

# 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 Communica-  
tion*

Par

**Léo LAVAU**

## 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 après la soutenance.

Examineurs : Sonia Ben Mokhtar  
Pierre-François Gimenez  
Vincent Nicomette  
Fabien AUTREL  
Marc-Oliver PAHL

Dir. de thèse : Yann BUSNEL

Directrice de Recherche CNRS au laboratoire LIRIS

Maître de Conférence à CentraleSupélec

Professeur à l'INSA de Toulouse

Ingénieur de Recherche à IMT Atlantique

Directeur d'Étude à IMT Atlantique

Directeur de la Recherche et de l'Innovation (DRI) à IMT Nord Europe

### Invité(s) :

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



## 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

---

<b>Abstracts</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Table of Contents</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Context and Motivation . . . . .	3
1.2 Contributions . . . . .	5
1.3 Outline . . . . .	6
1.4 Publications . . . . .	7
<b>I Federated Learning to build CIDSs</b>	<b>9</b>
<b>2 Preliminaries</b>	<b>11</b>
2.1 Machine Learning for Intrusion Detection . . . . .	11
2.2 Fundamentals of Federated Learning . . . . .	11
2.3 Threats against Federated Learning . . . . .	11
<b>3 State of the Art</b>	<b>13</b>
3.1 Introduction and Motivation . . . . .	13
3.2 Methodology . . . . .	14
3.3 Quantitative Analysis . . . . .	18
3.4 Qualitative Analysis . . . . .	23
3.5 Related Work . . . . .	38
3.6 Discussion . . . . .	41
3.7 Conclusion and takeaways . . . . .	44
<b>4 Application – FIDSs Performance and Limitations</b>	<b>47</b>
<b>II Quantifying the Limitations of FIDSs</b>	<b>49</b>
<b>5 Studying Heterogeneity in Distributed Intrusion Detection with Topology Generation</b>	<b>51</b>

<b>6</b>	<b>Assessing the Impact of Label-Flipping Attacks on FL-based IDSs</b>	<b>53</b>
<b>III</b>	<b>Providing Solutions</b>	<b>55</b>
<b>7</b>	<b>Model Quality Assessment for Reputation-aware Collaborative Federated Learning</b>	<b>57</b>
<b>8</b>	<b>Solutions for the Future of FIDSs</b>	<b>59</b>
<b>9</b>	<b>Conclusion</b>	<b>61</b>
	<b>Bibliography</b>	<b>63</b>
	<b>List of Figures</b>	<b>63</b>
	<b>List of Tables</b>	<b>65</b>
	<b>Appendices</b>	<b>87</b>
	A   Additional figures . . . . .	87
	B   Résumé en français de la thèse . . . . .	87
	<b>Glossary</b>	<b>88</b>





PART I

# Federated Learning to build CIDSs

---



# PRELIMINARIES

---

## Contents

2.1	Machine Learning for Intrusion Detection . . . . .	11
2.2	Fundamentals of Federated Learning . . . . .	11
2.3	Threats against Federated Learning . . . . .	11

---

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







PART II

# Quantifying the Limitations of FIDSs

---









PART III

# Providing Solutions

---











# LIST OF FIGURES

---

1.1	Illustration of Federated Learning (FL) in a Collaborative IDS (CIDS) use case. . . . .	5
3.1	Search and selection processes. . . . .	15
3.2	Updated selection process. . . . .	17
3.3	Evolution of the topics and number of publications. . . . .	18
3.4	Distribution of the publications in the most recurring venues. . . . .	19
3.5	Distribution of the publications by affiliation. . . . .	20
3.6	Distribution of the publications by author and country. . . . .	21
3.7	Topics of interest in the field of Federated Intrusion Detection Systems (FIDSs). . . . .	22
3.8	Exploiting the topics of interest. . . . .	22
3.9	The proposed reference architecture for FIDSs. . . . .	24
3.10	Proposed taxonomy for FIDS. . . . .	26
9.1	Topic embedding of the FIDS literature using a Non-negative Matrix Factorization (NMF) model with 20 topics. Each point represents a paper, and each are labelled with the topic they are the most associated with. . . . .	87



# LIST OF TABLES

---

3.1	Comparative overview of selected works in the original study—approach and objectives (1/2). . . . .	27
3.2	Comparative overview of selected works in the original study—algorithms and performance (2/2). . . . .	32
3.3	Related literature reviews, their topics, contributions, and number of citations. . . . .	39







---

**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.