Hoc inmaturo interitu ipse quoque sui pertaesus excessit e vita aetatis nono anno atque vicensimo cum quadriennio imperasset. natus apud Tuscos in Massa Veternensi, patre Constantio Constantini fratre imperatoris, matreque Galla. Thalassius vero ea tempestate praefectus praetorio praesens ipse quoque adrogantis ingenii, considerans incitationem eius ad multorum augeri discrimina, non maturitate vel consiliis mitigabat, ut aliquotiens celsae potestates iras principum molliverunt, sed adversando iurgandoque cum parum congrueret, eum ad rabiem potius evibrabat, Augustum actus eius exaggerando creberrime docens, idque, incertum qua mente, ne lateret adfectans. quibus mox Caesar acrius efferatus, velut contumaciae quoddam vexillum altius erigens, sine respectu salutis alienae vel suae ad vertenda opposita instar rapidi fluminis irrevocabili impetu ferebatur. Hae duae provinciae bello quondam piratico catervis mixtae praedonum.

Title: titre (en anglais).....

Keywords: de 3 à 6 mots clefs

Abstract: Eius populus ab incunabulis primis ad usque pueritiae tempus extremum, quod annis circumcluditur fere trecentis, circummurana pertulit bella, deinde aetatem ingressus adultam post multiplices bellorum aerumnas Alpes transcendit et fretum, in iuvenem erectus et virum ex omni plaga quam orbis ambit inmensus, reportavit laureas et triumphos, iamque vergens in senium et nomine solo aliquotiens vincens ad tranquilliora vitae discessit. Hoc inmaturo interitu ipse quoque sui pertaesus excessit e vita aetatis nono anno atque vicensimo cum quadriennio imperasset. natus apud Tuscos in Massa Veternensi, patre Constantio Constantini fratre imperatoris, matreque Galla. Thalassius vero ea tempestate praefectus praetorio praesens ipse quoque adrogantis ingenii, considerans incitationem eius ad multorum augeri discrimina, non maturitate vel consiliis mitigabat, ut aliquotiens celsae potestates iras principum molliverunt, sed adversando iurgandoque cum parum congrueret, eum ad rabiem potius evibrabat, Augustum actus eius exaggerando creberrime docens, idque, incertum qua mente, ne lateret adfectans. quibus mox Caesar acrius efferatus, velut contumaciae quoddam vexillum altius erigens, sine respectu salutis alienae vel suae ad vertenda opposita instar rapidi fluminis irrevocabili impetu ferebatur. Hae duae provinciae bello quondam piratico catervis mixtae praedonum.