

HEC MONTRÉAL

**Amélioration du jeu de données CERT Insider Threat
pour la détection proactive des menaces internes malveillantes**

**par
Mariam Fatima Coulibaly**

**Sciences de la gestion
(Spécialisation Intelligence d'affaires)**

*Mémoire présenté en vue de l'obtention
du grade de maîtrise en intelligence d'affaires
(M. Sc.)*

Alina Dulipovici
HEC Montréal
Directrice de recherche

Denis Larocque
HEC Montréal
Co-directeur de recherche

Décembre 2025
© Mariam Fatima Coulibaly, 2025

CERTIFICAT D'APPROBATION ÉTHIQUE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait aux exigences de notre politique en cette matière.

Projet # : 2026-6541

Titre du projet de recherche : Amélioration du jeu de données CERT pour la détection des menaces internes

Chercheur principal : Mariam Fatima Coulibaly

Directeur/codirecteurs : Alina Maria Dulipovici ; Denis Larocque, Professeur - HEC Montréal

Date d'approbation du projet : 11 septembre 2025

Date d'entrée en vigueur du certificat : 11 septembre 2025

Date d'échéance du certificat : 01 septembre 2026



Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2025-09-11 à 16:54

Sommaire

La détection des menaces internes malveillantes constitue aujourd'hui un enjeu majeur en cybersécurité, en raison à la fois de la gravité potentielle de leurs impacts et de la complexité des comportements humains qui les sous-tendent. Contrairement aux attaques externes, ces menaces prennent naissance au sein même des organisations et s'inscrivent généralement dans des dynamiques progressives, ambiguës et fortement dépendantes du contexte organisationnel. Cette complexité limite l'efficacité des approches de détection traditionnelles, souvent fondées sur des modèles statiques ou sur des jeux de données simulés ne capturant qu'imparfaitement la réalité des environnements professionnels.

Dans ce contexte, le jeu de données CERT Insider Threat s'est imposé dans la littérature comme un banc d'essai de référence pour l'évaluation de méthodes de détection des menaces internes malveillantes. Toutefois, son utilisation soulève des critiques récurrentes concernant le réalisme des comportements simulés et la faible prise en compte des dimensions contextuelles propres aux organisations réelles

Ce mémoire a pour premier objectif d'évaluer la capacité du jeu de données CERT à soutenir des approches de détection proactive fondées sur l'analyse fonctionnelle des données. Le deuxième objectif vise à examiner la cohérence entre les comportements simulés dans CERT et ceux observés en milieu organisationnel réel, à partir d'une analyse qualitative menée auprès de professionnels de la cybersécurité. Enfin, le troisième objectif vise à formuler des recommandations concrètes afin d'améliorer la représentativité du jeu de données CERT et de renforcer sa pertinence méthodologique, tant pour la recherche que pour la pratique en cybersécurité.

Mots-clés : menace interne malveillante, CERT, analyse fonctionnelle des données, FPCA, détection proactive, jeux de données simulés, cybersécurité.

Remerciements

Je souhaite tout d'abord exprimer ma profonde gratitude à mes deux co-directeurs de mémoire, Mme Alina Dulipovici et M. Denis Larocque, pour leur accompagnement rigoureux et bienveillant tout au long de ce travail.

Merci à Mme Alina Dulipovici, qui a su orienter et approfondir ma réflexion sur les idées centrales de ce mémoire et sur la manière de les formuler avec clarté et cohérence. Par sa rigueur intellectuelle, ses conseils avisés et son écoute constante, elle m'a permis de structurer ma réflexion, de gagner en confiance et de progresser tant sur le plan scientifique que personnel. Sa bienveillance, sa disponibilité et son exigence ont été des piliers essentiels dans l'aboutissement de ce mémoire. Merci pour votre soutien, votre temps et les apprentissages précieux que vous m'avez transmis.

Merci à M. Denis Larocque, dont la disponibilité, l'engagement et la bienveillance ont marqué mon parcours à la maîtrise. J'ai eu la chance de suivre l'un de ses cours, qui m'a profondément marquée par sa capacité à rendre accessibles des notions complexes et à susciter l'intérêt pour des sujets exigeants. Son approche pédagogique, à la fois claire et humaine, ainsi que ses encouragements tout au long de ce mémoire ont grandement contribué à mon cheminement académique.

Merci à mes amis pour leur patience, leur compréhension et leur soutien constant, notamment lors des périodes plus intenses de rédaction.

Merci chaleureux à mon oncle, Maharouf Oyolola, pour son accompagnement, ses précieux conseils et ses encouragements, qui ont été une véritable source de motivation tout au long de la rédaction de ce mémoire.

Finalement, je souhaite également adresser un merci tout particulier à mon époux, dont le soutien indéfectible, la motivation et la présence ont été essentiels lorsque je doutais ou que je baissais les bras. Je ne saurais tout de même pas terminer ces mots de remerciements sans une mention spéciale à ma famille, pour les prières, les appels de réconfort et l'amour constant qui m'ont accompagnée tout au long de ce parcours. Votre soutien a été une source de force précieuse dans l'aboutissement de ce projet.

Table des matières

Sommaire	3
Remerciements.....	4
Liste des figures	8
Liste des tableaux	9
Liste des abréviations	10
Chapitre 1. Introduction	11
1. Problématique	11
2. Objectifs et question de recherche	15
3. Structure du mémoire	16
Chapitre 2. Revue de la littérature	18
1. La menace interne en cybersécurité.....	19
1.1. Typologies et profils des acteurs internes	20
1.2. Facteurs humains et organisationnels.....	21
1.3. Motivations et comportements	23
1.4 Conclusion	26
2. Approches et méthodes de détection de la menace interne malveillante.....	27
2.1. Approches basées sur les règles et signatures.....	28
2.2. Analyse comportementale et modèles statistiques	29
2.3. Approches basées sur l'apprentissage automatique et intelligence artificielle	30
2.4 Approches basées sur l'analyse fonctionnelle des données (FDA).....	39
2.5 Conclusion	43
3. Les jeux de données	44
3.1. Types de jeux de données pour la détection de la menace interne malveillante.....	46
3.2. Jeux de données de référence dans la littérature	51
3.3. Utilisations du jeu de données CERT dans la littérature.....	61
3.4 Conclusion	66
Chapitre 3. Méthodologie	67
1. Choix de la méthodologie et justifications	67
1.1 Justification du protocole expérimental.....	67
1.2 Justification du recours aux entretiens qualitatifs	68
2. Protocole expérimental sur le jeu de données CERT	69
2.1 Préparation et agrégation des données CERT	70
2.2 Construction des trajectoires fonctionnelles	73
2.3 Réduction dimensionnelle par FPCA	75
2.4 Modèles de classification basés sur les composantes fonctionnelles.....	76
2.5 Critères d'évaluation et interprétation des résultats	80
3. Entrevues auprès d'experts	82

3.1 Sélection des participants	82
3.2 Identification et recrutement des participants	84
3.3 Quelques détails sur le protocole d’entrevue	85
3.4 Quelques détails sur l’entrevue et sa structure	86
4. Analyse des données.....	89
4.1 Transcription et préparation du corpus.....	90
4.2 Codifier les données collectées.....	90
4.3 Analyser les données (intra-répondants et inter-répondants)	94
4.4 Comparer l'analyse à la littérature	94
4.5 Considérations éthiques	95
4.6 Conclusion	98
Chapitre 4. Résultats et Analyse	99
1. Approche quantitative	99
1.1 Description du corpus	99
1.2 Préparation des données et construction des trajectoires fonctionnelles	116
1.3 Analyse en composantes principales fonctionnelles.....	118
1.4 Modélisation et validation croisée.....	120
1.5 Résultats de la régression logistique	122
1.6 Résultats de la forêt aléatoire	125
2. Approche qualitative	127
2.1 Profil des participants et crédibilité du panel	127
2.2 Codage des données.....	128
2.3 Thème : nature et complexité de la menace interne malveillante	129
2.4 Thème : comportements malveillants et profils d'utilisateurs à risque.....	129
2.5 Thème : signaux faibles et limites des outils de détection actuels.....	130
2.6 Thème : perception des jeux de données simulés et mise en perspective du CERT	131
Chapitre 5. Discussion	132
1. Discussion des résultats quantitatifs	133
2. Discussion des résultats qualitatifs	135
Chapitre 6. Conclusion.....	139
1. Retour sur la question de recherche et de l’approche méthodologique.....	139
2. Implications pour les professionnels	143
3. Implications pour la recherche	144
4. Limites.....	147
5. Avenue de recherche	148
Bibliographie	153
Annexes	157
Annexe 1 : Protocole d’entrevue	157
Annexe 2 : Scénarios de menaces internes présentés aux participants	161

Annexe 3 : Distribution horaire des connexions (table Device)	162
Annexe 4 : Distribution des activités Logon et Logoff sur toute la période (table Logon)	163
Annexe 5 : Top 10 des utilisateurs actifs après 20h en fonction du nombre de fichiers envoyés (table File).....	164
Annexe 6 : Trajectoires hebdomadaires brutes et lissées pour la variable total_events_logon	165
Annexe 7 : Effet des composantes principales fonctionnelles sur les trajectoires hebdomadaires (ex : activité logon)	166
Annexe 8 : Variance fonctionnelle cumulée expliquée par les composantes principales pour les douze variables comportementales	167
Annexe 9 : Harmoniques principales fonctionnelles pour trois variables comportementales (FPCA)	168
Annexe 10 : Distribution des seuils de Youden par fold de validation croisée (régression logistique)	169
Annexe 11 : Variabilité des performances du modèle fonctionnel sur l'ensemble des partitions de validation croisée	170
Annexe 12 : Importance relative des composantes principales fonctionnelles dans le modèle de régression logistique.....	171
Annexe 13 : Importance relative des composantes principales fonctionnelles dans la forêt aléatoire	172
Annexe 14 : Déclaration sur l'utilisation de l'intelligence artificielle générative (IAG)..	173

Liste des figures

Figure 4-1 : Distribution hebdomadaire du nombre de connexions (nb_connect)	103
Figure 4-2 : Distribution hebdomadaire du nombre de déconnexions (nb_disconnect)	103
Figure 4-3 : Distribution hebdomadaire du nombre de postes utilisés (nb_pc_device)	104
Figure 4-4 : Distribution hebdomadaire du nombre total d'événements Device (total_events_device)	105
Figure 4-5 : Distribution hebdomadaire du nombre de connexions (nb_logon)	107
Figure 4-6 : Distribution hebdomadaire du nombre de déconnexions (nb_logoff)	107
Figure 4-7 : Distribution hebdomadaire du nombre total d'événements d'authentification (total_events_logon)	108
Figure 4-8 : Distribution hebdomadaire du nombre total de courriels envoyés (total_emails)	110
Figure 4-9 : Distribution hebdomadaire du nombre de pièces jointes envoyées (nb_attachments)	111
Figure 4-10 : Distribution hebdomadaire du nombre de postes utilisés pour les accès aux fichiers (nb_pc_file)	112
Figure 4-11 : Distribution hebdomadaire du nombre total d'événements liés aux fichiers (total_events_file)	113

Liste des tableaux

Tableau 2-1 : Description des principales colonnes du jeu de données Enron Email	
.....	51
Tableau 2-2 : Description des principales colonnes du jeu de données RUU	54
Tableau 2-3 : Description des principales colonnes du jeu de données WUL	56
Tableau 2-4 : Description des principales colonnes du jeu de données CERT	59
Tableau 3-1 : Variables comportementales hebdomadaires retenues pour la FDA	74
Tableau 3-2 : Grille de codification thématique	91
Tableau 4-1 : Statistiques descriptives des variables comportementales hebdomadaires de la table Device	102
Tableau 4-2 : Statistiques descriptives des variables comportementales hebdomadaires de la table Logon	106
Tableau 4-3 : Statistiques descriptives des variables comportementales hebdomadaires de la table Email	109
Tableau 4-4 : Statistiques descriptives des variables comportementales hebdomadaires de la table File	112
Tableau 4-5 : Statistiques descriptives des variables comportementales hebdomadaires de la table http	115
Tableau 4-6 : Comparaison des performances des modèles (FPCA) pour la détection des menaces internes malveillantes	123
Tableau 6-1 : Recommandations structurées pour améliorer la représentativité du jeu de données CERT	141

Liste des abréviations

AE	Autoencoder
AUC	Area Under the Curve
CERT	Computer Emergency Response Team
CISA	Cybersecurity and Infrastructure Security Agency
CNN	Convolutional Neural Network
DARPA	Defense Advanced Research Projects Agency
DL	Deep Learning
ENISA	European Union Agency for Cybersecurity
FDA	Functional Data Analysis
FPCA	Functional Principal Component Analysis
GCN	Graph Convolutional Network
IA	Intelligence artificielle
IoT	Internet of Things
LSTM	Long Short-Term Memory
ML	Machine Learning
MITRE	MITRE Corporation
NIST	National Institute of Standards and Technology
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SIEM	Security Information and Event Management
UEBA	User and Entity Behavior Analytics

Chapitre 1. Introduction

1. Problématique

Les avancées technologiques récentes, notamment dans le domaine de l'intelligence artificielle, du cloud computing et de l'Internet des objets (IoT), transforment profondément la manière dont les organisations gèrent et protègent leurs données. Comme le souligne Jain (2025), cette transformation crée à la fois de nouvelles opportunités et de nouvelles vulnérabilités, amplifiant la surface d'attaque potentielle des systèmes d'information.

Dans ce contexte, la cybersécurité s'impose comme un pilier stratégique. Selon le dictionnaire Larousse, le terme cybersécurité se définit comme, « l'ensemble des techniques destinées à protéger les systèmes informatiques et les données contre toute intrusion, altération ou destruction », regroupant les mesures de prévention, de détection et de réponse face aux événements susceptibles de compromettre la confidentialité, l'intégrité ou la disponibilité des ressources numériques.

Ces événements, communément désignés comme des menaces, sont définis par le NIST (2023) comme « tout événement ou entité susceptible d'exploiter une vulnérabilité et de causer un préjudice ». Les menaces peuvent émaner de l'extérieur de l'organisation (menaces externes) ou de l'intérieur (menaces internes), c'est-à-dire par des individus disposant d'un accès légitime aux ressources, mais pouvant les utiliser de manière abusive (Theoharidou et al., 2005 ; Sinclair & Smith, 2008).

Les menaces internes représentent un type de risque particulièrement insidieux et souvent sous-estimé. Selon la Cybersecurity and Infrastructure Security Agency (CISA, 2023), une menace interne est « un individu autorisé à accéder aux ressources d'une organisation qui utilise cet accès, intentionnellement ou non, pour compromettre la sécurité, la confidentialité, l'intégrité ou la disponibilité des systèmes et données » (CISA, 2023 ; NIST, 2023). Ces acteurs peuvent être des employés, des sous-traitants, des consultants ou des partenaires disposant de privilèges légitimes.

L'impact des menaces internes est souvent plus élevé que celui des menaces externes. Leur proximité avec les ressources critiques et la connaissance approfondie des systèmes leur permettent de causer des dommages significatifs tout en restant difficiles à détecter (CISA, 2023 ; Greitzer & Frincke, 2010). L'ampleur du phénomène est illustrée par ces chiffres : le rapport Insider Threats 2023 du CERT de l'Université Carnegie Mellon estime que les menaces internes représentent près de 60 % des incidents de cybersécurité signalés, avec un coût moyen d'environ 15 millions USD par incident (CERT, 2023). En comparaison, le rapport Cost of a Data Breach 2023 d'IBM évalue le coût moyen d'une violation liée à une menace externe à environ 5 millions USD (IBM, 2023). Ces chiffres soulignent non seulement le potentiel destructeur des menaces internes mais aussi l'importance d'outils et de méthodes de détection proactive capables d'identifier rapidement les comportements à risque.

Les menaces internes peuvent prendre différentes formes : intentionnelles ou non intentionnelles (Cappelli, Mohr & Trzeciak, 2012 ; Greitzer et al., 2013). Les menaces intentionnelles (ou malveillantes) correspondent à des actions délibérées, motivées par la vengeance, le gain financier ou l'espionnage industriel. Les menaces non intentionnelles (ou malhabiles) résultent quant à elles d'erreurs, de négligences ou de mauvaises pratiques, comme la gestion imprudente des accès ou l'inattention aux protocoles de sécurité. Indépendamment de leur type, leur détection est difficile par le fait que les acteurs internes connaissent parfaitement les processus, disposent souvent d'autorisations étendues et peuvent dissimuler leurs actions dans des comportements en apparence légitimes.

Pour y répondre, les organisations mettent en œuvre des stratégies spécifiques : surveillance comportementale, suivi des accès en temps réel, analyse basée sur l'intelligence artificielle, mais aussi formation et sensibilisation des employés (Nasir et al., 2021 ; ENISA, 2022). Ces mesures visent à identifier les comportements suspects, malhabiles ou malveillants, avant qu'un incident ne survienne et à intervenir rapidement pour limiter les dommages. Toutefois, l'efficacité des approches de détection proactive, comme la surveillance comportementale, dépend fortement de la qualité des environnements de test et de l'accès à des données pertinentes et représentatives. Pour entraîner et évaluer des modèles de détection proactive, il est nécessaire de disposer de

jeux de données qui reflètent fidèlement la complexité et la diversité des comportements réels des utilisateurs au sein de l'organisation.

En pratique, l'accès à des données réelles d'incidents internes malveillants est extrêmement limité. Les incidents sont souvent confidentiels et leur diffusion est restreinte par des contraintes de sécurité et de protection des informations sensibles, ce qui empêche le partage libre entre organisations ou au sein de la communauté scientifique (Greitzer et al., 2010 ; Cappelli et al., 2012). De plus, la variabilité des contextes organisationnels rend chaque incident unique, compliquant la création de bases de données comparables ou exhaustives. Cette rareté de données fiables constitue un défi majeur pour la recherche et le développement de méthodes de détection proactive : sans environnements de test réalistes et riches en informations comportementales et contextuelles, il est difficile de valider les modèles et de garantir leur performance dans des situations réelles.

Pour surmonter ces limitations, la communauté scientifique s'appuie sur des jeux de données simulés ou semi-simulés, conçus pour reproduire des scénarios variés de menaces internes tout en préservant la confidentialité. Parmi eux, le jeu de données CERT Insider Threat développé par le Computer Emergency Response Team de l'Université Carnegie Mellon constitue une référence largement adoptée. Il fournit des traces détaillées d'activité utilisateur, malhabile et malveillante, incluant les connexions aux systèmes, la consultation de fichiers, les échanges de courriels et d'autres interactions numériques, ainsi que des événements simulés représentant différents types de menaces internes malveillantes (Glasser & Lindauer, 2013 ; Cappelli et al., 2012). Ces caractéristiques permettent aux chercheurs de tester et d'entraîner des modèles de détection proactive dans un cadre contrôlé tout en préservant la confidentialité des données réelles.

Cependant, malgré sa richesse et son adoption répandue, le jeu de données CERT Insider Threat présente des limites importantes qui influencent directement son utilisation. La diversité limitée des scénarios, autant malhabiles que malveillants, peut contraindre la capacité des modèles à généraliser à des situations réelles variées, et la représentativité imparfaite des comportements humains empêche de reproduire

fidèlement certaines actions complexes ou contextuelles (Legg et al., 2017 ; Tuor et al., 2017). La granularité temporelle parfois insuffisante et la rareté des indicateurs précoces compliquent également la détection proactive des incidents, qui est pourtant l'objectif central des approches modernes de cybersécurité (Greitzer et al., 2010 ; Cappelli et al., 2012).

Pour pallier ces limites, les chercheurs adaptent l'utilisation de ce jeu de données de plusieurs manières : en complétant les données avec d'autres sources simulées ou réelles anonymisées, en ajustant les algorithmes pour tenir compte des signaux faibles mal représentés, ou en développant des scénarios supplémentaires pour tester la robustesse des modèles face à des comportements non couverts par le jeu de données (Glasser & Lindauer, 2013 ; Legg et al., 2017). Néanmoins, ces solutions sont partielles et introduisent leurs propres défis : le risque de sur-apprentissage sur des scénarios artificiels, la difficulté à valider les modèles dans des contextes réellement opérationnels, et l'incertitude quant à la performance en production demeurent élevés (Tuor et al., 2017 ; Greitzer et al., 2010).

Ces constats soulignent l'importance de mieux structurer et enrichir le jeu de données CERT afin d'améliorer la qualité des environnements de test et, surtout, de soutenir plus efficacement la détection proactive des menaces internes malveillantes. Celles-ci s'avèrent en effet nettement plus dangereuses et coûteuses que les comportements malhabiles, comme l'indiquent les rapports récents du CERT et de la CISA : les incidents malveillants sont associés à des impacts financiers, opérationnels et stratégiques significativement plus élevés, en raison de leur caractère intentionnel, dissimulé et souvent prolongé (CERT, 2023 ; CISA, 2023). Une amélioration du jeu de données est donc essentielle pour permettre aux modèles de mieux capter ces signaux faibles et dynamiques comportementales complexes.

2. Objectifs et question de recherche

Ce mémoire vise à étudier l'adéquation et le potentiel d'amélioration du jeu de données CERT pour la détection proactive des menaces internes malveillantes en entreprise. Dans le cadre de ce travail de recherche, l'analyse se concentrera exclusivement sur les employés malveillants. Étant donné la complexité d'une telle étude, les autres formes d'acteurs internes (fournisseurs, partenaires, sous-traitants) ainsi que les menaces non intentionnelles (erreurs, négligences) ne seront pas prises en compte. En s'appuyant à la fois sur l'analyse des données, des méthodes de détection avancées, et sur les retours d'expérience des professionnels de la cybersécurité, cette recherche souhaite contribuer à optimiser l'utilisation de ce jeu de données dans un contexte opérationnel.

La question centrale de cette recherche est la suivante :

Comment le jeu de données CERT peut-il être amélioré afin de mieux soutenir la détection proactive des menaces internes malveillantes ?

Il s'agira d'évaluer le jeu de données CERT pour la détection proactive des menaces internes malveillantes, identifier ses limites au regard des comportements observés en entreprise, et proposer des pistes d'amélioration concrètes pour enrichir sa représentativité et sa pertinence opérationnelle.

Sous-objectifs

S.-O.1 – Tester et évaluer différentes méthodes de détection (ex. FDA) sur le jeu CERT afin d'identifier les approches les plus performantes pour la détection proactive.

Cet objectif permet de situer le potentiel du jeu de données CERT face à des approches variées de détection et de déterminer dans quelle mesure il peut être utilisé avec des modèles avancés. Il offre ainsi une première évaluation technique qui servira de base à la comparaison avec les retours des praticiens.

S.-O.2 – Recueillir auprès de professionnels de la cybersécurité des informations sur les types de comportements malveillants rencontrés en entreprise afin de comparer avec ceux modélisés dans le jeu CERT.

L'intégration de la perspective des experts permettra d'identifier les écarts entre les scénarios simulés dans le jeu de données et les situations réellement rencontrées en milieu organisationnel. Cela contribuera à évaluer la représentativité et la pertinence des données CERT.

S.-O.3 – Analyser la représentativité, les forces et les limites du jeu de données CERT en lien avec les données terrain et la littérature scientifique.

Cet objectif vise à croiser les résultats de l'expérimentation (S.O.1) et des entrevues (S.O.2) avec les connaissances issues des travaux académiques. Il permettra d'identifier les dimensions bien couvertes par CERT, mais aussi les aspects manquants ou sous-représentés.

S.-O.4 – Formuler des recommandations pour améliorer la qualité et la richesse du jeu de données CERT, notamment en intégrant de nouvelles variables ou scénarios issus du terrain.

À partir des constats dégagés, il s'agira de proposer des améliorations concrètes et réalisables, afin d'accroître l'utilité opérationnelle de CERT dans la détection proactive des menaces internes malveillantes. Ces recommandations pourront servir à orienter tant la recherche que les applications pratiques.

3. Structure du mémoire

Ce mémoire est structuré en six chapitres, chacun contribuant à l'atteinte de l'objectif principal, soit l'évaluation critique du jeu de données CERT et l'identification de pistes d'amélioration pour soutenir la détection proactive des menaces internes malveillantes.

Le premier chapitre présente le contexte de la recherche, la problématique étudiée, la question de recherche ainsi que les objectifs poursuivis. Le deuxième chapitre propose une revue de la littérature portant sur les menaces internes malveillantes, les approches de détection existantes et les jeux de données utilisés en recherche, en mettant en évidence le rôle central du jeu de données CERT.

Le troisième chapitre décrit la méthodologie retenue, fondée sur une approche combinant une expérimentation quantitative sur le jeu CERT r4.2 et des entrevues qualitatives auprès de professionnels de la cybersécurité. Le quatrième chapitre présente les résultats issus de ces deux volets d'analyse, en examinant à la fois les performances des méthodes de détection et les perceptions des experts quant au réalisme des comportements simulés.

Le cinquième chapitre discute ces résultats à la lumière de la littérature et des objectifs de recherche, afin d'évaluer la pertinence méthodologique du jeu CERT et les limites de généralisation des approches testées. Enfin, le sixième chapitre conclut le mémoire en synthétisant les principaux apports de l'étude, en exposant ses limites et en proposant des recommandations et avenues de recherche pour améliorer la représentativité des jeux de données dédiés à la détection des menaces internes malveillantes.

Chapitre 2. Revue de la littérature

L'objectif de ce deuxième chapitre est de passer en revue la littérature scientifique et professionnelle afin de fournir un cadre solide pour cette étude. Plus spécifiquement, il s'agit de synthétiser les connaissances existantes sur les approches de détection des menaces internes malveillantes en cybersécurité, en mettant l'accent sur les méthodes et technologies utilisées pour identifier les menaces internes intentionnelles et à risque provenant d'utilisateurs légitimes.

Pour ce faire, différentes sources de données ont été consultées. Les recherches académiques ont été menées dans des bases de données telles que Web of Science, IEEE Xplore et ACM Digital Library, permettant de rassembler des publications pertinentes issues de divers contextes organisationnels. Des rapports d'organismes spécialisés, notamment CISA et ENISA, ainsi que les publications du CERT de l'Université Carnegie Mellon, ont été mobilisés afin de compléter les données académiques par une perspective pratique sur les menaces internes malveillantes (CERT, 2023 ; ENISA, 2022). Des recherches supplémentaires ont été effectuées sur Google Scholar, en se concentrant sur les articles récents liés à la détection proactive et aux jeux de données simulés pour menaces internes malveillantes (Glasser & Lindauer, 2013 ; Voss, McCulley & Park, 2015 ; Nasir et al., 2021).

Les mots-clés utilisés pour la recherche documentaire incluaient, entre autres : « insider threat detection », « malicious insider », « proactive cybersecurity », « user behavior analytics », « UEBA », ainsi que des combinaisons de termes tels que « jeu de données CERT » et « simulated insider threats ». La sélection des publications s'est faite d'abord par lecture des titres, puis par analyse des résumés afin de confirmer la pertinence des articles. Les références des articles retenus ont ensuite été explorées pour identifier des sources supplémentaires et enrichir la revue de la littérature.

Ce chapitre est structuré en trois sections principales et chaque section a un objectif précis qui permet de répondre à l'objectif général du chapitre qui est de recenser la littérature au sujet de la détection de la menace interne en cybersécurité. La première section approfondit certains concepts clés et la terminologie utilisée tout au long du mémoire. La deuxième section présente les techniques existantes pour la détection des

menaces internes malveillantes intentionnelles, en détaillant leurs principes, leurs limites et les technologies employées. Enfin, la troisième section examine le jeu de données CERT, en mettant en lumière sa structure, ses usages et ses limites, afin de nourrir notre compréhension du contexte et de répondre aux objectifs de recherche.

1. La menace interne en cybersécurité

Les menaces internes constituent un enjeu majeur pour la cybersécurité en raison de leur complexité et de leur diversité. Contrairement aux attaques externes, elles proviennent d'individus disposant d'un accès légitime aux ressources de l'organisation, ce qui les rend particulièrement difficiles à anticiper et à détecter, et les conséquences sont significativement plus sérieuses que pour la menace externe. Comme le soulignent Cappelli, Moore et Trzeciak (2012), « insiders are trusted individuals who use their legitimate access in a way that negatively affects the organization ». Cette confiance accordée aux collaborateurs devient ainsi un facteur aggravant, puisqu'elle permet aux acteurs internes, ici les employés, d'exploiter leurs privilèges sans éveiller de soupçons immédiats.

Plusieurs travaux montrent que l'impact des menaces internes peut dépasser celui des menaces externes. Greitzer et Frincke (2010) insistent sur le fait que les acteurs internes disposent d'une « intimate knowledge of organizational systems and processes », leur donnant la capacité de contourner les mécanismes de contrôle classiques et de maximiser les dommages. De manière complémentaire, Eberle et Holder (2009) soulignent que leur aptitude à « blend malicious actions with normal user behavior » complique fortement la détection automatique des comportements suspects.

En outre, Gamachchi, Sun et Boztas (2017) rappellent que les menaces internes malveillantes ne sont pas seulement difficiles à identifier : elles « pose a persistent and evolving risk » en raison de la variété de leurs formes et de leurs motivations (fraude, vol de propriété intellectuelle, sabotage, négligence). Autrement dit, elles ne se limitent pas à un scénario unique, mais englobent un spectre large et hétérogène de comportements.

1.1. Typologies et profils des acteurs internes

La littérature distingue généralement deux grandes catégories de menaces internes : les menaces intentionnelles et non intentionnelles. Les premières regroupent les comportements malveillants délibérés tels que le vol de données sensibles, l'espionnage industriel, la fraude ou le sabotage, motivés par des objectifs financiers, personnels ou idéologiques (Cappelli, Mohr & Trzeciak, 2012 ; Greitzer et al., 2013). Ces acteurs exploitent leur connaissance des processus organisationnels et leurs droits d'accès pour dissimuler leurs actions, rendant leur détection particulièrement ardue.

Les menaces non intentionnelles, quant à elles, découlent d'erreurs humaines, de négligences ou de pratiques inadaptées. Elles incluent, par exemple, le partage accidentel d'informations sensibles, l'envoi de fichiers à des destinataires non autorisés, une mauvaise configuration de systèmes critiques ou encore l'usage imprudent d'appareils personnels connectés au réseau de l'organisation (Greitzer et al., 2014 ; Rahman, Al Naeem & Abubakar, 2022). Bien qu'elles ne reposent pas sur une volonté malveillante, ces actions peuvent avoir des conséquences tout aussi graves : pertes financières, interruptions opérationnelles ou atteintes à la réputation.

Cette distinction entre comportements intentionnels et involontaires est déterminante, car elle conditionne, par exemple, les stratégies de détection et de mitigation. Ainsi, les menaces intentionnelles requièrent souvent des techniques avancées de surveillance comportementale et de corrélation d'événements afin d'identifier des signaux faibles ou des anomalies (Gamachchi, Sun & Boztas, 2017 ; Yuan et al., 2018), tandis que les menaces non intentionnelles peuvent être réduites grâce à des programmes de formation, des procédures normalisées et des contrôles automatisés renforcés (Nasir et al., 2021 ; Greitzer et al., 2014).

Cependant, plusieurs auteurs rappellent que la frontière entre ces deux types de menaces peut s'avérer poreuse. Une erreur répétée ou mal gérée peut évoluer vers un comportement intentionnel si l'individu exploite consciemment la situation (Cappelli et al., 2012 ; Greitzer & Frincke, 2010). Cette complexité met en évidence la nécessité d'adopter une approche multidimensionnelle, intégrant non seulement des outils techniques, mais aussi une compréhension des facteurs humains et organisationnels.

Au-delà de la distinction entre menaces intentionnelles et non intentionnelles, la littérature souligne l'importance de considérer le rôle de l'acteur interne au sein de l'organisation. Les risques varient en effet selon le type d'acteur et l'étendue de ses privilèges d'accès.

- Les employés permanents représentent la majorité des incidents et peuvent avoir un accès étendu aux systèmes critiques.
- Les consultants et sous-traitants, bien qu'externes à l'organisation, disposent souvent d'accès temporaires ou spécialisés qui, mal utilisés, peuvent engendrer des dommages significatifs.
- Les partenaires ou fournisseurs tiers peuvent également introduire des vulnérabilités, notamment lorsqu'ils interagissent avec des systèmes sensibles ou partagent des données stratégiques.

Comprendre ces distinctions permet d'adapter les stratégies de surveillance, de contrôle et de formation selon le profil de l'insider, afin de réduire le risque global de menaces internes malveillantes (Theoharidou et al., 2005 ; Sinclair & Smith, 2008).

1.2. Facteurs humains et organisationnels

Un aspect central des menaces internes malveillantes réside dans la psychologie et le comportement des acteurs internes. La littérature montre que les actions intentionnellement nuisibles ne dépendent pas uniquement des privilèges techniques, mais sont fortement influencées par des dynamiques humaines et organisationnelles (Greitzer & Frincke, 2010 ; Nasir, Afzal, Latif & Iqbal, 2021). En effet, derrière chaque incident se trouvent des motivations, des émotions et des contextes organisationnels qui façonnent la manière dont les individus détournent leur accès légitime aux ressources pour nuire à l'organisation.

Facteurs humains

Les facteurs humains correspondent aux éléments liés aux caractéristiques personnelles, émotionnelles et psychologiques des individus. Ils jouent un rôle déterminant à la fois dans les comportements malveillants et malhabiles :

- Stress professionnel et surcharge de travail : Un stress chronique ou une charge excessive peut nourrir de la frustration et un sentiment de rancœur. Ces conditions peuvent conduire certains employés à envisager des gestes de représailles, comme l'exfiltration ou la destruction de données.
- Insatisfaction et sentiment d'injustice : Lorsqu'un employé estime être traité de manière inéquitable ou insuffisamment reconnu, ce ressentiment peut devenir un puissant moteur de comportements malveillants, tels que le sabotage, la fuite d'informations ou la collaboration avec des acteurs externes hostiles.
- Manque de reconnaissance et isolement social : Le déficit de valorisation professionnelle ou l'exclusion sociale fragilise le lien de loyauté avec l'organisation. L'isolement peut alimenter un sentiment d'exclusion et motiver des actions malveillantes en guise de rébellion ou de revanche.
- Profil psychologique et traits de personnalité : Certains individus présentent des traits comme l'impulsivité, le goût du risque ou l'opportunisme. Ces caractéristiques augmentent la probabilité qu'un employé exploite volontairement des failles de sécurité pour obtenir un gain personnel ou causer un dommage intentionnel à l'organisation.

Facteurs organisationnels

Les facteurs organisationnels renvoient aux pratiques de gestion, à la culture et à la gouvernance internes qui encadrent la sécurité de l'information. Ils créent le contexte dans lequel des comportements malveillants peuvent émerger ou, au contraire, être limités.

- Culture de sécurité : Une culture organisationnelle qui minimise l'importance de la cybersécurité envoie un signal implicite de tolérance aux comportements

déviant. À l'inverse, une culture robuste et bien ancrée réduit les opportunités pour les acteurs malveillants en instaurant des normes claires et partagées.

- **Politiques et procédures :** L'absence de politiques claires ou la perception de règles injustes peut générer un sentiment de frustration ou d'oppression, qui nourrit le risque de comportements malveillants. Des politiques équilibrées et transparentes sont essentielles pour réduire ces tensions.
- **Formation et sensibilisation des employés :** Si les employés ne comprennent pas les enjeux de sécurité ou perçoivent les règles comme inutiles, cela crée un terrain fertile pour l'émergence de comportements hostiles. À l'inverse, une sensibilisation régulière favorise l'adhésion et réduit les justifications que pourraient se donner des individus malveillants.
- **Surveillance et mécanismes de contrôle internes :** Une surveillance inexistante ou trop laxiste laisse le champ libre aux comportements malveillants, car les individus savent qu'ils risquent peu d'être détectés. Cependant, une surveillance excessive peut aussi générer de la méfiance et inciter certains employés à défier le système. Trouver un équilibre est donc essentiel pour réduire les menaces malveillantes sans fragiliser la confiance organisationnelle.

En somme, les facteurs humains et organisationnels jouent un rôle déterminant dans l'émergence des menaces internes malveillantes. Une gestion efficace repose donc sur une compréhension fine des motivations psychologiques des individus et du climat organisationnel, combinée à des mesures techniques adaptées.

1.3. Motivations et comportements

Plusieurs études indiquent que les motivations des acteurs internes sont fréquemment multifactorielles, combinant dimensions émotionnelles, financières et idéologiques. Ces motivations interagissent avec les dynamiques organisationnelles et la culture d'entreprise pour façonner des comportements à risque (Cappelli, Mohr & Trzeciak, 2012 ; Greitzer et al., 2013). Par exemple, un employé frustré par une gestion perçue comme injuste peut d'abord commettre des erreurs non intentionnelles, mais celles-ci peuvent évoluer en actes délibérés si le contexte organisationnel s'y prête.

Notre mémoire se concentre sur la menace interne malveillante en raison de son impact disproportionné sur la sécurité organisationnelle et de la complexité particulière qu'elle représente pour les systèmes de défense. Contrairement aux erreurs non intentionnelles, qui peuvent être atténuées par des formations ou des contrôles procéduraux, les actes malveillants reposent sur une volonté explicite de nuire et s'appuient sur une connaissance intime des systèmes. Comme le rappellent Greitzer et Frincke (2010), ce type d'initié constitue une menace singulière car il « combine privileged access with intent », créant ainsi un risque difficilement détectable et souvent durable. De plus, le rapport du Ponemon Institute (2020) souligne que les incidents causés par des acteurs internes malveillants figurent parmi les plus coûteux pour les organisations, tant en termes de pertes financières que de durée de résolution. Ces constats justifient la nécessité d'un examen approfondi de la menace interne malveillante, qui représente aujourd'hui un défi stratégique central pour les approches modernes de cybersécurité.

En approfondissant cette dimension, la littérature distingue plusieurs types de comportements malveillants, caractérisés par leur finalité, leur méthodologie et leur impact sur l'organisation (Cappelli, Moore & Trzeciak, 2012 ; Kowalski, Cappelli & Moore, 2008). Ces comportements exploitent les accès privilégiés et la connaissance intime des systèmes, et se manifestent de manière souvent subtile et progressive.

Type de comportements malveillants

Vol de propriété intellectuelle et données sensibles

Le vol de propriété intellectuelle constitue l'un des comportements les plus fréquents et les plus problématiques. Il englobe l'exfiltration de secrets commerciaux, de brevets, de plans stratégiques ou de bases de données clients, souvent à des fins financières ou concurrentielles. Les acteurs internes exploitent leur position pour accéder à des informations normalement protégées et les transférer à des tiers ou pour un usage personnel. Les motivations financières sont prédominantes, mais des raisons

idéologiques ou personnelles peuvent également intervenir (Greitzer et al., 2013 ; Shaw & Fischer, 2005).

Les méthodes sont souvent dissimulées au sein des activités normales de l'utilisateur, ce qui rend la détection difficile. Comme le soulignent Eberle et Holder (2009), les acteurs internes savent « masquer leurs actions en les intégrant aux comportements normaux ». L'impact sur l'organisation peut être considérable, affectant la compétitivité, générant des pertes financières et portant atteinte à la réputation (Ponemon Institute, 2020).

Espionnage industriel et sabotage idéologique

L'espionnage industriel et le sabotage idéologique concernent des actions destinées à affaiblir l'organisation pour un concurrent ou pour des raisons politiques ou idéologiques. Contrairement au vol de propriété intellectuelle à visée économique, ces actes sont souvent motivés par un objectif stratégique ou par une conviction personnelle. Les individus peuvent agir par loyauté envers un concurrent, par adhésion à une cause idéologique, ou par ressentiment envers l'organisation (Greitzer et al., 2013 ; Cappelli et al., 2012).

Les acteurs internes peuvent installer des logiciels malveillants, corrompre des données ou divulguer des informations stratégiques à des parties externes. La planification et la progression des actions visent à réduire les risques de détection, en combinant des accès systèmes légitimes et des actions dissimulées (Gamachchi, Sun & Boztas, 2017). Ces comportements peuvent entraîner des pertes stratégiques importantes et compromettre des infrastructures critiques ou la prise de décision organisationnelle.

Fraude financière et manipulation des systèmes financiers

La fraude interne implique la falsification de documents, la manipulation de transactions ou l'usage abusif de ressources financières pour un gain personnel. Les motivations sont principalement économiques, mais peuvent s'accompagner d'une

rationalisation morale ou d'un sentiment de rancœur envers l'organisation (Cappelli et al., 2012).

Les acteurs internes utilisent leur connaissance des processus financiers pour contourner les contrôles internes, par exemple en modifiant des fichiers comptables, en créant des transactions fictives ou en réalisant des virements non autorisés. L'impact se manifeste par des pertes financières directes, des sanctions légales potentielles et une érosion de la confiance au sein et à l'extérieur de l'entreprise (Ponemon Institute, 2020).

Sabotage des systèmes et attaques « low-and-slow »

Le sabotage se traduit par des actions visant à perturber les opérations normales, à détruire des systèmes ou à compromettre la disponibilité de ressources critiques. Ces comportements sont souvent associés à des motivations émotionnelles ou idéologiques, comme la frustration, le ressentiment ou des convictions personnelles (Shaw & Fischer, 2005 ; Greitzer et al., 2013).

L'acteur interne peut supprimer ou altérer des fichiers critiques, introduire des logiciels malveillants ou perturber les systèmes de production. Dans les attaques « low-and-slow », l'attaquant étale ses actions sur plusieurs semaines ou mois pour maximiser l'impact tout en réduisant les risques de détection. Ces actions sont intégrées aux comportements normaux, ce qui rend inefficaces les outils de détection classiques (Eberle & Holder, 2009). Les conséquences sont durables, car l'acteur interne peut laisser des portes dérobées, copier des données sensibles ou compromettre des relations de confiance, créant ce que Gamachchi, Sun et Boztas (2017) décrivent comme une « evolving risk ». Selon le Ponemon Institute (2020), ces incidents figurent parmi les plus coûteux et durables pour les organisations.

1.4 Conclusion

La menace interne malveillante ne peut être comprise uniquement sous l'angle technique. Elle résulte d'une interaction complexe entre les motivations individuelles, les dynamiques humaines et psychosociales, ainsi que les vulnérabilités

organisationnelles. Les tensions liées au stress professionnel, au sentiment d'injustice ou au manque de reconnaissance se combinent souvent avec des faiblesses structurelles, comme des politiques de sécurité inadaptées, une culture organisationnelle déficiente ou un contrôle insuffisant des accès. Ces conditions créent un terrain propice à l'émergence de comportements à risque, qu'il s'agisse de vol de données, de fraude ou de sabotage.

Dans ce contexte, l'intégration de signaux organisationnels, tels que les changements dans les habitudes de travail, les tensions hiérarchiques ou les anomalies d'accès aux systèmes, apparaît comme un levier essentiel pour compléter les approches techniques de détection. Toutefois, l'exploitation de ces données sensibles soulève d'importants enjeux éthiques et de confidentialité, qui limitent leur usage direct dans la pratique courante.

Ainsi, la compréhension de la menace interne requiert une approche multidimensionnelle qui dépasse la simple analyse comportementale pour inclure les facteurs humains et organisationnels. Cette perspective intégrée permet non seulement de mieux anticiper les risques, mais aussi de concevoir des stratégies de prévention et de détection plus adaptées à la réalité des environnements organisationnels. Elle ouvre également la voie à de futures recherches visant à réduire l'écart entre données simulées et dynamiques réelles, afin de renforcer la robustesse et la pertinence des modèles de détection.

2. Approches et méthodes de détection de la menace interne malveillante

La détection de la menace interne malveillante repose sur une diversité de méthodes allant des règles simples aux approches hybrides intégrant l'intelligence artificielle et l'analyse organisationnelle. Toutes partagent un même défi : distinguer les comportements légitimes des employés des activités réellement malveillantes, souvent dissimulées dans la normalité.

2.1. Approches basées sur les règles et signatures

Les approches traditionnelles de détection reposent sur des règles prédéfinies, des politiques de sécurité et des signatures d'événements connus. Par exemple, une règle peut être configurée pour générer une alerte lorsqu'un employé accède à un fichier sensible en dehors des heures ouvrables, copie un volume de données supérieur à un seuil établi ou connecte un périphérique externe non autorisé. De même, les signatures permettent de comparer un comportement observé à une base de modèles d'incidents déjà répertoriés, comme une tentative d'exfiltration de données utilisant un outil connu (Greitzer et al., 2012). Ces systèmes sont relativement simples à mettre en place et permettent de détecter efficacement des comportements déjà identifiés comme suspects.

Cependant, leur principal inconvénient est leur incapacité à détecter des attaques nouvelles ou non répertoriées. En effet, ces approches produisent également un grand nombre de faux positifs, car les règles et signatures reposent sur des seuils ou des modèles fixes qui ne tiennent pas toujours compte du contexte. Par exemple, un transfert de fichiers important peut être une activité légitime liée au travail quotidien d'un employé, comme la préparation d'un rapport volumineux ou la mise à jour d'une base de données, mais sera tout de même signalé comme suspect. Dans certaines organisations, ce type d'alerte peut se produire des dizaines de fois par jour, ce qui contribue à saturer les équipes de sécurité et à diluer leur attention sur les signaux réellement critiques. Ils sont donc vulnérables face aux menaces émergentes et aux comportements subtilement malveillants (Eberle et al., 2010).

Ce volume élevé de faux positifs constitue un inconvénient majeur : il réduit la crédibilité des systèmes de détection, surcharge les équipes de sécurité qui doivent analyser manuellement un grand nombre d'alertes inutiles, et peut entraîner une lassitude face aux alarmes. Ce phénomène, appelé parfois « alert fatigue », augmente le risque que de véritables incidents passent inaperçus ou soient traités avec retard, compromettant ainsi l'efficacité globale de la défense de l'organisation.

2.2. Analyse comportementale et modèles statistiques

Pour dépasser les limites des règles figées, de nombreux chercheurs et praticiens se sont intéressés à l'analyse comportementale, qui constitue aujourd'hui l'un des piliers des approches modernes de détection des menaces internes malveillantes. L'idée centrale est de modéliser le comportement habituel d'un utilisateur, ses horaires de connexion, le volume et la nature des courriels envoyés, les fichiers consultés, ou encore les périphériques utilisés, afin d'identifier les écarts significatifs pouvant indiquer une activité suspecte (Magklaras & Furnell, 2001). Contrairement aux approches basées sur les signatures, qui ne détectent que ce qui est déjà connu, l'analyse comportementale vise à repérer l'inédit : un changement soudain ou progressif dans la manière dont un individu interagit avec le système.

Concrètement, ces approches reposent sur des modèles statistiques et probabilistes permettant de définir une « baseline » comportementale. Par exemple, l'utilisation de chaînes de Markov peut servir à analyser les séquences d'actions réalisées par un utilisateur (ex. ouverture de fichiers, envoi de courriels, connexions réseau) et à estimer la probabilité d'occurrence d'une action donnée en fonction du comportement passé. Une action improbable dans cette séquence, comme l'accès soudain à une base de données sensible en dehors des heures normales, sera identifiée comme une anomalie (Eberle & Holder, 2009). De même, les modèles gaussiens mixtes permettent de regrouper les comportements usuels en « clusters » et de signaler tout comportement qui s'en éloigne fortement (Legg et al., 2015).

Ces méthodes présentent un intérêt particulier pour la détection des menaces internes malveillantes dites « low and slow ». Dans ces scénarios, l'utilisateur malveillant agit progressivement et de manière discrète, par exemple en exfiltrant de petites quantités de données sur une longue période pour éviter de déclencher des alertes. Là où des règles statiques resteraient silencieuses, une approche comportementale peut détecter la tendance anormale dans la durée et attirer l'attention des analystes.

En raison de leur capacité à traiter des scénarios inédits, ces approches connaissent une popularité croissante, tant dans les travaux académiques que dans les solutions commerciales de type UEBA (User and Entity Behavior Analytics). Gartner (2015) a

d'ailleurs contribué à la diffusion de ce paradigme en soulignant que la modélisation comportementale représentait une avancée majeure pour la détection des menaces internes malveillantes, en permettant de contextualiser les alertes et de réduire les faux positifs. Plusieurs éditeurs de solutions de cybersécurité intègrent désormais des algorithmes statistiques dans leurs plateformes SIEM, offrant ainsi aux organisations une capacité d'analyse plus dynamique et adaptative.

Toutefois, ces approches ne sont pas exemptes de limites. Premièrement, elles exigent une grande quantité de données historiques pour construire des modèles fiables. Or, dans certaines organisations, la collecte et la conservation de données à grande échelle peuvent poser des problèmes techniques ou juridiques, notamment en matière de protection de la vie privée. Deuxièmement, elles peuvent être biaisées par des comportements atypiques mais légitimes : un employé en période de projet intensif peut travailler tard le soir ou manipuler un volume inhabituel de données, déclenchant à tort des alertes. Enfin, leur efficacité dépend fortement de la qualité des données disponibles : des enregistrements incomplets ou bruités peuvent fausser la baseline et réduire la pertinence des résultats (Glasser & Lindauer, 2013).

Ces limites expliquent pourquoi l'analyse comportementale tend à être combinée à d'autres approches, comme l'apprentissage automatique ou l'analyse contextuelle organisationnelle, afin d'améliorer la précision et la robustesse des détections. Ainsi, les modèles statistiques constituent aujourd'hui une étape incontournable dans l'évolution des systèmes de détection, mais ils trouvent toute leur puissance lorsqu'ils sont intégrés à des cadres hybrides qui conjuguent intelligence artificielle, analyse des signaux organisationnels et règles expertes.

2.3. Approches basées sur l'apprentissage automatique et intelligence artificielle

L'apprentissage automatique (machine learning) s'est imposé comme un pilier central de la détection des menaces internes malveillantes, en raison de sa capacité à s'adapter dynamiquement aux nouveaux comportements. Ces approches se déclinent

principalement en trois catégories : supervisées, non supervisées et séquentielles (deep learning).

2.3.1 Approches supervisées

Les approches supervisées reposent sur des données annotées, c'est-à-dire des exemples connus de comportements normaux et de comportements malveillants. L'algorithme apprend à distinguer ces deux types d'actions afin de prédire la probabilité qu'un nouvel événement soit suspect. Parmi les techniques les plus utilisées figurent :

- XGBoost : une méthode de gradient boosting appliquée aux arbres de décision (Chen & Guestrin, 2016). XGBoost se distingue par sa scalabilité, sa capacité à traiter de grands ensembles de données hétérogènes et son aptitude à gérer les données déséquilibrées via un poids différent pour chaque classe.
- Support Vector Machines (SVM) : ces modèles cherchent à maximiser la marge entre les classes dans un espace de haute dimension (Cortes & Vapnik, 1995). Les SVM sont efficaces même avec un nombre limité d'échantillons et permettent de traiter des données non linéairement séparables grâce à l'utilisation de fonctions noyaux.
- Forêts aléatoires (Random Forests) : en combinant plusieurs arbres de décision, elles réduisent le surapprentissage et améliorent la généralisation sur de nouveaux ensembles de données.
- Réseaux de neurones classiques : capables de modéliser des relations complexes entre variables explicatives, ils sont particulièrement utiles pour détecter des schémas subtils de comportements malveillants.

Ces approches offrent plusieurs avantages : elles donnent des résultats précis lorsque les données sont abondantes et bien étiquetées, permettent de traiter de grands volumes de logs et peuvent intégrer une variété de variables explicatives (fréquence des connexions, volumes de données transférées, types d'accès aux fichiers, etc.).

Cependant, elles présentent aussi des limites importantes :

- Dépendance aux données étiquetées : sans exemples fiables d'attaques internes, la performance chute fortement (Liu et al., 2018).
- Détection des comportements émergents : elles peuvent peiner à identifier des attaques inédites ou des tactiques nouvelles qui n'ont pas été rencontrées lors de l'entraînement.
- Explicabilité : certains modèles, notamment les réseaux de neurones et XGBoost complexes, fonctionnent comme des « boîtes noires », rendant difficile la justification d'une alerte auprès des équipes de sécurité ou de la direction.

Malgré ces limites, les approches supervisées restent très pertinentes pour la détection de comportements connus et répétitifs, et elles servent souvent de base à des systèmes hybrides combinant intelligence humaine, règles métiers et autres techniques de machine learning.

Approches non supervisées

Lorsque les incidents étiquetés sont rares, ce qui est fréquent dans le cas de la menace interne malveillante, les approches non supervisées sont privilégiées. Contrairement aux méthodes supervisées, qui nécessitent des exemples de menaces connues pour l'entraînement, l'apprentissage non supervisé repose sur l'analyse des comportements habituels des utilisateurs afin de détecter des écarts significatifs. Cette caractéristique en fait un outil pertinent dans des environnements où les données de menaces sont peu nombreuses ou difficiles à obtenir.

Deux grandes catégories de méthodes dominent cette approche : le clustering et la détection d'anomalies. Le clustering regroupe les utilisateurs en fonction de comportements similaires. Des algorithmes tels que K-Means ou DBSCAN permettent ainsi de former des groupes représentant des comportements "normaux". Lorsqu'une activité ne correspond à aucun cluster identifié, elle peut être considérée comme suspecte et potentiellement indicative d'une menace interne malveillante.

La détection d'anomalies repose, quant à elle, sur la modélisation de la distribution des comportements considérés comme normaux. Des techniques comme Isolation Forest,

Local Outlier Factor (LOF) ou encore les Autoencoders détectent toute activité s'écartant fortement de cette distribution. Par exemple, une connexion à un horaire inhabituel, un accès répété à des dossiers sensibles ou un transfert massif de fichiers peuvent être signalés comme anomalies et faire l'objet d'une investigation.

L'un des avantages majeurs des approches non supervisées est leur capacité à détecter des attaques inédites, que les méthodes supervisées ne sauraient identifier faute d'exemples préalables. Toutefois, elles présentent une limite importante : leur manque d'interprétabilité. En effet, il peut être difficile pour les analystes de comprendre et de justifier pourquoi un comportement a été signalé comme suspect, ce qui freine parfois leur adoption dans des environnements sensibles comme les entreprises ou les institutions publiques.

2.3.2 Approches séquentielles et deep learning

Les approches séquentielles reposent sur le deep learning pour analyser les journaux d'activité comme des séries temporelles, ce qui permet de détecter des anomalies subtiles dans le comportement des utilisateurs sur le temps. Ces méthodes sont particulièrement adaptées à la menace interne malveillante, car elles peuvent capturer des déviations progressives ou des schémas complexes qui échappent aux techniques classiques.

1. Réseaux de Neurones Récurrents (RNN) et Long Short-Term Memory (LSTM)

Les RNN sont conçus pour traiter des données séquentielles en mémorisant les observations précédentes afin de prédire les comportements futurs. Les LSTM, une variante des RNN, sont capables de gérer la dépendance à long terme, ce qui les rend efficaces pour détecter des écarts progressifs ou des séquences inhabituelles sur de longues périodes (Yuan et al., 2018).

- Exemple concret : un employé qui commence à accéder progressivement à des fichiers sensibles sur plusieurs semaines peut être identifié grâce à un LSTM analysant l'évolution des accès.

- Avantages : capacité à modéliser des comportements complexes et évolutifs, réduction des faux positifs par apprentissage des modèles comportementaux spécifiques à l'organisation.
- Limitations : nécessite de grandes quantités de données et des ressources de calcul importantes. L'explicabilité des prédictions reste faible, ce qui complique la justification des alertes auprès de la direction.

2. Auto-encodeurs (AE)

Les auto-encodeurs sont des réseaux de neurones qui apprennent à reconstruire leurs entrées. Lorsqu'ils sont entraînés uniquement sur des comportements normaux, ils signalent les anomalies en détectant des écarts significatifs entre l'entrée et sa reconstruction (Pantelidis et al., 2021).

- Exemple concret : un transfert massif de fichiers vers un périphérique externe ou un accès hors horaires habituels peut être détecté comme anomalie.
- Avantages : efficace pour repérer des comportements rares ou nouveaux sans étiquetage préalable, robuste face au bruit dans les données.
- Limitations : complexité du modèle et besoin de données massives pour un entraînement performant.

3. Combinaison LSTM-CNN (Convolutional Neural Networks)

Certains travaux combinent LSTM et CNN pour bénéficier à la fois de la capacité de détection temporelle des LSTM et de la puissance d'extraction de caractéristiques des CNN (Yuan et al., 2018). Cette combinaison est utilisée pour analyser les séquences d'activité des utilisateurs et identifier des motifs d'attaque subtils qui seraient invisibles à des modèles isolés.

- Exemple concret : détection en temps réel de comportements de phishing interne ou d'exfiltration de données répartie sur plusieurs sessions.
- Avantages : haute précision et faible taux de faux positifs, détection en quasi-temps réel.
- Limitations : forte consommation en ressources de calcul et en mémoire, nécessitant souvent des infrastructures spécialisées.

2.3.3 Applications et défis pratiques

Les approches séquentielles et fondées sur le deep learning trouvent plusieurs applications concrètes dans le domaine de la cybersécurité, notamment pour la détection des acteurs internes malveillants. Elles permettent, par exemple, d'identifier des comportements d'exfiltration progressive de données qui échappent aux règles fixes, comme lorsqu'un employé transfère régulièrement de petits volumes de fichiers sensibles pour ne pas déclencher les seuils traditionnels. Elles sont également utilisées dans la surveillance des comptes à privilèges, où une séquence d'actions inhabituelles, connexion depuis un terminal non habituel, escalade de privilèges, consultation de fichiers confidentiels, peut indiquer une compromission ou un abus de droits. Enfin, elles trouvent leur application dans la prévention de fraudes internes, par exemple pour repérer des manipulations subtiles de données financières ou de dossiers clients.

Malgré ces applications prometteuses, la mise en œuvre de ces approches rencontre plusieurs obstacles concrets :

- **Besoin de données volumineuses et de qualité** : les modèles séquentiels exigent des journaux riches et continus pour bien apprendre les dynamiques comportementales. Or, dans la pratique, les données sont souvent incomplètes, bruitées ou cloisonnées entre différents systèmes (RH, sécurité, réseau), ce qui limite la performance des modèles et rend difficile la généralisation à des environnements réels, comme ceux simulés dans le jeu de données CERT.
- **Complexité technique élevée** : le déploiement de modèles deep learning nécessite non seulement une expertise avancée en intelligence artificielle, mais aussi une forte compréhension des contextes métier et sécurité. Sans cette double compétence, il est difficile d'obtenir des modèles exploitables opérationnellement, même s'ils sont performants en laboratoire.
- **Problème d'explicabilité** : ces modèles fonctionnent comme des "boîtes noires", produisant des alertes sans explication claire. Dans des environnements sensibles, par exemple des institutions financières ou gouvernementales, cette opacité limite l'adoption, car les équipes doivent justifier les décisions, parfois devant des instances légales ou réglementaires. L'absence de transparence peut

également générer de la confusion ou une méfiance vis-à-vis des alertes produites.

- Le volume élevé d’alertes, combiné à la complexité des modèles et aux limites des données, constitue un défi majeur. Les équipes de sécurité peuvent être submergées par des alertes faussement positives, phénomène appelé « alert fatigue », ce qui augmente le risque que de véritables incidents passent inaperçus ou soient traités avec retard. Ces enjeux soulignent que la performance des modèles ne dépend pas uniquement de l’algorithme, mais aussi de la qualité et de la représentativité des données, ainsi que de l’accompagnement organisationnel.

En résumé, les approches séquentielles et deep learning offrent un potentiel considérable pour la détection proactive des menaces internes malveillantes, notamment en permettant de repérer des comportements subtils ou atypiques qui échappent aux méthodes traditionnelles basées sur les règles et signatures. Toutefois, leur efficacité réelle est étroitement liée à la qualité des données, aux compétences des équipes et à l’intégration dans un cadre organisationnel approprié. Ces limites suggèrent qu’une approche hybride, combinant méthodes traditionnelles et approches avancées, reste actuellement la stratégie la plus pragmatique.

Enfin, la mise en œuvre de ces technologies doit s’accompagner de politiques claires, de formations adaptées et de mécanismes de contrôle appropriés, afin de maximiser la valeur des systèmes de détection tout en minimisant les risques liés aux faux positifs ou à l’incompréhension des alertes. Cette réflexion est directement pertinente pour le travail de ce mémoire, qui cherche à enrichir et adapter le jeu de données CERT afin d’optimiser la détection proactive des comportements malveillants des employés.

Bien que les approches supervisées, non supervisées et séquentielles offrent des outils puissants pour détecter les comportements malveillants, chacune présente des limites spécifiques. Les méthodes supervisées dépendent fortement de données étiquetées et peuvent manquer des comportements émergents, les approches non supervisées sont parfois difficiles à interpréter, et les modèles séquentiels exigent des volumes de données importants et restent souvent des “boîtes noires”. Pour pallier ces contraintes

et maximiser l'efficacité de la détection, les chercheurs privilégient aujourd'hui des stratégies hybrides et intégrées, combinant plusieurs techniques et sources de données.

2.3.4 Approches hybrides et intégrées

Les approches hybrides et intégrées visent à combiner de manière complémentaire différentes méthodes, règles métier, apprentissage supervisé et non supervisé, deep learning, ainsi que signaux organisationnels et comportementaux, afin d'améliorer la précision et la couverture des systèmes de détection des menaces internes malveillantes (Nurse et al., 2020). En intégrant plusieurs perspectives, elles permettent de repérer des comportements suspects qui pourraient échapper aux méthodes isolées et d'optimiser la détection précoce des incidents.

L'une des dimensions clés de ces approches repose sur l'analyse relationnelle à l'aide de graphes, qui relient utilisateurs et ressources pour enrichir la compréhension des interactions et identifier des anomalies subtiles d'accès ou de transfert de données. Gamachchi, Sun et Boztas (2017) soulignent que ces modèles permettent de détecter des anomalies difficiles à repérer avec des méthodes classiques. L'application de réseaux de neurones convolutifs sur graphes (Graph Convolutional Networks – GCN) améliore encore la précision de la détection en exploitant ces relations complexes (Yuan et al., 2023). Cette contextualisation est essentielle pour aider les équipes de sécurité à interpréter les incidents, prioriser les investigations et justifier les décisions auprès de la direction ou des instances réglementaires.

Parallèlement, la combinaison de signaux techniques et humains constitue un autre avantage majeur. Le modèle proposé par MITRE (2019) illustre comment l'analyse conjointe de journaux systèmes, d'authentifications, de transferts de données et de facteurs humains comme le stress ou l'insatisfaction professionnelle permet de détecter des schémas plus complexes et d'anticiper des menaces qui se manifestent progressivement ou de manière subtile. Cette approche multidimensionnelle étend la portée de la détection et contribue à identifier des comportements atypiques qui resteraient invisibles avec des méthodes purement techniques.

Des initiatives à grande échelle, comme le programme ADAMS (Anomaly Detection at Multiple Scales) développé par la DARPA (Stolfo et al., 2012), montrent la pertinence de combiner différentes méthodes et niveaux d'analyse, technique, comportementale et organisationnelle, pour identifier des signaux faibles et détecter précocement des comportements malveillants. Ces applications illustrent que l'intégration de plusieurs perspectives améliore non seulement la précision des systèmes de détection, mais réduit également le nombre de faux positifs, limitant ainsi la fatigue des analystes et le risque que de véritables incidents passent inaperçus.

Cependant, la mise en œuvre de ces approches présente plusieurs défis. Sur le plan technique, elle exige une coordination interdisciplinaire entre spécialistes de la cybersécurité, analystes de données et experts en comportement organisationnel, ainsi que des ressources suffisantes pour traiter des données variées, parfois cloisonnées ou hétérogènes (Nurse et al., 2020). La complexité des modèles, notamment en deep learning et graphes, nécessite une expertise avancée pour garantir des résultats fiables et exploitables opérationnellement. Sur le plan éthique, l'exploitation de signaux humains et la surveillance accrue soulèvent des questions relatives à la vie privée, à la confiance des employés et à la conformité aux réglementations. Enfin, l'efficacité de ces systèmes dépend fortement de leur acceptabilité organisationnelle : la collaboration des différents services, y compris les ressources humaines et la sécurité informatique, est essentielle pour que les informations générées soient interprétées et utilisées correctement (Nurse et al., 2020).

En synthèse, les approches hybrides et intégrées offrent un potentiel considérable pour améliorer la détection des menaces internes malveillantes, en combinant précision, contextualisation et couverture étendue (Gamachchi, Sun & Boztas, 2017 ; Yuan et al., 2023 ; MITRE, 2019 ; Stolfo et al., 2012 ; Nurse et al., 2020). Leur succès repose sur un équilibre entre performance technique, coordination interdisciplinaire et respect des enjeux éthiques et organisationnels. Lorsqu'elles sont déployées avec soin, ces approches permettent non seulement de repérer des comportements atypiques, mais aussi de renforcer la confiance et l'efficacité des systèmes de sécurité au sein de l'organisation.

2.4 Approches basées sur l'analyse fonctionnelle des données (FDA)

Les approches basées sur l'analyse fonctionnelle des données (*Functional Data Analysis*, FDA) constituent une famille de méthodes statistiques destinées à analyser des données qui évoluent dans le temps sous la forme de courbes ou de trajectoires continues. Contrairement aux techniques traditionnelles qui transforment les séries temporelles en valeurs résumées, telles que des moyennes ou des totaux, la FDA considère la forme complète des comportements, ce qui permet de préserver la dimension dynamique des phénomènes étudiés. Cette perspective est particulièrement utile dans le cadre des menaces internes malveillantes, où les variations progressives, les routines quotidiennes et les dérives comportementales jouent un rôle central dans l'identification des risques. Comme le rappellent Ramsay et Silverman (2005), la FDA permet d'aborder des phénomènes observés sous forme de fonctions continues tout en offrant des outils d'analyse comparables à ceux mobilisés pour les données classiques, notamment la classification, la régression ou l'analyse de variance. Plus récemment, Gertheiss, Goldsmith et Greven (2024) ont démontré que la FDA est particulièrement appropriée pour analyser des comportements évoluant sur des périodes prolongées et présentant des variations d'amplitude ou de structure difficiles à capturer avec des approches discrètes.

2.4.1 Fondements et principes de la FDA

La FDA repose sur l'idée que le comportement d'un individu peut être représenté par une fonction $X(t)$, où t correspond au temps. À partir d'observations discrètes, telles que le nombre de fichiers consultés quotidiennement ou la fréquence des connexions réseau, cette fonction peut être reconstruite au moyen de techniques de lissage qui réduisent le bruit tout en préservant la tendance générale du comportement. Ce processus permet de représenter les activités sous forme de courbes continues, révélant ainsi des tendances sous-jacentes qui seraient invisibles dans une analyse ponctuelle. Cette représentation fonctionnelle est particulièrement utile pour identifier des dérives progressives typiques des menaces internes malveillantes qui se développent de manière subtile et graduelle. Ferraty et Vieu (2006) ainsi que Kokoszka et Reimherr (2017)

soulignent d'ailleurs que cette capacité à modéliser des séquences lissées et continues constitue l'une des forces majeures de la FDA.

La Functional Principal Component Analysis (FPCA) constitue un outil central dans ce cadre. Elle permet de réduire la dimensionnalité des trajectoires en les décomposant en quelques composantes principales qui capturent l'essentiel de la variabilité observée, facilitant ainsi la comparaison entre individus. Dai et Müller (2018) ont montré que la FPCA offre une représentation compacte et fidèle de la dynamique temporelle des comportements, ce qui en fait une méthode particulièrement adaptée à l'analyse de séries complexes. En combinant lissage, représentation continue et réduction dimensionnelle, la FDA offre ainsi un cadre analytique performant pour repérer des anomalies fonctionnelles subtiles, comme l'ont montré Febrero-Bande, Galeano et González-Manteiga (2008) ainsi que Hubert, Rousseeuw et Segaert (2015) dans le contexte de la détection d'outliers.

2.4.2 Représentation des comportements d'utilisateurs comme des trajectoires

Les comportements d'un employé peuvent être représentés sous forme de trajectoires fonctionnelles construites à partir de différentes sources de journaux d'activité : accès à des fichiers sensibles, volume de courriels envoyés, fréquence des connexions en dehors des heures régulières, transferts vers des supports externes ou encore activités Web. Plutôt que d'observer ces signaux séparément, la FDA permet d'intégrer l'ensemble de ces événements dans une courbe comportementale continue qui reflète la dynamique globale d'un utilisateur. Ce type de représentation rend possible l'identification de motifs temporels complexes, tels que des dérives progressives ou des modifications subtiles dans les habitudes de travail.

Par exemple, deux employés peuvent présenter un volume similaire d'accès à des répertoires sensibles. Cependant, si l'un d'entre eux augmente progressivement la fréquence ou l'intensité de ces accès sur plusieurs semaines, cette évolution sera visible dans la trajectoire fonctionnelle, alors qu'elle resterait indétectable dans une analyse ponctuelle. Cette capacité à représenter la structure temporelle des

comportements constitue un avantage majeur de la FDA dans la détection de signaux faibles.

2.4.3 Outils clés : FPCA et alignement temporel

La FPCA joue un rôle central en permettant d'identifier les tendances générales des comportements, les modes principaux de variation et les trajectoires atypiques. Comme le montrent Gertheiss et al. (2024), cet outil facilite la comparaison des comportements entre individus et la détection d'écarts significatifs, qu'il s'agisse de variations anormales, de routines inhabituelles ou de dérives annonçant un risque.

La FDA s'appuie également sur des techniques d'alignement temporel, telles que le *time warping*, qui rendent possible la comparaison de courbes même lorsque les événements ne surviennent pas aux mêmes moments. Cet aspect est essentiel dans un contexte organisationnel où les employés ne suivent pas des patterns d'activité parfaitement synchrones. Le *warping* permet ainsi d'isoler la structure des comportements indépendamment des décalages temporels, ce qui améliore la précision de la détection d'anomalies.

2.4.5 Intérêt de la FDA pour la détection des menaces internes malveillantes

L'un des avantages majeurs de la FDA réside dans sa capacité à analyser les comportements de manière continue. Les attaques de type *low and slow*, caractérisées par une accumulation discrète d'actions suspectes, sont généralement difficiles à détecter à l'aide d'approches reposant sur des seuils fixes ou des agrégations ponctuelles. La FDA permet au contraire de mettre en évidence ces évolutions graduelles en considérant la forme globale des trajectoires. De plus, les techniques de lissage réduisent l'impact du bruit présent dans les journaux d'activité, ce qui améliore la précision de l'analyse même lorsque les données sont irrégulières ou fragmentées.

La réduction dimensionnelle réalisée grâce à la FPCA contribue également à faciliter la classification des comportements atypiques ou potentiellement malveillants en s'appuyant sur une représentation compacte de la trajectoire complète de l'utilisateur. Horváth et Kokoszka (2012) ont ainsi montré que cette approche est particulièrement efficace pour analyser des séries temporelles complexes et identifier des écarts structurels indiquant un comportement suspect.

2.4.6 Applications en cybersécurité

Bien que la FDA ait d'abord été largement mobilisée dans des domaines comme la biomécanique, les neurosciences ou la médecine (Gertheiss et al., 2024), elle suscite un intérêt croissant dans la cybersécurité, notamment dans l'analyse comportementale. Ullah et Finch (2013) ont démontré que les modèles fonctionnels sont particulièrement efficaces pour classifier des séquences temporelles et détecter des anomalies dans des données où la structure évolutive est essentielle. Cette capacité ouvre de nouvelles perspectives pour l'étude des menaces internes malveillantes, car elle permet d'identifier des trajectoires d'activité qui s'écartent des comportements attendus, de regrouper des séquences en fonction de leur similarité temporelle ou encore d'analyser des dérives progressives précédant un acte malveillant.

L'application de la FDA à des jeux de données simulés comme CERT permet par ailleurs d'évaluer la crédibilité et la cohérence temporelle des comportements qu'ils reproduisent. Cette perspective est particulièrement pertinente dans un contexte où les jeux de données synthétiques tendent parfois à représenter les comportements malveillants de manière trop abrupte ou artificielle.

En somme, la FDA constitue une approche à la fois puissante et complémentaire pour l'analyse des menaces internes malveillantes. En intégrant pleinement la dimension temporelle des comportements, elle permet de repérer des évolutions subtiles que les méthodes traditionnelles peinent à capturer. Dans ce mémoire, la FDA joue ainsi un double rôle : d'une part, comme méthode de détection permettant d'étudier les trajectoires comportementales des utilisateurs, et d'autre part, comme outil

d'évaluation critique destiné à analyser dans quelle mesure le jeu CERT parvient à reproduire fidèlement les dynamiques associées aux menaces internes malveillantes.

2.5 Conclusion

La détection des menaces internes malveillantes a connu une évolution marquée, passant de mécanismes traditionnels basés sur des règles et des signatures vers des approches de plus en plus sophistiquées mobilisant l'analyse comportementale, l'apprentissage automatique et le deep learning. Les méthodes classiques demeurent efficaces pour identifier des comportements déjà connus, mais elles peinent à saisir les attaques subtiles, inédites ou progressives, tout en générant un volume important de faux positifs. Les approches comportementales et statistiques représentent un progrès notable en permettant de détecter des anomalies et des dérives temporelles, mais leur efficacité dépend étroitement de la qualité, de la granularité et de la densité des données disponibles. Quant aux techniques de machine learning, supervisées, non supervisées ou séquentielles, elles offrent une capacité accrue à détecter des schémas complexes et émergents, tout en soulevant des défis liés à leur besoin en données massives, leur opacité et la difficulté de les interpréter dans un contexte opérationnel.

C'est dans ce paysage que les approches hybrides et intégrées s'imposent progressivement comme une voie prometteuse. En combinant règles métier, analyses comportementales, signaux organisationnels et intelligence artificielle, elles permettent de contextualiser les alertes, de réduire les faux positifs et de capter des comportements atypiques qui échappent aux méthodes isolées. Leur efficacité repose toutefois sur un équilibre délicat entre performance technique, expertise interdisciplinaire, gouvernance des données et respect des principes éthiques, notamment en matière de vie privée et de confiance organisationnelle.

Dans cette évolution, les approches basées sur l'analyse fonctionnelle des données (FDA) offrent une contribution particulièrement intéressante. En représentant les comportements sous forme de trajectoires continues plutôt que comme des points discrets, elles permettent de saisir la dimension temporelle des actions, d'identifier des

dérives progressives, des ruptures subtiles ou des fluctuations inhabituelles, et de comparer de manière plus fine les comportements entre utilisateurs. La FDA constitue ainsi un complément puissant aux méthodes existantes : elle éclaire non seulement la dynamique des comportements, mais fournit également un outil rigoureux pour évaluer la crédibilité et la cohérence temporelle des jeux de données utilisés dans la recherche, comme le CERT.

Cette complémentarité met en évidence un enjeu crucial : la dépendance structurelle des approches modernes à la disponibilité de jeux de données riches, réalistes et représentatifs. Or, l'accès à des données réelles de menaces internes malveillantes demeure difficile en raison des contraintes légales, éthiques et organisationnelles, de la sensibilité des informations internes, et de la rareté des incidents avérés. Les jeux de données simulés, tels que CERT, deviennent alors des outils essentiels pour entraîner, tester et comparer les méthodes de détection, mais ils doivent eux-mêmes être évalués de manière critique afin d'en comprendre les limites.

La section suivante s'attache précisément à cet enjeu. Elle analyse les différents types de jeux de données disponibles, leurs avantages et leurs limites, et présente en quoi leur qualité conditionne directement la capacité des organisations et des chercheurs à développer des systèmes de détection robustes, réalistes et véritablement prédictifs.

3. Les jeux de données

La détection de la menace interne malveillante repose fortement sur la disponibilité de jeux de données de qualité, permettant d'entraîner, de valider et de comparer différents modèles de détection. Contrairement à d'autres domaines de la cybersécurité, tels que la détection de malwares ou d'intrusions réseau, l'accès à des données pertinentes sur les comportements internes malveillants reste particulièrement limité (Glasser & Lindauer, 2013). Plusieurs facteurs structurels expliquent cette rareté.

Un premier frein est lié à la sensibilité des informations organisationnelles. Les journaux d'activité des employés (courriels, accès aux fichiers, navigation web, connexions système) contiennent des données critiques pouvant révéler la structure hiérarchique

d'une organisation, ses processus internes ou même ses actifs stratégiques. Partager ces données exposerait non seulement des secrets commerciaux mais aussi des vulnérabilités potentielles exploitables par des attaquants (Legg et al., 2015).

Un second facteur est lié aux contraintes légales et éthiques de confidentialité. Dans la plupart des juridictions, les activités des employés sont considérées comme des données personnelles sensibles. Le Règlement général sur la protection des données (RGPD) en Europe ou encore les lois nord-américaines sur la vie privée imposent des restrictions strictes sur la collecte, l'utilisation et le partage de telles données. Même lorsque des logs peuvent être anonymisés, ce processus tend à réduire leur richesse informationnelle, en supprimant ou en transformant des attributs essentiels pour la détection des comportements suspects (CMU-SEI, 2021).

Enfin, un troisième obstacle réside dans la rareté des incidents malveillants étiquetés en pratique. Comparativement aux cyberattaques externes, les cas avérés de menaces internes malveillantes sont beaucoup moins fréquents, et lorsqu'ils se produisent, leur documentation reste partielle et difficile à exploiter scientifiquement. La frontière entre erreur humaine et intention malveillante est souvent floue, ce qui complique la constitution de bases de données représentatives et fiables pour l'entraînement de modèles supervisés (Nurse et al., 2020).

Par conséquent, les chercheurs et praticiens s'appuient généralement sur trois grandes catégories de jeux de données :

- Jeux de données synthétiques ou simulés, produits dans des environnements contrôlés pour reproduire le comportement d'utilisateurs et générer artificiellement des scénarios d'attaques.
- Jeux de données réels ou semi-réels, collectés dans des organisations partenaires, souvent anonymisés et limités pour des raisons de confidentialité.
- Jeux de données hybrides, combinant données réelles et scénarios simulés afin d'équilibrer réalisme et diversité des menaces.

Ces catégories répondent chacune à des besoins distincts que nous détaillerons dans la section ci-dessous. Toutefois, aucun jeu de données ne couvre parfaitement l'ensemble des défis liés à la menace interne malveillante : complexité des

comportements humains, diversité des contextes organisationnels et rareté des cas d'attaques malveillantes confirmées (Legg et al., 2015).

3.1. Types de jeux de données pour la détection de la menace interne malveillante

La recherche sur la détection des menaces internes malveillantes s'appuie sur différents types de jeux de données, qui se distinguent par leur origine, leur degré de réalisme, et leur accessibilité. Ces différences conditionnent directement la pertinence des analyses et la robustesse des modèles développés.

Jeux de données synthétiques

Les jeux synthétiques sont générés artificiellement à l'aide de modèles statistiques, de règles logiques ou de simulateurs multi-agents. Leur principal avantage est la possibilité de créer des scénarios variés et contrôlés. Cette caractéristique est cruciale, car dans la réalité, les comportements malveillants internes sont très rares et imprévisibles : sans données simulées, il serait difficile de tester un modèle sur un large éventail de situations (Chattopadhyay et al., 2018). Par exemple, simuler un employé qui tente progressivement de transférer des fichiers sensibles permet de vérifier si un algorithme peut détecter cette menace avant qu'elle ne devienne critique. Cela favorise également la reproductibilité des expériences, car les mêmes scénarios peuvent être générés plusieurs fois, permettant aux chercheurs de comparer de manière fiable la performance de différents algorithmes (Camino et al., 2020).

Un autre avantage est la flexibilité. Les chercheurs peuvent contrôler précisément la complexité des comportements et ajuster les paramètres pour tester des cas limites ou rares. Par exemple, ils peuvent augmenter la fréquence de comportements malveillants extrêmes pour vérifier si un système de détection les identifie correctement (Camino et al., 2020). Cette flexibilité est particulièrement utile pour entraîner des modèles d'apprentissage automatique, qui nécessitent souvent un grand nombre d'exemples pour apprendre efficacement.

Cependant, ces jeux synthétiques présentent des limites importantes. La première est la faible représentativité des comportements humains réels. Les modèles simulés reposent sur des règles ou hypothèses simplifiées qui ne capturent pas toujours les motivations, les imprévus ou les interactions sociales qui influencent réellement le comportement des employés (Magklaras & Furnell, 2020). En conséquence, un modèle performant sur des données synthétiques peut échouer dans le monde réel, où les schémas d'attaque sont souvent plus subtils ou inattendus.

La deuxième limite concerne le risque de biais. Si les règles ou paramètres utilisés pour générer les données ne sont pas correctement calibrés, certains types de comportements peuvent être surreprésentés et d'autres sous-représentés. Cela peut induire un apprentissage biaisé : par exemple, le modèle peut apprendre à détecter uniquement certains types d'anomalies et manquer d'autres formes d'attaques internes (Magklaras & Furnell, 2020).

Enfin, les données synthétiques manquent souvent de contexte organisationnel et social, comme les structures hiérarchiques, les politiques internes ou les interactions interpersonnelles qui peuvent influencer les comportements malveillants (Camino et al., 2020). Par conséquent, un modèle entraîné uniquement sur des données synthétiques peut détecter des anomalies statistiques, mais sans pouvoir distinguer celles qui sont réellement dangereuses pour l'organisation.

Jeux de données réels

Contrairement aux données synthétiques, les jeux de données réels sont particulièrement précieux car ils reflètent fidèlement les comportements réels des utilisateurs dans un contexte organisationnel. Leur principal avantage réside dans leur réalisme : les modèles entraînés sur ces données sont exposés à la vraie complexité des interactions humaines, des séquences d'accès aux fichiers, des communications entre employés et des usages des systèmes informatiques. Cette authenticité permet d'évaluer la performance des modèles dans des conditions proches de celles rencontrées dans la pratique, ce qui est essentiel pour s'assurer que les algorithmes

détectent correctement les comportements malveillants et minimisent les faux positifs (Greitzer et al., 2019).

Un autre avantage est la valeur empirique que ces jeux offrent pour la validation des modèles. Contrairement aux jeux synthétiques, qui reposent sur des hypothèses ou des règles artificielles, les données réelles permettent de tester la robustesse des algorithmes face à des comportements inattendus ou à des scénarios non prévus par les chercheurs. Par exemple, l'analyse de journaux anonymisés provenant d'une banque ou d'un réseau d'entreprise peut révéler des modèles subtils de fraude interne, que les données simulées ne captureraient pas forcément (Stolfo et al., 2012).

Cependant, ce type de jeux de données présente des limites significatives. La première est leur rareté : les contraintes éthiques, juridiques et de confidentialité rendent extrêmement difficile l'accès à des informations sensibles sur le comportement réel des employés. Les organisations hésitent souvent à partager ces données, même anonymisées, par crainte de violations de la vie privée ou de risques légaux (Greitzer et al., 2019).

La deuxième limite concerne la diffusion et la reproductibilité. Comme ces jeux sont rares et spécifiques à certaines organisations, il est difficile pour les chercheurs d'y accéder en grand nombre, ce qui limite la comparabilité des résultats et la possibilité de répéter les expériences à grande échelle (Mundie et al., 2021). De plus, même lorsque des données réelles sont disponibles, elles doivent être soigneusement anonymisées et adaptées, ce qui peut réduire leur richesse contextuelle et certaines informations cruciales pour la détection des menaces.

Enfin, la collecte et l'utilisation de données réelles impliquent des coûts élevés en termes de temps, de coordination et de conformité légale, ce qui limite leur utilisation pour des expérimentations extensives ou pour entraîner des modèles nécessitant de larges volumes de données.

Jeux de données simulés publics

Face à la rareté des données réelles, les chercheurs s'appuient très largement sur des jeux simulés publics, qui constituent une alternative pratique et accessible. Leur principal avantage réside dans leur disponibilité et leur accessibilité : contrairement aux données réelles, souvent inaccessibles pour des raisons de confidentialité et de sécurité, ces jeux peuvent être utilisés librement par la communauté scientifique, ce qui favorise la reproductibilité des recherches et la comparabilité des résultats entre différents modèles (Glasser & Lindauer, 2013 ; Legg et al., 2017).

Un autre avantage clé est la possibilité de créer et de contrôler des scénarios spécifiques. Par exemple, le jeu de données CERT Insider Threat contient des comportements simulés malveillants et bénins, comme le vol de données, le sabotage ou la fuite d'informations, générés à partir d'activités loguées (courriels, navigation, fichiers). De même, le jeu de données iTrust Insider Threat reproduit des comportements dans une plateforme bancaire fictive, tandis que le jeu de données Enron Email, bien qu'axé sur les communications, permet d'analyser les interactions entre employés (Parveen et al., 2011 ; Klimt & Yang, 2004). Ces jeux permettent aux chercheurs de tester et d'entraîner des algorithmes dans un cadre structuré et connu, de manière à évaluer la performance des modèles face à des scénarios variés et reproductibles.

Cependant, ces jeux présentent des limites importantes, principalement liées à leur réalisme. Étant simulés, ils ne capturent pas pleinement la complexité et la variabilité des comportements humains observés dans des environnements organisationnels réels. Par exemple, des schémas de communication ou des comportements d'accès aux fichiers qui surviennent dans des contextes spécifiques ou avec des motivations diverses peuvent être absents ou simplifiés dans les jeux simulés, ce qui limite la généralisation des modèles entraînés sur ces données (Nurse et al., 2020).

De plus, la simulation repose sur des hypothèses ou des règles définies par les chercheurs, ce qui peut introduire des biais artificiels si ces règles ne reflètent pas fidèlement les comportements réels. Les modèles formés sur ces données peuvent ainsi développer des stratégies de détection trop centrées sur des scénarios connus et manquer des comportements malveillants atypiques ou émergents.

Jeux de données hybrides

Les jeux de données hybrides sont conçus pour pallier les limites propres à chaque type de données : manque de réalisme des données simulées, rareté et contraintes d'accès des données réelles, ainsi que faible représentativité des données synthétiques. Leur principe consiste à combiner différentes sources, par exemple, des journaux réels anonymisés enrichis par des scénarios simulés ou générés artificiellement, afin d'obtenir à la fois un haut degré de réalisme et une couverture suffisante des cas de menace interne malveillante (Camino et al., 2020 ; Gavai et al., 2015 ; Nurse et al., 2020).

L'avantage majeur de cette approche réside dans sa capacité à offrir le meilleur des deux mondes. En intégrant des données réelles, les jeux hybrides conservent la richesse contextuelle et la crédibilité des comportements humains, ce qui renforce la validité externe des modèles entraînés. En parallèle, l'ajout de scénarios simulés permet d'introduire des menaces rarement observables dans la réalité, par exemple, un sabotage interne ou une fuite massive d'information, qui seraient pratiquement impossibles à collecter de manière empirique en raison de leur rareté ou de leur caractère sensible (Legg et al., 2017). Ainsi, les jeux hybrides accroissent la diversité des comportements représentés, ce qui améliore la généralité et la robustesse des modèles de détection, notamment dans la gestion de situations nouvelles ou atypiques (Tuor et al., 2017 ; Homoliak et al., 2019).

De plus, ces jeux offrent un équilibre entre réalisme et flexibilité : d'un côté, ils bénéficient de l'ancrage empirique apporté par les données organisationnelles réelles ; de l'autre, ils permettent aux chercheurs de manipuler et d'adapter les scénarios pour tester des hypothèses précises. Par exemple, certaines extensions du jeu de données CERT incorporent des traces comportementales réelles pour renforcer la crédibilité des scénarios simulés, ce qui permet de mieux reproduire la dynamique d'un environnement de travail (Legg et al., 2017).

Cependant, les limites des jeux hybrides sont à la hauteur de leurs ambitions. D'abord, la cohérence entre des sources hétérogènes est difficile à garantir : fusionner des journaux réels avec des données générées artificiellement peut créer des artefacts ou des incohérences temporelles qui nuisent à la crédibilité des scénarios. Ensuite, le

recours à des simulations ouvre la porte à l'introduction de biais artificiels, si les règles ou hypothèses sous-jacentes ne reflètent pas fidèlement les comportements humains (Nurse et al., 2020). Enfin, l'exigence de maintenir des scénarios plausibles et représentatifs impose une charge méthodologique importante. La création de jeux hybrides nécessite une collaboration étroite entre experts techniques, analystes de sécurité et spécialistes organisationnels pour concilier diversité des menaces, réalisme opérationnel et respect des contraintes de confidentialité (Baracaldo et al., 2021).

3.2. Jeux de données de référence dans la littérature

Dans le domaine de la détection des menaces internes malveillantes, plusieurs jeux de données dits de référence sont fréquemment mobilisés par la communauté scientifique. Ces jeux de données se distinguent par leur nature (réels, simulés, hybrides), leur granularité, et les types d'activités utilisateur qu'ils couvrent. Nous présentons ci-dessous les plus cités dans la littérature, notamment le jeu de données CERT Insider Threat, le jeu de données Enron Email, ainsi que des jeux plus spécifiques comme les jeux de données RUU et WUIL. Chacun apporte une perspective particulière sur la modélisation des comportements normaux et déviants en milieu organisationnel.

Jeu de données Enron Email

Le jeu de données Enron Email est l'un des corpus les plus utilisés dans la recherche sur les menaces internes malveillantes et l'analyse de la communication organisationnelle. Le tableau 2.1 présente ses principales colonnes, dont l'utilité analytique est détaillée dans les paragraphes suivants.

Tableau 2-1 : Description des principales colonnes du jeu de données Enron Email

Colonne principale	Description
Message-ID	Identifiant unique du courriel
Date	Date et heure d'envoi
From	Expéditeur

To, Cc, Bcc	Destinataires
Subject	Objet du message
Body	Contenu textuel courriel du courriel
Folder	Dossier de classement (Inbox, Sent, etc.)

Ce corpus regroupe entre 0,5 et 0,6 million de courriels, selon les versions retraitées, issus de la correspondance de plus de 150 employés de l'entreprise Enron. Ces messages ont été collectés dans le cadre de l'enquête judiciaire sur la faillite de l'entreprise par la Federal Energy Regulatory Commission (FERC), puis publiés par la Carnegie Mellon University (Klimt & Yang, 2004). Il constitue ainsi l'une des rares bases massives et authentiques de communications électroniques organisationnelles, ce qui en fait une ressource unique pour l'étude des interactions sociales, des réseaux internes et de l'analyse textuelle appliquée à la détection d'anomalies comportementales.

L'un des principaux avantages de ce corpus réside dans son authenticité. Contrairement aux jeux de données synthétiques ou simulés, il reflète des échanges réels entre employés, avec toute la spontanéité, l'ambiguïté et la richesse linguistique caractéristiques de la communication humaine. Cette authenticité permet aux chercheurs d'analyser des dynamiques interactionnelles crédibles et de modéliser des réseaux sociaux internes avec une précision difficile à atteindre avec des données artificielles (Klimt & Yang, 2004). De plus, le volume important des données (plus d'un demi-million de courriels) favorise le développement et la validation de modèles de traitement automatique du langage naturel et d'apprentissage automatique à grande échelle, ce qui constitue un atout pour des tâches telles que la détection de collusion, de fraude ou d'abus de confiance. Enfin, son caractère public et largement diffusé facilite la reproductibilité des recherches, un aspect central dans le domaine académique et scientifique (Diesner et al., 2005).

Cependant, le corpus présente plusieurs limites importantes. Premièrement, il se concentre exclusivement sur les communications par courriel, sans inclure d'autres sources comme les journaux d'authentification, les logs systèmes ou les traces d'accès

aux fichiers. Cette restriction limite sa capacité à capturer la diversité des signaux techniques et comportementaux associés aux menaces internes malveillantes (Diesner et al., 2005). Deuxièmement, le jeu de données n'est pas annoté pour les comportements malveillants : il ne contient pas de labels indiquant explicitement les cas de fraude ou d'abus. Cette absence d'étiquetage contraint les chercheurs à investir des efforts supplémentaires dans l'annotation manuelle, à recourir à des méthodes semi-supervisées, ou à enrichir le corpus par des scénarios simulés afin de rendre l'analyse exploitable pour la détection (Klimt & Yang, 2004). Enfin, le jeu de données présente des biais contextuels liés à son origine : il provient d'une seule organisation, évoluant dans un secteur particulier (l'énergie), et dans une situation de crise exceptionnelle. Ces conditions limitent la généralisation des résultats à d'autres environnements organisationnels plus stables ou à d'autres industries (Diesner et al., 2005).

Malgré ces limites, le jeu de données Enron Email demeure une ressource précieuse et incontournable pour la recherche en cybersécurité et en sciences sociales computationnelles. Sa disponibilité dans plusieurs versions retraitées et annotées, notamment via la Carnegie Mellon University et des plateformes comme Kaggle, facilite son exploitation dans des expérimentations d'apprentissage automatique et dans des projets appliqués cherchant à modéliser la communication organisationnelle et à détecter des signaux faibles de comportements anormaux.

Jeu de données RUU

Le jeu de données RUU (Relative User Utilisation) a été conçu pour analyser les comportements liés à l'usage anormal des systèmes informatiques, en mettant l'accent sur l'exploitation des ressources par les utilisateurs. Le tableau 2.2 présente ses principales colonnes, dont l'utilité analytique est détaillée dans les paragraphes suivants. Contrairement à des bases massives et multi-sources comme le jeu de données CERT Insider Threat, il adopte une approche ciblée, fondée sur les journaux d'activité relatifs aux sessions utilisateur, aux commandes exécutées et aux accès réalisés. Cette orientation permet une granularité fine, utile pour repérer des abus de privilèges ou des déviations par rapport au profil normal d'un individu (Tuor et al., 2017).

Tableau 2-2 : Description des principales colonnes du jeu de données RUU

Colonne principale	Description
User_ID	Identifiant anonymisé de l'utilisateur.
Timestamp	Date et heure précises de l'événement.
Activity_Type	Type d'activité réalisée (ex. login, file_access, courriel_sent, http_request).
Resource	Ressource ou objet cible de l'action (nom du fichier, URL, base de données).
Access_Result	Résultat de l'action (succès, échec, accès refusé).
Device/Terminal	Poste de travail ou terminal utilisé.
Location/IP	Adresse IP ou localisation du poste lors de l'activité.
Action_Detail	Métadonnées spécifiques (ex. taille du fichier transféré, sujet de courriel).
Label/Scenario	Indication si l'événement fait partie d'un scénario malveillant ou non.

Un des principaux avantages de ce jeu de données réside dans sa structuration détaillée des sessions utilisateur, qui inclut des variables comme l'heure de connexion, la durée des sessions, les processus exécutés et les patterns d'usage du système. Cette richesse facilite la modélisation de l'« empreinte numérique » d'un utilisateur, c'est-à-dire une représentation de son comportement habituel. Grâce à cela, les algorithmes d'apprentissage automatique peuvent mieux distinguer les anomalies, ce qui favorise une détection plus précise et personnalisée des comportements suspects. Par ailleurs, la simplicité et l'organisation claire des données en font un support méthodologique intéressant : il permet de tester rapidement différents algorithmes et de comparer leurs performances dans des conditions expérimentales bien contrôlées. Cela accélère les cycles de recherche et favorise la reproductibilité des résultats, un aspect essentiel dans un domaine où la validation scientifique est souvent freinée par l'hétérogénéité des jeux de données.

Cependant, ces atouts sont contrebalancés par des limites notables. Tout d'abord, le jeu de données RUU reste synthétique : les comportements injectés ne capturent pas la complexité des interactions humaines réelles dans un environnement de travail. En conséquence, les modèles entraînés sur ces données risquent de manquer de robustesse lorsqu'ils sont appliqués à des environnements réels où les menaces internes malveillantes adoptent des stratégies beaucoup plus subtiles et variées. Par exemple, le vocabulaire utilisé dans les commandes ou les séquences d'accès peut sembler artificiel, ce qui introduit un écart de réalisme entre les comportements simulés et ceux observés en production. Ensuite, le jeu de données couvre un nombre limité de scénarios, ce qui restreint sa capacité à représenter la diversité des comportements malveillants. Des menaces sophistiquées, comme l'exfiltration lente de données ou le sabotage dissimulé, ne sont pas représentées, ce qui réduit sa valeur pour l'entraînement d'algorithmes destinés à une application pratique large. Enfin, il souffre d'un manque d'informations contextuelles : il se focalise exclusivement sur les données techniques liées aux sessions et aux commandes, mais ignore des canaux essentiels comme les courriels, la messagerie instantanée ou les réseaux sociaux internes. Or, ces dimensions contextuelles sont cruciales pour interpréter les motivations des comportements et comprendre les dynamiques sociales à l'origine des menaces internes malveillantes.

Ainsi, le jeu de données RUU peut être considéré comme un outil complémentaire : il est particulièrement pertinent pour étudier l'anomalie technique et les usages systèmes dans un cadre académique ou expérimental, mais il ne saurait, à lui seul, offrir une vision complète de la menace interne. Pour cela, il doit être enrichi par des données plus réalistes et multi-dimensionnelles, qui capturent les interactions sociales et organisationnelles. Enfin, son accessibilité restreinte, puisqu'il n'est disponible qu'auprès des chercheurs originaux et sous conditions spécifiques, limite sa diffusion. Cela freine son adoption dans des projets de recherche de grande envergure et réduit la possibilité de comparaisons entre équipes de recherche.

Jeu de données WUIL

Le jeu de données WUIL (Windows User and Intruder Logs) a été conçu pour l'étude des attaques par usurpation d'identité (masquerade attacks). Il enregistre principalement la

navigation dans le système de fichiers sous Windows et insère, au sein de sessions légitimes, des activités d'intrus simulées (Camiña, Monroy, Trejo & Hernández-Gracidas, 2014). Le tableau 2.3 présente ses principales colonnes, dont l'utilité analytique est détaillée dans les paragraphes suivants. L'approche méthodologique adoptée repose sur une hypothèse centrale : ce qui distingue réellement un utilisateur légitime d'un intrus n'est pas tant la commande exécutée que l'objet ciblé, fichier, répertoire ou ressource manipulée. Cette perspective a conduit les auteurs à privilégier une modélisation centrée sur les objets plutôt que sur les actions, afin de mieux capturer les déviations fines dans les schémas d'accès (Camiña et al., 2014).

Tableau 2-3 : Description des principales colonnes du jeu de données WUL

Type de log	Colonnes principales
Web logs	Timestamp, UserID, URL, Action
Application logs	Timestamp, App name, Action performed
File logs	Timestamp, FileName, Operation
System logs	UserID, Computer, Logon/Logoff

Le corpus rassemble des journaux détaillés d'accès au système de fichiers, incluant l'horodatage, l'identifiant utilisateur, la machine concernée, le chemin de l'objet, le type d'opération (lecture, écriture, suppression), le succès ou l'échec de l'action, et parfois le processus appelant. Ces données, collectées par l'audit Windows sur des postes utilisateurs pendant plusieurs jours, ont ensuite été enrichies par des scénarios d'intrusion conçus par des opérateurs selon des stratégies réalistes d'exfiltration ou de repérage. Le jeu de données WUIL a été mobilisé dans diverses évaluations expérimentales, notamment pour la détection via partitionnement de graphes d'événements d'accès fichiers, atteignant des performances élevées ($AUC \approx 0,94$ et $TPR > 95 \%$ selon les paramètres), ce qui illustre son intérêt comme banc d'essai pour comparer différentes approches (Homoliak, Huraj & Hanuliak, 2018).

Sur le plan méthodologique, le WUIL présente plusieurs atouts clés. Sa granularité fine, chaque opération de fichier étant enregistrée avec son contexte précis, facilite la

modélisation des régularités d'usage. Cela permet de capter la localité spatiale et temporelle des accès, c'est-à-dire la tendance des utilisateurs à consulter certains ensembles de fichiers dans des fenêtres temporelles limitées. Ce niveau de détail rend possible l'apprentissage d'empreintes comportementales précises, utiles pour identifier des déviations même subtiles (Camiña et al., 2014).

De plus, sa focalisation sur les objets plutôt que sur les commandes rend l'approche robuste à la variabilité syntaxique : deux utilisateurs peuvent réaliser la même tâche avec des commandes différentes, mais les objets ciblés révèlent des régularités plus stables. Cela en fait un jeu de données particulièrement adapté aux modèles séquentiels (RNN, LSTM) et aux graphes d'interactions objets-utilisateurs (Homoliak et al., 2018). Enfin, la présence conjointe de traces légitimes et de sessions d'intrus scénarisées est un avantage important : elle permet de tester la capacité discriminante des algorithmes dans des scénarios réalistes, y compris dans le cas de masquerades « low-and-slow », où l'intrus agit de manière furtive pour rester indétectable par des règles simples.

Cependant, plusieurs limites réduisent sa portée. Premièrement, même si les traces reposent sur des journaux Windows réels, les intrusions sont simulées par des opérateurs. Cela introduit un risque d'artificialité : les styles d'accès peuvent paraître moins « organiques » que ceux d'un véritable intrus, et les scénarios se concentrent sur un périmètre restreint (accès fichiers) au détriment d'autres canaux pertinents comme les courriels, les logs d'authentification ou le trafic réseau. Cette focalisation limite la transférabilité du jeu de données à des environnements organisationnels complexes (Zhou et al., 2021).

Deuxièmement, son accessibilité restreinte pose un problème méthodologique : il est disponible uniquement sous NDA et sur demande directe auprès des auteurs, et sa documentation publique reste partielle. Cela rend la reproductibilité des expériences plus difficile et empêche la comparaison transparente des prétraitements entre équipes de recherche (Camiña et al., 2014 ; Homoliak et al., 2018).

Enfin, il est important de rappeler que le WUIL n'avait pas été conçu initialement pour la détection des menaces internes malveillantes, mais pour l'étude des usages collaboratifs et de la productivité (Zhou et al., 2021). Son exploitation en cybersécurité

implique donc une adaptation des scénarios et des objectifs d'analyse, ce qui peut conduire à des décalages entre les hypothèses de départ et les besoins réels en matière de détection.

En somme, le jeu de données WUIL constitue une ressource précieuse pour l'étude des anomalies comportementales dans les systèmes de fichiers grâce à son niveau de détail et sa capacité à supporter des approches avancées (séquentielles, graphes, apprentissage d'empreintes). Toutefois, ses limites structurelles, caractère simulé des intrusions, portée réduite aux accès fichiers, disponibilité restreinte, appellent à une utilisation prudente. Dans le cadre des menaces internes malveillantes, il doit souvent être complété par d'autres sources pour refléter la complexité réelle des environnements organisationnels.

Un exemple notable d'application du jeu de données WUIL est proposé par Homoliak, Huraj et Hanuliak (2018), qui ont utilisé ce corpus pour tester une approche de détection fondée sur le partitionnement de graphes d'événements d'accès fichiers. Dans leur modèle, les utilisateurs et les objets (fichiers, répertoires) sont représentés comme des nœuds dans un graphe, et chaque interaction (lecture, écriture, suppression) constitue une arête pondérée. Cette représentation permet d'analyser les schémas relationnels entre utilisateurs et ressources, et d'identifier des anomalies sous forme de sous-graphes atypiques.

L'expérience a montré que le WUIL est particulièrement adapté à ce type d'approches car il fournit des traces détaillées de navigation dans les fichiers, permettant de détecter des déviations comportementales subtiles. Les chercheurs rapportent des performances élevées, avec une AUC avoisinant 0,94 et un TPR (taux de vrais positifs) supérieur à 95 %, ce qui illustre la valeur opérationnelle du jeu de données comme banc d'essai pour comparer différentes méthodes de détection.

Cet exemple indique que, malgré ses limites (intrusions simulées, périmètre restreint aux accès fichiers), le WUIL peut servir de laboratoire expérimental pour tester des techniques avancées telles que l'apprentissage séquentiel (RNN, LSTM) ou les graphes d'interactions, en particulier dans le contexte des attaques de type masquerade, où

l'objectif est de distinguer un intrus d'un utilisateur légitime sur la base de l'usage des ressources.

Jeu de données CERT Insider Threat

Le jeu de données CERT Insider Threat, développé par le Software Engineering Institute de Carnegie Mellon University, est reconnu comme l'une des références majeures pour l'étude des menaces internes malveillantes.

Le tableau 2.4 présente ses principales colonnes, dont l'utilité analytique est détaillée dans les paragraphes suivants. L'un de ses principaux atouts réside dans sa richesse et sa diversité. En effet, il existe en plusieurs versions successives (v6, v7, v9, v10, v11), chacune intégrant de nouveaux scénarios pour couvrir différents types d'incidents internes. Les logs incluent des activités variées (courriels, navigation web, transferts de fichiers, accès système, authentications), ce qui en fait un jeu multi-sources très complet (Glasser & Lindauer, 2013). Cette granularité et cette documentation détaillée des scénarios constituent un atout majeur, car elles permettent de tester différentes approches de détection (règles, apprentissage supervisé, détection d'anomalies) dans un cadre contrôlé et reproductible (Harilal et al., 2017). Cela facilite notamment la comparaison des modèles entre chercheurs et favorise la répliquabilité scientifique, ce qui est souvent difficile avec des données réelles, rarement accessibles pour des raisons de confidentialité et de sécurité.

Tableau 2-4 : Description des principales colonnes du jeu de données CERT

Type de log	Colonnes principales	Description
Courriel logs	Date, UserID, Sender, Recipient, Subject, Content	Métadonnées et contenu des courriels échangés entre employés
Web logs	Date, UserID, URL, Activity	Navigation web des utilisateurs
Logon/logoff logs	Date, UserID, Computer, Activity (Logon/Logoff)	Authentications aux postes de travail

File access logs	Date, UserID, FileName, Activity (Read/Write/Delete)	Interactions avec les fichiers du serveur
Device logs	Date, UserID, DeviceType (USB, etc.), Activity (Connect/Disconnect)	Usage de périphériques amovibles
Psychological tests	UserID, Survey responses	Données contextuelles sur les traits de personnalité (dans certaines versions)

Un exemple concret d'application est fourni par Tuor et al. (2017), qui ont utilisé le jeu de données CERT v6 pour entraîner un modèle basé sur des réseaux de neurones récurrents (RNN). Leur objectif était de détecter les comportements suspects à partir des séquences d'actions des employés. Grâce à la richesse multi-sources des données, leur modèle a pu apprendre des motifs temporels (ex. : accès anormal aux fichiers après plusieurs échecs d'authentification) et a montré de bonnes performances comparées à des approches plus classiques. Cette étude démontre que le jeu de données CERT constitue un excellent terrain d'expérimentation pour développer et évaluer de nouvelles techniques de détection.

Cependant, ce même caractère synthétique constitue également son principal inconvénient. Comme il est simulé, le jeu de données ne reflète pas fidèlement la complexité des comportements humains réels. Par exemple, les courriels générés automatiquement utilisent un langage artificiel, beaucoup plus standardisé et moins ambigu que dans de véritables échanges professionnels, ce qui peut biaiser l'efficacité des modèles entraînés (CMU-SEI, 2021). De plus, même si plusieurs scénarios d'attaques internes (fraude, sabotage, fuite de données) sont intégrés, ils restent prédéfinis par les concepteurs. Cela limite la représentativité de la diversité des comportements réels, qui sont souvent plus subtils, imprévisibles et influencés par des dynamiques organisationnelles complexes (Nurse et al., 2020). Ainsi, le jeu de données CERT est un outil très utile pour expérimenter et établir des benchmarks, mais il doit être complété par des données réelles pour renforcer la validité externe des recherches.

Un exemple notable d'application du jeu de données CERT Insider Threat est présenté par Tuor et al. (2017), qui ont exploité le corpus pour tester une approche de détection des menaces internes malveillantes basée sur les réseaux de neurones récurrents (RNN). Dans leur modèle, chaque utilisateur est représenté par une séquence chronologique d'actions, incluant les accès à des fichiers, les transferts de données, les connexions aux systèmes et l'envoi de courriels. L'objectif est d'identifier des comportements anormaux par rapport aux motifs habituels d'utilisation, en capturant les variations subtiles dans le temps et les combinaisons d'activités qui peuvent indiquer une menace interne malveillante.

L'expérience a montré que le jeu de données CERT est particulièrement adapté à ce type d'approches, car il fournit des logs multi-sources et des scénarios documentés de comportements malveillants simulés. Cela permet au modèle d'apprendre des patterns complexes et d'évaluer des métriques de performance telles que le taux de vrais positifs (TPR) et le taux de faux positifs (FPR), offrant ainsi un cadre rigoureux pour tester et comparer différentes méthodes de détection.

Cet exemple illustre que, malgré ses limites (scénarios simulés, langage artificiel des courriels, comportement humain moins varié que dans la réalité), le jeu de données CERT constitue un outil précieux pour développer et expérimenter des techniques avancées d'apprentissage supervisé ou non supervisé, en particulier dans des contextes où la combinaison de multiples sources de données est essentielle pour détecter des menaces internes malveillantes complexes.

3.3. Utilisations du jeu de données CERT dans la littérature

Le jeu de données CERT Insider Threat a été largement exploité dans la littérature pour évaluer et développer des méthodes de détection des menaces internes malveillantes, en raison de sa richesse multi-sources et de la diversité des scénarios proposés. Plusieurs études se sont appuyées sur ce jeu de données pour tester des approches variées, allant des modèles classiques d'apprentissage supervisé à des architectures plus sophistiquées d'apprentissage profond.

Apprentissage supervisé et méthodes classiques

De nombreuses études se sont appuyées sur CERT pour entraîner des modèles supervisés standards tels que la régression logistique, les machines à vecteurs de support (SVM) ou encore les forêts aléatoires (Chandola et al., 2009 ; Harilal et al., 2017). Ces méthodes exploitent directement les variables issues des logs (par exemple : nombre de fichiers accédés, fréquence des connexions, volume de courriels envoyés).

L'un des avantages majeurs de ces modèles est leur interprétabilité : les coefficients de la régression logistique ou l'importance des variables dans une forêt aléatoire permettent de comprendre quels comportements contribuent le plus à la détection d'une menace. Cela favorise leur utilisation dans un cadre organisationnel, où les analystes de sécurité exigent souvent des explications transparentes pour justifier les alertes générées.

Cependant, ces approches montrent leurs limites lorsqu'il s'agit de capturer la dimension temporelle ou les séquences d'actions corrélées. Par exemple, accéder à un fichier sensible peut être normal pour un employé, mais répété juste après plusieurs échecs de connexion, cela devient suspect, ce type de dynamique échappe largement aux modèles classiques. C'est ce constat qui a ouvert la voie à des approches séquentielles et profondes.

Approches basées sur l'apprentissage séquentiel et profond

Un tournant majeur dans la littérature est illustré par l'étude de Tuor et al. (2017), qui ont utilisé CERT v6 pour entraîner des réseaux de neurones récurrents (RNN). Leur objectif était de modéliser les séquences chronologiques d'actions (authentifications, accès fichiers, transferts de données, courriels). Les RNN se sont révélés capables de capturer des motifs temporels complexes, tels qu'une succession inhabituelle de connexions à des heures non ouvrées suivie d'un téléchargement massif de fichiers.

D'autres recherches ont étendu cette logique :

- Des autoencodeurs ont été utilisés pour reconstruire les comportements « normaux » et identifier les écarts significatifs (Nurse et al., 2020) ;

- Des modèles basés sur les graphes ont représenté les relations entre utilisateurs, fichiers et ressources afin de détecter des anomalies structurelles dans les interactions (Le et al., 2019) ;
- Des architectures hybrides CNN-LSTM ont combiné l'extraction de motifs locaux avec la capture des dépendances temporelles à long terme (Yuan et al., 2018).

Ces travaux montrent que le CERT se prête bien à l'expérimentation avec l'apprentissage profond, grâce à la richesse de ses logs multi-sources. Toutefois, ces modèles nécessitent d'importantes ressources computationnelles et leur opacité (boîte noire) complique leur adoption en contexte opérationnel.

Détection d'anomalies et méthodes non supervisées

Un autre courant de recherche a privilégié l'apprentissage non supervisé, exploitant CERT sans recourir à des labels explicites. L'idée est de modéliser les comportements normaux et de signaler toute anomalie.

Par exemple, Glasser & Lindauer (2013) ont utilisé CERT pour identifier des anomalies dans les séquences d'accès et les volumes de données transférées, sans connaissances préalables sur les menaces. Ces approches sont particulièrement utiles face à des attaques nouvelles ou rares, qui ne sont pas incluses dans les scénarios prédéfinis.

Les méthodes mobilisées incluent :

- Le clustering (K-means, DBSCAN) pour regrouper les comportements similaires et détecter ceux qui s'écartent fortement ;
- Les modèles probabilistes (Hidden Markov Models, Gaussian Mixture Models) pour estimer la vraisemblance des séquences d'actions ;
- Des techniques modernes comme l'Isolation Forest ou les autoencodeurs variationnels pour capturer des anomalies multivariées complexes.

Toutefois, la principale limite de ces travaux est liée au caractère synthétique des données CERT : les anomalies détectées sont conditionnées par les règles de

simulation, et il existe un risque que les modèles surréagissent à des écarts artificiels, peu représentatifs de la réalité organisationnelle.

Limites observées dans la littérature

Malgré son importance, le jeu de données CERT fait l'objet de critiques récurrentes dans la littérature scientifique. Plusieurs auteurs soulignent que les courriels générés artificiellement, les scénarios prédéfinis et la nature synthétique des comportements humains introduisent des biais susceptibles de fausser l'évaluation de la performance des modèles (CMU-SEI, 2021). Les courriels, par exemple, sont produits selon des règles simplifiées et manquent de la richesse linguistique, des ambiguïtés et des subtilités relationnelles présentes dans des échanges professionnels réels. Cette artificialité peut induire en erreur les modèles de traitement du langage naturel, qui apprennent à repérer des signaux inexistant dans la pratique et dont la robustesse s'avère limitée en contexte opérationnel. De la même manière, les scénarios prédéfinis enferment les comportements malveillants dans un nombre restreint de cas, vol de données, sabotage ou fuite d'informations, ce qui limite la capacité des modèles à détecter des attaques atypiques ou inédites. Les algorithmes tendent alors à se spécialiser sur des profils de menaces connus, mais échouent à reconnaître des comportements émergents qui s'écartent de ces schémas. Enfin, le caractère entièrement synthétique des comportements humains simulés constitue une autre faiblesse : les employés fictifs réagissent selon des scripts déterminés, ce qui empêche de capturer la diversité des motivations, des contradictions et des influences sociales qui caractérisent le comportement réel des individus en entreprise (Magklaras & Furnell, 2020).

Un second point critique est le manque de réalisme organisationnel. Dans CERT, un employé peut changer de comportement de manière soudaine, passant d'un usage bénin du système à un acte malveillant, uniquement parce qu'un scénario le prévoit. Dans les organisations réelles, de tels comportements ne surgissent pas de façon abrupte, mais s'inscrivent généralement dans une dynamique sociale et professionnelle plus complexe : conflits hiérarchiques, pressions liées à la performance, frustrations accumulées ou réorganisations internes. Cette absence de gradualité réduit la

pertinence des modèles entraînés sur CERT, qui peinent ensuite à détecter les signaux faibles annonciateurs d'un passage à l'acte (Nurse et al., 2020).

Enfin, plusieurs travaux insistent sur la difficulté à généraliser les résultats obtenus à partir du jeu de données CERT à d'autres environnements organisationnels (Baracaldo et al., 2021). Les logs informatiques, les cultures d'entreprise et les modes de communication varient considérablement d'une organisation à l'autre. Ainsi, un modèle performant sur CERT peut échouer lorsqu'il est appliqué à un contexte différent, révélant une validité externe limitée. Cette contrainte rappelle la nécessité de compléter CERT par des données plus contextuelles ou hybrides, capables de refléter de manière plus fidèle la diversité et la complexité des environnements organisationnels réels.

À ces limites s'ajoutent plusieurs biais structurels relevés dans la littérature. D'abord, la distribution des types d'événements dans CERT est loin d'être uniforme : certaines catégories d'activité (par exemple les requêtes HTTP ou les connexions) sont surreprésentées, tandis que d'autres, pourtant centrales en milieu réel, comme les changements d'habilitations ou les interactions avec des systèmes critiques, sont peu fréquentes ou peu détaillées. Ensuite, la labellisation strictement binaire des utilisateurs ("malveillant" versus "non malveillant") efface la zone grise entre erreur, négligence et malveillance intentionnelle, pourtant largement documentée dans les études empiriques sur les menaces internes. Enfin, plusieurs auteurs, notamment Brown et al. (2017), Tuor et al. (2017) ainsi que Nurse et al. (2020), soulignent que les comportements malveillants simulés dans CERT ont tendance à se concentrer dans des fenêtres temporelles courtes et fortement visibles, ce qui contraste avec la nature plus diffuse, graduelle et bruitée des trajectoires observées dans les journaux organisationnels réels.

Dans la littérature, le jeu de données CERT apparaît comme un banc d'essai incontournable pour les modèles de détection de menaces internes malveillantes. Il a permis d'explorer une grande variété d'approches, des plus simples aux plus avancées, et de stimuler les progrès méthodologiques dans le domaine. Cependant, les limites soulevées invitent à le considérer davantage comme un point de départ que comme une solution définitive.

C'est dans cette perspective que s'inscrit notre projet : proposer des pistes d'amélioration du jeu de données CERT, afin de combler l'écart entre données simulées et dynamiques réelles. Cet enrichissement permettra de développer et tester des modèles plus robustes, capables de mieux refléter la diversité et la complexité des comportements internes malveillants dans des environnements organisationnels concrets.

3.4 Conclusion

La disponibilité et la qualité des jeux de données constituent des facteurs déterminants pour la détection des menaces internes malveillantes. Comme le montrent les travaux cités, chaque type de jeu de données, synthétique, réel, simulé public ou hybride, présente des avantages et des limites distincts. Les jeux synthétiques offrent flexibilité et reproductibilité, mais manquent de réalisme ; les données réelles reflètent fidèlement les comportements humains mais restent rares, coûteuses et difficiles à partager ; les jeux simulés publics permettent la comparabilité entre études mais simplifient la complexité des interactions organisationnelles ; enfin, les jeux hybrides combinent les points forts des précédents, tout en exigeant un équilibre délicat pour garantir cohérence et crédibilité.

Dans ce paysage, le jeu de données CERT Insider Threat occupe une place de référence : sa richesse multi-sources, la diversité de ses scénarios et sa structuration en font un banc d'essai incontournable pour comparer et développer des modèles de détection. Néanmoins, son caractère simulé et les biais liés à la nature artificielle des comportements qu'il génère rappellent qu'il doit être complété par d'autres ressources. Des corpus comme Enron Email, RUU ou WUIL élargissent les perspectives, mais aucun ne reflète pleinement la complexité des environnements organisationnels réels.

Ces constats justifient l'importance d'enrichir les bases de données existantes et de concevoir des ressources hybrides, plus représentatives et éthiquement exploitables. C'est dans cette logique que s'inscrit notre travail, qui vise à proposer des pistes d'amélioration du jeu de données CERT afin de réduire l'écart entre dynamiques simulées et comportements observés en pratique.

Chapitre 3. Méthodologie

1. Choix de la méthodologie et justifications

Cette section présente et justifie les choix méthodologiques retenus pour répondre à la question de recherche, centrée sur l'évaluation de la représentativité et de la pertinence du jeu de données CERT dans le contexte des menaces internes malveillantes. La démarche repose sur une approche mixte combinant, d'une part, un protocole expérimental visant à tester différentes méthodes de détection sur le jeu CERT, et d'autre part, une collecte de données qualitatives auprès d'experts en cybersécurité. Cette combinaison permet de mettre en perspective les résultats obtenus dans un environnement contrôlé avec les expériences et observations issues du terrain. Deux éléments sont donc détaillés dans cette section : (1) la justification du recours à un protocole expérimental sur le jeu CERT et (2) la justification du recours à des entrevues qualitatives auprès d'experts.

1.1 Justification du protocole expérimental

Le choix d'un protocole expérimental s'explique par la nature même de la question de recherche, qui implique de tester et d'évaluer la performance de différentes méthodes de détection sur un jeu de données largement utilisé dans la littérature. CERT constitue en effet un banc d'essai reconnu, standardisé et reproductible, permettant de comparer objectivement les résultats obtenus avec différents algorithmes (CMU-SEI, 2021). Cette logique s'inscrit dans une démarche positiviste où les faits observés, ici la performance mesurée à l'aide d'indicateurs précis, permettent d'analyser la robustesse et les limites des méthodes existantes.

De plus, le protocole expérimental répond à la nécessité de valider la fiabilité des approches de détection dans un environnement artificiel, avant de confronter ces résultats à des données contextuelles issues du terrain. En ce sens, la logique adoptée est déductive : il s'agit de tester des hypothèses issues de la littérature, selon lesquelles CERT peut servir de référence, et d'évaluer dans quelle mesure ces hypothèses se

vérifient ou se trouvent limitées. Enfin, ce choix est justifié par l'avantage de reproductibilité qu'offre le protocole expérimental, un critère essentiel pour comparer les résultats de ce mémoire avec ceux d'études antérieures.

1.2 Justification du recours aux entrevues qualitatives

Afin de compléter l'expérimentation sur le jeu de données CERT, des entrevues semi-structurées seront menées auprès de professionnels de la cybersécurité. Le recrutement des participants sera réalisé par courriel, à partir d'une invitation formelle envoyée aux contacts identifiés dans des réseaux professionnels et institutionnels. Ce choix repose sur le constat que CERT, bien qu'utile comme référence académique, ne reflète pas fidèlement la complexité des dynamiques organisationnelles où s'inscrit la menace interne malveillante (Nurse et al., 2020). Les entrevues permettent ainsi de recueillir des perceptions, expériences et jugements professionnels qui viennent enrichir et nuancer l'analyse expérimentale.

La logique suivie est ici inductive : plutôt que de tester des hypothèses préétablies, les entrevues visent à faire émerger des éléments de compréhension ancrés dans la réalité organisationnelle, tels que les pratiques de sécurité, les comportements d'employés ou les contraintes techniques qui échappent à un jeu de données synthétique. Cette approche qualitative permet de pallier la faible validité externe de CERT en offrant une perspective complémentaire issue du terrain.

Enfin, cette combinaison méthodologique permet non seulement de confronter différentes méthodes de détection sur le jeu de données CERT (S.O.1) et d'intégrer la perspective des praticiens en cybersécurité (S.O.2), mais également de croiser ces deux sources d'information avec les connaissances issues de la littérature scientifique afin d'analyser de manière approfondie la représentativité, les forces et les limites de CERT (S.O.3). Ce dispositif méthodologique vise donc une compréhension plus complète du phénomène des menaces internes malveillantes, en tenant compte à la fois des apports de l'expérimentation, des retours du terrain et des travaux académiques. Enfin, cette démarche offre un cadre propice à la formulation de recommandations concrètes pour enrichir le jeu de données CERT par l'ajout de nouvelles variables ou scénarios issus des

réalités organisationnelles, contribuant ainsi à accroître sa valeur scientifique et opérationnelle (S.O.4).

2. Protocole expérimental sur le jeu de données CERT

L'évaluation du réalisme et de la pertinence du jeu de données CERT r4.2 repose sur la mise en œuvre d'un protocole expérimental rigoureux, conçu pour examiner la dynamique temporelle des comportements simulés et leur cohérence avec les mécanismes décrits dans la littérature scientifique et par les experts rencontrés en entrevue. Les travaux en détection des menaces internes soulignent que les comportements malveillants ne se manifestent généralement pas sous forme d'actions isolées, mais émergent au contraire de séquences progressives, d'accumulations discrètes ou de dérives comportementales qui s'étendent dans le temps (Cappelli et al., 2012 ; Nurse et al., 2020). Dans cette perspective, l'analyse fonctionnelle des données (Functional Data Analysis, FDA) constitue un cadre particulièrement adapté, puisqu'elle permet de représenter les comportements non pas comme des points indépendants, mais comme des trajectoires continues décrivant l'évolution d'un employé sur plusieurs semaines.

Le choix de la FDA répond à un double enjeu méthodologique. D'une part, il s'agit de déterminer dans quelle mesure les comportements simulés dans CERT présentent une structure temporelle réaliste, c'est-à-dire des routines, des fluctuations et des dérives compatibles avec les observations rapportées dans les environnements organisationnels réels. D'autre part, ce protocole vise à évaluer si les signaux malveillants intégrés dans CERT reproduisent les caractéristiques progressives décrites dans la littérature, notamment les attaques de type *low and slow*, les phases préparatoires d'exfiltration ou encore les changements subtils dans les habitudes de travail, ou s'ils résultent au contraire de ruptures artificielles produites par la simulation.

Afin de répondre à ces objectifs, le protocole s'articule en plusieurs étapes successives et interdépendantes. La première consiste à préparer et transformer les données brutes de CERT en un corpus cohérent, structuré et exploitable pour une analyse fonctionnelle

(section 1.3.1). La deuxième étape vise à reconstruire, pour chaque utilisateur et pour chaque type d'activité, des trajectoires temporelles lissées qui serviront de fondement à la FDA (section 1.3.2). La troisième repose sur l'application d'une Functional Principal Component Analysis (FPCA), qui résume la variabilité des comportements à travers un petit nombre de composantes principales fonctionnelles (section 1.3.3). La quatrième mobilise des méthodes de classification supervisée, notamment la régression logistique et les forêts aléatoires, appliquées aux scores issus de la FPCA (section 1.3.4). Enfin, les résultats sont interprétés à la lumière de la littérature et des propos des experts, afin d'évaluer de manière critique le réalisme des comportements simulés et la structure des signaux malveillants (section 1.3.5).

Ainsi structuré, le protocole expérimental vise non seulement à mesurer la performance des méthodes fonctionnelles sur CERT, mais surtout à évaluer la capacité du jeu à reproduire de manière réaliste les dynamiques organisationnelles et comportementales liées aux menaces internes malveillantes.

2.1 Préparation et agrégation des données CERT

La préparation du jeu CERT commence par l'importation des différentes tables qui composent l'environnement simulé : journaux d'authentification (logon/logoff), d'utilisation de périphériques (USB), d'accès à des fichiers, d'échanges de courriels, de navigation Web (HTTP), ainsi que données contextuelles sur les employés (LDAP, psychométrie). Ces tables sont intégrées dans un environnement de traitement unifié, puis harmonisées au niveau des formats de date, des identifiants d'utilisateurs et des structures de colonnes.

a) Nettoyage et contrôle de qualité des données

Une fois importées, les données doivent être soumises à un processus de nettoyage systématique afin de corriger ou documenter les incohérences. Les dates sont converties au format datetime standardisé (ISO 8601), ce qui permet une harmonisation dans le traitement des horodatages entre fichiers. Des contrôles sont effectués pour identifier les valeurs manquantes (par exemple, utilisateurs anonymes ou champs URL

vides), les doublons (enregistrements identiques répétés), ainsi que les incohérences de type (par exemple, activité USB non associée à un utilisateur).

Dans un cadre académique, cette étape vise à minimiser le risque de biais liés à la qualité des données, tout en documentant les choix méthodologiques : exclusion des lignes incohérentes, remplacement par des valeurs neutres, ou création d'indicateurs binaires signalant les données manquantes (Harilal et al., 2017). Le nettoyage contribue à renforcer la validité interne de l'expérimentation en s'assurant que les comportements analysés correspondent effectivement à des activités plausibles dans un contexte organisationnel.

b) Construction d'une granularité temporelle cohérente

La dimension temporelle est centrale dans l'analyse des menaces internes malveillantes, car les comportements suspects se déploient rarement en un seul événement isolé. Ils émergent plutôt de séquences répétées ou progressives, comme plusieurs échecs de connexion suivis d'un accès réussi à des ressources sensibles, ou une activité USB inhabituelle précédant un envoi massif d'e-mails externes. Pour capturer cette dynamique, il est nécessaire de définir une granularité temporelle cohérente qui permette à la fois d'agréger l'information et de préserver les régularités comportementales.

Dans ce travail, une nouvelle variable `year_week` a été créée à partir de l'horodatage de chaque événement, afin de regrouper l'ensemble des activités d'un employé par semaine. L'unité d'analyse retenue est donc l'utilisateur-semaine : chaque observation correspond au profil comportemental d'un individu pour une période standardisée de sept jours. Cette structuration permet de comparer les comportements d'un même utilisateur dans le temps, tout en facilitant la détection d'anomalies relatives à son rythme habituel.

Le choix de la granularité hebdomadaire repose sur un compromis méthodologique. Une agrégation journalière aurait pu fournir une vision très fine des comportements, mais au prix d'une forte variabilité quotidienne (bruit), qui aurait rendu la détection instable. À

l'inverse, une agrégation mensuelle aurait dilué les signaux faibles et masqué des séquences d'attaques progressives de type low and slow, caractéristiques des menaces internes malveillantes (Tuor et al., 2017). La semaine constitue une unité intermédiaire pertinente, car elle reflète à la fois les cycles naturels de travail en organisation (horaires, jours ouvrés, rythmes hebdomadaires) et les fenêtres typiques d'évolution des comportements suspects (accumulation d'événements avant une action malveillante).

Cette granularité présente aussi des avantages pratiques pour l'expérimentation. Elle permet d'obtenir un volume suffisant d'observations pour l'entraînement des modèles tout en maintenant un niveau de détail adapté à la détection d'anomalies. Enfin, elle garantit la comparabilité entre individus, puisque tous les profils sont observés selon la même fenêtre temporelle.

c) Étiquetage des événements et identification des utilisateurs malveillants

La constitution de la variable cible (y) repose sur les scénarios fournis par CERT dans le dossier "answers", qui identifie explicitement les employés impliqués dans des comportements malveillants simulés. Ces utilisateurs sont étiquetés comme malicious (1), tandis que l'ensemble des autres reçoivent le label benign (0). Cet étiquetage est appliqué uniformément à tous les enregistrements associés à un même utilisateur, garantissant une vérité terrain simulée (ground truth) qui rend possible l'entraînement de modèles supervisés.

Toutefois, le jeu CERT ne met pas directement à disposition de variables prédictives exploitables : il fournit des journaux bruts d'activités (connexions, envoi de courriels, accès à des fichiers, navigation Web, etc.). Une étape de "feature engineering" s'est donc avérée indispensable afin de transformer ces logs en indicateurs quantitatifs pertinents. Dans ce travail, chaque profil hebdomadaire utilisateur a été résumé au moyen de douze variables comportementales dérivées, élaborées pour capturer les dimensions essentielles des activités simulées.

Les premières variables concernent les comportements d'authentification, mesurés par le nombre de connexions (nb_connect), de déconnexions (nb_disconnect), le volume

total d'événements de logon (`total_events_logon`), ainsi que les connexions et déconnexions réussies (`nb_logon`, `nb_logoff`). Les interactions avec les périphériques externes sont quantifiées à travers le nombre total d'événements correspondants (`total_events_device`) et le nombre distinct de périphériques utilisés (`nb_pc_device`). Les activités liées au système de fichiers sont représentées par le volume total d'événements (`total_events_file`) et le nombre distinct de fichiers consultés ou modifiés (`nb_pc_file`).

Les comportements en matière de messagerie sont décrits par le nombre total de courriels envoyés (`total_emails`), le nombre de pièces jointes (`nb_attachments`). Enfin, l'activité Web est représentée par le volume total de requêtes HTTP (`http_total_events`).

Cette démarche présente un double intérêt. D'une part, elle réduit la complexité des données brutes en créant des variables structurées et comparables d'un utilisateur à l'autre. D'autre part, elle introduit la possibilité d'exploiter des combinaisons de comportements pour détecter des menaces : par exemple, deux utilisateurs peuvent présenter un nombre total d'événements comparable sur une semaine donnée, tout en se distinguant fortement par la composition de leurs activités, l'un concentrant ses événements sur les connexions et l'autre sur les accès aux fichiers ou la navigation Web. Cette hétérogénéité des profils comportementaux est précisément ce que les variables agrégées hebdomadaires permettent de capturer.

Néanmoins, cette labellisation présente aussi une limite : elle ne reflète pas toujours les zones grises ou comportements ambigus, où un employé peut manifester des activités atypiques sans pour autant être malveillant (Nurse et al., 2020). Le recours au “feature engineering” vise justement à enrichir la modélisation par des variables intermédiaires permettant de capter ces signaux faibles, tout en conservant une logique supervisée fondée sur les scénarios définis dans CERT.

2.2 Construction des trajectoires fonctionnelles

À partir de cette structuration en unités *utilisateur-semaine*, l'ensemble des variables comportementales dérivées précédemment est agrégé de manière à constituer, pour

chaque employé, un profil hebdomadaire synthétisant ses activités. Cette représentation permet de comparer des profils hebdomadaires entre utilisateurs tout en fournissant une base standardisée et directement exploitable par les modèles supervisés

Le tableau 3.1 présente l'ensemble des variables hebdomadaires utilisées dans la construction des trajectoires fonctionnelles, en indiquant leur signification opérationnelle.

Tableau 3-1 : Variables comportementales hebdomadaires retenues pour la FDA

Variable	Description
nb_connect	Nombre de connexions (<i>logon</i>) effectuées durant la semaine.
nb_disconnect	Nombre de déconnexions (<i>logoff</i>) durant la semaine.
total_events_device	Nombre total d'événements liés aux périphériques (ex. USB).
nb_pc_device	Nombre distinct de périphériques utilisés par l'utilisateur.
total_events_file	Nombre total d'événements d'accès ou modification de fichiers.
total_events_logon	Nombre total d'événements d'authentification.
nb_pc_file	Nombre distinct de fichiers ouverts/consultés durant la semaine.
nb_logon	Nombre de connexions réussies.
nb_logoff	Nombre de déconnexions réussies.
nb_attachments	Nombre de pièces jointes envoyées par courriel.
total_emails	Nombre total de courriels envoyés durant la semaine.
http_total_events	Volume total de requêtes HTTP (activité Web).

Dans une perspective d'analyse fonctionnelle, ces séries hebdomadaires sont interprétées comme des trajectoires temporelles. Pour chaque variable comportementale et pour chaque utilisateur, on construit une série ordonnée sur l'ensemble des semaines observées, puis on applique un lissage à l'aide d'une base de B-splines afin de reconstruire une fonction continue $X_i(t)$ sur l'axe des semaines. Ce

lissage permet de réduire le bruit inhérent aux logs tout en préservant les tendances structurelles, ce qui constitue un prérequis pour la FDA (Ramsay & Silverman, 2005 ; Ferraty & Vieu, 2006). Chaque utilisateur est ainsi représenté par un ensemble de courbes (une courbe par variable) décrivant l'évolution de son activité au fil des semaines.

Toutes les courbes sont construites sur la même grille temporelle hebdomadaire, ce qui assure une base temporelle commune pour la comparaison des trajectoires. Il ne s'agit pas d'un time warping complexe, mais d'un alignement structurel sur des périodes homogènes, cohérent avec la granularité hebdomadaire choisie.

2.3 Réduction dimensionnelle par FPCA

Une fois les trajectoires reconstruites, une Functional Principal Component Analysis (FPCA) est appliquée pour réduire la dimensionnalité de l'information et extraire les principaux modes de variation des comportements. Pour chaque variable comportementale, la FPCA identifie quelques composantes principales qui résument la tendance moyenne des utilisateurs, les modes de variation les plus fréquents et, le cas échéant, la présence de trajectoires atypiques (Gertheiss, Goldsmith & Greven, 2024 ; Dai & Müller, 2018).

Dans le contexte de CERT, cette étape permet :

- D'identifier les patterns dominants dans les routines simulées ;
- De comparer la variabilité des comportements entre utilisateurs malveillants et bénins ;
- De repérer les trajectoires associées à des séquences d'actions suspectes ;
- D'observer si les comportements malveillants se traduisent par des dérives progressives ou par des ruptures abruptes.

Les scores issus de la FPCA pour chaque utilisateur (par exemple `nb_connect_fpca1`, `nb_connect_fpca2`, etc.) sont ensuite utilisés comme variables explicatives dans les

modèles de classification. Ils offrent une représentation compacte mais riche de la dynamique temporelle des comportements.

2.4 Modèles de classification basés sur les composantes fonctionnelles

Une fois les trajectoires fonctionnelles reconstruites et résumées par la Functional Principal Component Analysis (FPCA), les scores fonctionnels obtenus servent de variables explicatives dans un ensemble de modèles supervisés destinés à évaluer la capacité du jeu CERT à permettre la détection des comportements malveillants. Cette section présente l'intégration de ces composantes dans deux familles de méthodes complémentaires : la régression logistique et les forêts aléatoires. Ces modèles ont été retenus en raison de leur large utilisation dans la littérature sur la détection d'anomalies et d'insider threats, mais aussi pour leur complémentarité en termes de capacité de modélisation, d'interprétabilité et de robustesse.

2.4.1 Régression logistique

La régression logistique constitue l'un des modèles supervisés les plus anciens et les plus employés dans le domaine de la cybersécurité, notamment pour la détection des menaces internes malveillantes (Chandola et al., 2009). Son principe repose sur l'estimation de la probabilité qu'un événement appartienne à une classe donnée, par exemple, une activité malveillante ou une activité normale, à partir d'un ensemble de variables explicatives. La fonction logistique ramène toute combinaison de variables à une valeur comprise entre 0 et 1, correspondant à la probabilité d'appartenance à la classe cible.

L'un des principaux avantages de cette méthode réside dans son caractère paramétrique : les coefficients associés aux variables explicatives fournissent une interprétation directe de leur contribution au risque global. Chaque paramètre peut être lu comme un odds ratio, indiquant comment une variation de la variable affecte les chances d'appartenir à la classe « malveillante ». Par exemple, dans l'expérimentation réalisée sur CERT, le modèle a révélé qu'un accroissement du nombre de fichiers

sensibles consultés multipliait par environ 1,35 le risque de comportement malveillant, toutes choses égales par ailleurs. De même, les connexions effectuées en dehors des heures ouvrables étaient fortement associées à une hausse du risque, confirmant l'importance de ce signal. Cette capacité à expliquer clairement les relations entre comportements et probabilité de menace constitue un atout majeur dans un contexte organisationnel où les analystes doivent justifier leurs décisions auprès des gestionnaires ou des instances juridiques (Harilal et al., 2017).

La régression logistique présente d'autres qualités notables. Elle s'adapte relativement bien aux situations de déséquilibre des classes, fréquentes dans CERT où les activités malveillantes sont minoritaires, grâce à l'introduction de pondérations différentielles ou à l'usage de techniques comme le sur-échantillonnage (SMOTE). Elle est également peu coûteuse en ressources computationnelles et simple à implémenter, ce qui facilite sa reproductibilité et les comparaisons entre études.

Ses limites sont cependant importantes. La régression logistique repose sur l'hypothèse de linéarité entre les variables explicatives et la transformation logit de la probabilité de l'événement cible, ce qui peut s'avérer réducteur dans un contexte où les comportements suspects émergent souvent de combinaisons non linéaires de signaux. Par exemple, un accès inhabituel peut sembler bénin isolément, mais devient suspect lorsqu'il est associé à un transfert massif de fichiers et à une navigation simultanée vers un service de stockage externe. Le modèle logistique peine à capturer ce type d'interactions complexes. De plus, il analyse chaque observation comme un événement indépendant, ce qui limite sa performance face aux menaces progressives dites « low and slow », caractérisées par une succession d'actions discrètes étalées dans le temps. Enfin, sa sensibilité à la qualité des données peut entraîner une chute rapide de performance en présence de variables bruitées ou manquantes, là où des méthodes plus robustes comme les forêts aléatoires tolèrent mieux ces imperfections.

Des ajustements permettent néanmoins d'enrichir l'usage de la régression logistique dans ce domaine. L'intégration de variables d'interaction améliore la capture des comportements complexes ; l'application de méthodes de régularisation (Lasso, Ridge) aide à réduire la multi-colinéarité et à renforcer la généralisation ; et l'usage de modèles logistiques hiérarchiques permet de mieux refléter la diversité des comportements selon

les départements ou types d'employés. Dans une perspective plus large, la régression logistique peut également servir de composant explicatif dans des approches hybrides combinant apprentissage supervisé et règles organisationnelles.

En somme, la régression logistique offre un compromis intéressant entre simplicité, transparence et efficacité, qui en fait un modèle de référence dans un travail comparatif comme celui mené sur CERT. Ses limites justifient toutefois sa confrontation à d'autres méthodes plus flexibles, telles que les forêts aléatoires afin de mieux cerner les apports et les limites de CERT dans l'évaluation des approches de détection des menaces internes malveillantes.

2.4.2 Forêts aléatoires (Random Forests)

Les forêts aléatoires, introduites par Breiman (2001), constituent l'une des méthodes d'ensemble les plus influentes et les plus largement utilisées dans le domaine de l'apprentissage supervisé. Leur principe repose sur la combinaison de multiples arbres de décision construits à partir de sous-échantillons aléatoires des données et de sous-ensembles de variables sélectionnés de manière aléatoire à chaque division. L'agrégation des prédictions issues de ces arbres, généralement par vote majoritaire pour les tâches de classification, permet d'obtenir un modèle global plus robuste et performant que les arbres individuels. Cette approche capitalise sur la diversité introduite par l'aléa pour réduire la variance et améliorer la capacité de généralisation, tout en limitant le risque de surapprentissage.

L'un des atouts majeurs des forêts aléatoires réside dans leur capacité à modéliser des relations non linéaires et à capturer des interactions complexes entre les variables. Contrairement à la régression logistique, qui repose sur des hypothèses de linéarité, les forêts aléatoires n'imposent aucune contrainte sur la forme de la relation entre variables explicatives et variable cible. Cela en fait un outil particulièrement adapté aux environnements hétérogènes et dynamiques comme ceux simulés dans le jeu de données CERT. Par exemple, un utilisateur qui se connecte à des heures inhabituelles peut ne pas représenter un risque isolé ; toutefois, si ce comportement est combiné à un transfert important de fichiers sensibles et à l'utilisation d'un périphérique externe, la

probabilité de menace augmente considérablement. Les forêts aléatoires sont capables de détecter ces combinaisons implicites sans qu'elles soient spécifiquement modélisées a priori.

Dans la détection des menaces internes malveillantes, plusieurs travaux ont démontré l'efficacité de cette approche. Legg et al. (2017) montrent que les forêts aléatoires permettent d'identifier des comportements déviants tels que des accès répétés à des ressources critiques en dehors des horaires habituels, ou des transferts massifs de données vers des destinations inhabituelles. Leur capacité à traiter un grand nombre de variables simultanément facilite également l'intégration de sources de données variées, comme les journaux de connexion, les courriels ou encore les historiques d'accès aux fichiers. Dans le contexte du jeu de données CERT, où les données sont multi-sources et hétérogènes, cette robustesse constitue un avantage déterminant.

Un autre aspect intéressant des forêts aléatoires tient à leurs propriétés d'explicabilité intermédiaire. Bien qu'il soit difficile d'interpréter la logique globale d'une forêt composée de centaines d'arbres, les mesures d'importance des variables offrent une visibilité sur les signaux les plus discriminants pour la classification. Ces indicateurs permettent, par exemple, de mettre en évidence que la fréquence d'accès à certains dossiers sensibles, le volume des courriels envoyés à l'extérieur de l'organisation ou encore les variations inhabituelles dans les privilèges d'un compte sont parmi les variables les plus contributives à la détection. Cette caractéristique renforce la valeur opérationnelle du modèle, car elle fournit aux analystes de sécurité des pistes concrètes pour prioriser leurs investigations et orienter leurs décisions.

Les forêts aléatoires se distinguent également par leur robustesse face au bruit et au surapprentissage. Chaque arbre de décision peut être sensible aux fluctuations locales des données, mais l'agrégation de centaines d'arbres atténue ces effets et permet de produire des prédictions plus stables. Cela est particulièrement pertinent dans des environnements où les données peuvent contenir des anomalies légitimes ou des comportements atypiques mais non malveillants, comme dans le jeu de données CERT. De plus, cette approche tolère relativement bien la présence de données manquantes ou de variables bruitées, ce qui constitue un atout dans les contextes organisationnels réels où la qualité des logs n'est pas toujours garantie.

Cependant, malgré ces avantages, certaines limites doivent être reconnues. La principale concerne le manque de transparence globale du modèle. Bien que chaque arbre soit simple et facilement interprétable, l'ensemble constitué de centaines d'arbres forme un système complexe, souvent qualifié de « boîte grise ». Cela pose un défi dans des environnements sensibles, comme les institutions financières ou gouvernementales, où la justification claire d'une alerte est cruciale. Une alerte générée par une forêt aléatoire peut être difficile à expliquer de manière intelligible aux décideurs non spécialistes.

Un autre défi réside dans la gestion des déséquilibres de classes, fréquents dans le cas des menaces internes malveillantes où les comportements malveillants sont largement minoritaires par rapport aux activités normales. Les forêts aléatoires peuvent avoir tendance à privilégier la classe majoritaire, ce qui entraîne une baisse de performance dans la détection des cas rares mais critiques. Des ajustements sont alors nécessaires, tels que le rééchantillonnage des données (oversampling de la classe minoritaire ou undersampling de la classe majoritaire), l'utilisation d'ensembles équilibrés ou encore l'application de pondérations spécifiques dans la fonction de coût.

En résumé, les forêts aléatoires représentent une méthode de détection puissante et flexible, capable de capturer des relations complexes et de traiter efficacement des données hétérogènes et bruitées. Leur capacité à fournir des mesures d'importance des variables en fait un outil particulièrement utile pour les analystes, à mi-chemin entre performance prédictive et explicabilité. Toutefois, leur opacité relative et leur sensibilité aux déséquilibres de classes rappellent qu'elles ne constituent pas une solution universelle. Dans le cadre de ce mémoire, leur utilisation sur le jeu de données CERT permettra d'examiner à quel point cette méthode concilie précision, robustesse et interprétabilité, et de comparer ses performances avec celles d'approches plus linéaires.

2.5 Critères d'évaluation et interprétation des résultats

L'évaluation des méthodes de détection vise à mesurer leur capacité à identifier correctement les utilisateurs malveillants simulés dans CERT, à partir de leurs

trajectoires comportementales hebdomadaires. Les indicateurs utilisés sont dérivés de la matrice de confusion :

- La précision (precision), qui mesure la proportion d'utilisateurs effectivement malveillants parmi ceux prédits comme tels. Elle se définit par :

Précision = $TP / (FP + TP)$ où TP représente les vrais positifs (malveillants correctement identifiés) et FP les faux positifs.

- Le rappel (recall), qui indique la proportion des malveillants correctement détectés par le modèle :

Rappel = $TP / (FN + TP)$ où FN correspond aux faux négatifs (malveillants non détectés).

- Le score F1 constitue une moyenne harmonique de la précision et du rappel, offrant une mesure synthétique équilibrée particulièrement pertinente dans un contexte de classes déséquilibrées (Sokolova & Lapalme, 2009) :
- Score F1 = $2 \times ((\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel}))$

En complément, l'aire sous la courbe ROC (AUC) est utilisée pour évaluer la capacité discriminante globale des modèles, indépendamment du seuil de classification choisi (He & Garcia, 2009). Ces métriques sont calculées pour chaque pli de validation croisée, puis moyennées pour obtenir des estimations stables des performances de la régression logistique et de la forêt aléatoire.

Les résultats sont ensuite interprétés à la lumière des connaissances théoriques issues de la littérature et des propos recueillis auprès des experts. Cette triangulation permet d'évaluer dans quelle mesure CERT représente fidèlement la variabilité des comportements d'utilisateurs, les dérives progressives observées dans des cas de menaces internes malveillantes, les ruptures comportementales associées à des activités réellement suspectes, ainsi que la diversité des profils d'employés en contexte organisationnel. Elle permet également de déterminer si le jeu CERT constitue un banc d'essai pertinent pour développer et comparer des méthodes de détection basées sur l'analyse fonctionnelle des données, ou si certaines limites structurelles doivent être prises en compte dans l'interprétation des résultats.

3. Entrevues auprès d'experts

La collecte de données qualitatives pour cette recherche s'est déroulée en plusieurs étapes successives visant à obtenir une compréhension approfondie des pratiques, perceptions et enjeux liés à la détection des menaces internes malveillantes. L'objectif de ces entrevues est de permettre une analyse comparative entre la littérature scientifique proposée et les propos des participants.

Pour assurer la confidentialité et le respect des exigences éthiques encadrant ce projet, la collecte de données qualitatives comprend la sélection des participants, leur recrutement, puis la réalisation d'entrevues semi-dirigées menées de manière individuelle. Plusieurs facteurs ont poussé la sélection de l'entrevue semi-dirigée pour la collecte de données. Premièrement, l'entrevue semi-dirigée se déroule majoritairement comme une conversation entre le chercheur et les répondants, conversation durant laquelle le chercheur peut plus facilement s'adapter et adapter ses questions au rythme de la conversation. Deuxièmement, la flexibilité que l'entrevue semi-dirigée apporte permet d'obtenir une compréhension plus riche du phénomène étudié (Kumar, 2011). Troisièmement, tous ces éléments permettent de stimuler la discussion encore plus comparativement à une entrevue structurée, mais permettent quand même de délimiter les frontières de l'entrevue contrairement à une entrevue non dirigée. Ces frontières sont délimitées à l'aide des questions ouvertes contenues dans le guide d'entrevue préparé à l'avance.

3.1 Sélection des participants

La population cible de cette étude est constituée de professionnels évoluant dans le domaine de la cybersécurité et ayant une exposition directe ou indirecte aux activités de surveillance, d'analyse ou de réponse aux comportements d'utilisateurs potentiellement malveillants. Ce champ d'expertise est largement représenté dans l'industrie montréalaise, qui compte plusieurs milliers de professionnels occupant des rôles liés à la gestion des menaces, à la réponse aux incidents et à la détection d'activités anormales.

Afin de constituer un échantillon pertinent et capable d'offrir un éclairage riche et varié, un échantillonnage intentionnel a été privilégié. Cette stratégie permet de sélectionner des participants sur la base de leur expertise reconnue et de leur capacité à contribuer de manière significative à l'analyse qualitative. Des critères d'inclusion et d'exclusion ont été définis pour assurer la pertinence des perspectives recueillies :

- Être âgés de 18 ans ou plus

Un critère essentiel dans toute démarche de recherche impliquant des êtres humains, afin de garantir leur capacité juridique à consentir librement et à comprendre les implications éthiques et méthodologiques de leur participation.

- Parler français ou anglais

La maîtrise de l'une de ces langues assurait une communication fluide, une compréhension précise des questions et la possibilité pour les experts d'exprimer de manière nuancée leurs expériences professionnelles.

- Occuper un poste en cybersécurité ayant un lien avec la détection ou l'analyse des comportements anormaux d'utilisateurs.

Ce critère garantissait que les participants possédaient une connaissance directe des processus organisationnels, des outils techniques et des limites inhérentes aux mécanismes de détection des menaces internes malveillantes. Cette expertise opérationnelle était indispensable pour obtenir des commentaires éclairés et applicables au cadre de recherche.

- Avoir une expérience professionnelle minimale d'un an dans le domaine des menaces internes ou dans des activités connexes comme la surveillance, la gestion d'accès ou l'analyse comportementale

Une expérience minimale garantissait la capacité du répondant à mobiliser des exemples concrets et des représentations informées. Les individus possédant une expertise insuffisante ou étant soumis à des contraintes professionnelles ou contractuelles limitant leur capacité à partager certaines informations étaient exclus de l'étude.

L'objectif de ces critères était de constituer un échantillon pertinent et crédible pour analyser de manière nuancée les limites méthodologiques des jeux de données simulés, tels que le jeu CERT, et les attentes des praticiens face à de tels outils.

3.2 Identification et recrutement des participants

Les participants potentiels ont été identifiés principalement selon une approche de convenance, en mobilisant le réseau professionnel de la ainsi que celui des directeurs du projet. Ce premier bassin de contacts a permis d'atteindre divers profils : analystes opérationnels, experts en modélisation des menaces, gestionnaires d'équipes SOC, consultants en sécurité ou spécialistes en enquêtes numériques. À cette première vague de recrutement s'est ajoutée une stratégie de type boule de neige, permettant aux participants déjà rencontrés de suggérer d'autres professionnels susceptibles d'apporter une contribution pertinente au projet. Cette méthode a facilité l'accès à des profils variés, provenant d'organisations publiques et privées, et exerçant dans des environnements techniques ou de gouvernance différente.

Les individus identifiés ont été sollicités directement par la par courriel. Le message d'invitation présentait l'objectif du mémoire, les thèmes abordés lors de l'entrevue et les modalités de déroulement, ainsi que les garanties de confidentialité accordées aux participants. La lettre d'invitation officielle envoyée à tous les participants précisait notamment que ni leur nom, ni celui de leur organisation ne seraient divulgués, et que l'entrevue se déroulerait en personne, dans un environnement calme et confidentiel, pour une durée de 30 à 45 minutes. La participation était entièrement volontaire et chaque répondant conservait la possibilité de se retirer à tout moment.

L'objectif de la démarche n'était pas d'atteindre un nombre prédéfini de participants, mais plutôt de réunir un groupe d'experts possédant des compétences complémentaires et une expérience suffisamment diversifiée pour éclairer les différentes dimensions des menaces internes malveillantes. Au fil du recrutement, ce sont finalement six professionnels qui ont accepté de participer, chacun avec une

expertise distincte. Cette diversité de profils a permis d'obtenir un éventail riche de perspectives, malgré la taille limitée de l'échantillon.

De plus, le domaine de la cybersécurité présente des contraintes importantes en matière de recrutement : la sensibilité des informations, les obligations de confidentialité et la réticence à discuter de cas réels ou simulés réduisent fortement le bassin de spécialistes disposés à participer à une étude académique. Ces facteurs ont renforcé la pertinence de cibler un petit nombre de participants qualifiés plutôt que de viser un échantillon plus large difficilement accessible. Afin d'assurer la crédibilité et la profondeur des propos recueillis, les critères de la section 3.1 ont été respectés, permettant de garantir que les participants disposaient du recul et de l'expertise nécessaires pour contribuer de manière substantielle à l'analyse.

Les entrevues se sont déroulées entre octobre et novembre 2025, période approuvée par le Comité d'éthique de la recherche (CER) de HEC Montréal.

3.3 Quelques détails sur le protocole d'entrevue

Les entrevues ont été réalisées individuellement selon une approche semi-dirigée, permettant à la fois d'orienter la discussion autour de thèmes préétablis et de laisser aux participants la liberté d'exprimer leurs réflexions et expériences personnelles. Une seule version du protocole d'entrevue (Annexe 1) a été élaborée pour tous les participants, peu importe leur poste au sein d'une entreprise. Cette décision a été prise, car il a été jugé que tout participant répondant aux critères de sélection pour l'étude serait en mesure de répondre adéquatement aux questions posées. Plus spécifiquement, le protocole d'entrevue a été conçu pour guider progressivement la discussion du participant, en commençant par son parcours professionnel et son exposition aux incidents internes, puis en explorant sa compréhension des comportements malveillants, et enfin en l'amenant à analyser en profondeur divers scénarios représentatifs. La structure générale du protocole s'inspire à la fois des recommandations méthodologiques présentes dans la littérature sur les entrevues semi-dirigées et des approches adoptées dans des travaux antérieurs réalisés à HEC Montréal. La construction du protocole a également été influencée par l'état actuel de la recherche sur les menaces internes

malveillantes, ainsi que par les limites reconnues des jeux de données synthétiques comme le CERT r4.2, qui constituent un enjeu central de cette étude.

3.4 Quelques détails sur l'entrevue et sa structure

Bien que l'utilisation de la structure semi-dirigée pour les entrevues présente plusieurs forces, soit entre autres la possibilité d'expliquer les questions aux participants si un élément n'est pas bien compris ou pas bien énoncé initialement, cette structure comporte également des faiblesses dont il faut tenir compte. Premièrement, des détails de l'entrevue peuvent être oubliés si la prise de note ou la retranscription des entrevues n'est pas réalisée de manière rigoureuse (Paré, 2004). D'autre part, la qualité des données dépend de la qualité des réponses obtenues. Pour atténuer ces limites, le protocole prévoit qu'une synthèse soit réalisée après chaque entrevue à l'aide d'une grille d'analyse individuelle, et que la pose des questions de relance lorsque nécessaire afin d'approfondir les propos des participants.

Les entrevues se déroulent en mode présentiel lorsque possible, ou à distance par visioconférence, afin de faciliter la coordination des rencontres avec des experts provenant de milieux professionnels variés. Dans les deux formats, un partage d'écran est utilisé afin de présenter les scénarios, les questions clés et l'aide-mémoire, assurant ainsi une compréhension commune du contenu discuté. Ce support visuel sert également à structurer l'entretien tout en maintenant un échange dynamique et ouvert aux digressions pertinentes.

Dans une perspective méthodologique, les participants ne reçoivent pas les questions à l'avance. Ce choix vise à favoriser l'émergence de réponses spontanées, authentiques et non influencées par une préparation préalable, ce qui contribue à renforcer la validité des données qualitatives et à limiter le biais de désirabilité sociale.

Le guide utilisé pour structurer ces échanges se compose de cinq sections complémentaires. La première vise à accueillir le participant, à le remercier pour sa contribution et à rappeler les objectifs de l'étude. C'est également le moment de préciser la durée prévue de l'entretien, de rappeler les modalités de confidentialité ainsi

que l'enregistrement audio. En plus, elle permet de répondre à tous les questionnements du participant au sujet de l'étude avant de commencer l'entrevue.

La deuxième section porte sur le profil professionnel du répondant. Elle a pour objectif de situer son expertise : son rôle actuel en cybersécurité, son expérience dans la détection ou l'analyse d'incidents, ainsi que son exposition préalable à des cas de menaces internes malveillantes. Cette contextualisation est essentielle pour interpréter nuancément ses perceptions et analyser la profondeur de son jugement professionnel.

La troisième section explore les menaces internes malveillantes rencontrées en pratique. Les participants sont invités à décrire les comportements qu'ils considèrent comme révélateurs de risques, les signaux faibles qu'ils observent le plus souvent, ainsi que les types de données qu'ils jugent essentiels pour soutenir une détection efficace. Cette partie permet de brosser un portrait riche de la compréhension experte du phénomène.

La quatrième section repose sur des scénarios élaborés spécifiquement pour l'étude, chacun représentant une forme distincte de malveillance interne. Dès cette étape, l'écran est partagé afin de présenter la mise en situation sélectionnée aléatoirement, suivie des différentes questions d'analyse. Bien que les participants disposent d'un aide-mémoire, cette présentation visuelle demeure essentielle pour stimuler la réflexion, permettre une interprétation détaillée des événements et encourager l'identification de signaux techniques, comportementaux et organisationnels associés au scénario. L'usage de situations narratives et contextualisées vise à reproduire la complexité des incidents réels, de manière à mobiliser le jugement professionnel du répondant plutôt que de susciter des réponses théoriques ou générales.

Pour chaque participant, deux mises en situation sont sélectionnées aléatoirement parmi les 3 mises en situation suivantes :

- Scénario 1 : Vol de données stratégiques par un employé insatisfait cherchant à exfiltrer progressivement des informations sensibles
- Scénario 2 : Sabotage des systèmes par un administrateur système profitant de ses privilèges pour modifier des configurations critiques

- Scénario 3 : Fraude financière par exploitation d'un défaut de séparation des tâches.

Ces mises en situation, présentées sous forme de diapositives (Powerpoint), sont accompagnées de séries de questions ouvertes portant notamment sur la plausibilité du scénario, les actions jugées typiques d'un comportement malveillant, les signaux techniques et comportementaux associés, et les types de données qui devraient être incluses dans un jeu de données simulé pour représenter adéquatement ce type d'incident. Cette étape vise également à documenter les écarts perçus entre les incidents réels et la manière dont ils sont simplifiés dans les jeux de données synthétiques existants, notamment le jeu CERT.

Les versions complètes des scénarios, telles que présentées aux participants, sont reproduites en Annexe 2.

Afin de garder le participant engagé tout au long de l'entrevue et favoriser la spontanéité des réponses, les questions sont présentées avant l'affichage du scénario, ce qui implique des allers-retours entre les diapositives.

Enfin, la dernière section du protocole d'entrevue aborde plusieurs dimensions transversales, dont les comportements perçus comme indicateurs de risque interne, les types de données jugés essentiels pour une détection efficace, les limites attribuées aux jeux de données synthétiques utilisés dans la recherche, ainsi que les attentes des experts envers des approches de détection plus contextualisées et représentatives des environnements réels. Cette structure souple offre aux participants l'espace nécessaire pour illustrer leurs propos à l'aide d'exemples concrets tirés de leur pratique, tout en leur permettant de commenter de manière nuancée les scénarios artificiels proposés par le jeu CERT.

Finalement, chaque entrevue se conclut par une phase d'ouverture invitant les participants à partager toute réflexion additionnelle ou recommandation concernant l'amélioration des jeux de données simulés ou des outils actuels de détection. Cette étape vise à faire émerger des éléments souvent absents des réponses directes, mais essentiels pour enrichir l'analyse, notamment en ce qui concerne les enjeux

organisationnels, les limites structurelles des outils ou les pratiques émergentes pouvant influencer la détection des menaces internes malveillantes dans les prochaines années.

Le participant est ensuite remercié, et la confirme son accord pour être recontacté au besoin. Il est également informé qu'il peut, s'il le souhaite, suggérer d'autres experts susceptibles de contribuer à l'étude.

Toutes les entrevues sont enregistrées avec le consentement explicite des participants, puis intégralement transcrites afin de permettre une analyse rigoureuse et systématique dans les étapes subséquentes. Aucune information permettant d'identifier les répondants n'est conservée dans les transcriptions, chaque participant étant désigné par un identifiant neutre au moment du traitement et de l'analyse

4. Analyse des données

Cette section présente en détail les étapes suivies pour analyser les données qualitatives recueillies dans le cadre de cette recherche portant sur la menace interne malveillante. Les données à analyser proviennent de six entrevues semi-dirigées menées auprès d'experts en cybersécurité, en détection d'intrusion, en gouvernance de la sécurité et en recherche universitaire. L'objectif principal de cette analyse est de mieux comprendre les types de comportements jugés préoccupants, les données réellement utilisées par les professionnels en cybersécurité, ainsi que les limites et défis associés à l'utilisation de jeux de données simulés.

L'unité d'analyse retenue dans cette étude correspond à la représentation experte de la menace interne malveillante, c'est-à-dire la manière dont les professionnels conceptualisent les comportements à risque, les signaux faibles, les trajectoires malveillantes et les difficultés liées à leur détection. Même si les entrevues sont réalisées individuellement, l'analyse cherche à dégager des régularités, des convergences et des divergences dans les perceptions collectives afin d'en tirer des recommandations pour la construction d'un jeu de données plus réaliste.

L'ensemble du processus d'analyse est guidé par une approche thématique inspirée de Braun & Clarke (2006), qui constitue une référence majeure pour les analyses qualitatives. Cette approche permet de structurer et d'interpréter les idées récurrentes tout en laissant émerger de nouveaux thèmes issus du discours des participants. Elle s'inscrit également dans la lignée des recommandations méthodologiques de Guest, MacQueen & Namey (2012) ainsi que de Miles, Huberman & Saldaña (2014), qui insistent sur la rigueur, la transparence et la systématique dans le traitement des données qualitatives. Les étapes du processus sont détaillées ci-dessous

4.1 Transcription et préparation du corpus

La première étape de l'analyse consiste à transcrire intégralement chacune des entrevues enregistrées. Cette transcription vise à préserver la richesse des propos, à assurer une uniformité dans le traitement des données et à permettre le repérage de nuances, d'exemples et d'explications contextuelles que la prise de notes seule ne permet pas de saisir.

Dans le domaine de la cybersécurité, où les raisonnements des experts mobilisent souvent des connaissances techniques complexes, des expériences situées et des interprétations fines des comportements, la fidélité des propos est essentielle pour garantir la validité de l'analyse. Cette assertion est aussi reconnue par Nowell et al. (2017), la fidélité du corpus est essentielle pour garantir la crédibilité et la validité d'une analyse thématique.

Les notes d'observation recueillies pendant les entrevues sont ensuite ajoutées aux transcriptions afin de fournir un complément contextuel utile à l'interprétation. Les entrevues et les notes d'observation sont ainsi regroupées et préparées pour être analysées sous forme de thèmes communs plutôt que par outil de collecte.

4.2 Codifier les données collectées

Une fois les transcriptions complétées, une codification systématique est effectuée. Le processus suit une démarche en deux temps. Une première grille de codification (voir le

tableau 3.2 – Grille de codification thématique ci-dessous) est élaborée à partir de la littérature sur la menace interne malveillante, notamment les travaux portant sur les comportements malveillants, les signaux faibles, les profils à risque et les limites des mécanismes de détection. Cette grille déductive assure une cohérence avec les cadres conceptuels existants.

Tableau 3-2 : Grille de codification thématique

Thème	Catégories	Codes	Description analytique
T1-Profil et positionnement des experts	T1.1 Expérience	EXP	Années d'expérience, ancienneté
	T1.2 Rôle dans l'incident	ROLE	Prévention, détection, réponse, gouvernance
	T1.3 Orientation	ORIENT	Technique / organisationnelle / humaine
	T1.4 Ambiguïté comportementale	AMBIG	Ambiguïté entre erreur et malveillance
T2-Typologie des comportements malveillants	T2.1 Exfiltration	EXFIL	Extraction de données sensibles
	T2.2 Abus de privilèges	PRIV	Accès non autorisés, élévation de droits
	T2.3 Identifiants	IDMG	Mauvaise gestion ou partage
	T2.4 Sabotage	SAB	Modifications malveillantes
	T2.5 Usage abusif des ressources	RESS	Détournement, usage personnel

	T2.6 Actions “grises”	GREY	Ambiguïté, comportements non malveillants
T3-Profiles d'utilisateurs à modéliser	T3.1 Employés réguliers	EMP	Accès sensibles non privilegiés
	T3.2 Administrateurs	ADMIN	Risque élevé, capacités de dissimulation
	T3.3 Gestionnaires	GEST	Accès stratégiques
	T3.4 Fonctions sensibles	FINRH	Finances, RH, recherche
	T3.5 Nouveaux employés	NEW	Stagiaires, contractuels
T4-Signaux observables et corrélations	T4.1 Signaux temporels	TIME	Horaires atypiques, fréquence
	T4.2 Signaux d'accès	ACCESS	Accès inhabituels, répétés
	T4.3 Périphériques	USB	Clés USB, stockage externe
	T4.4 Habilitations	HABIL	Changement de privileges
	T4.5 Communications	COMM	Emails internes, transferts
	T4.6 Corrélations	CORR	Multi-signaux, patterns séquentiels
T5-Outils de détection et limites	T5.1 Outils utilisés	TOOLS	SIEM, IDS, EDR, UEBA
	T5.2 Limites techniques	TECHL	Faux positifs, rigidité

	T5.3 Limites contextuelles	CONTEXT	Faible contextualisation métier
	T5.4 Limites humaines	HUM	Charge cognitive, expertise
T6-Enjeux des jeux de données simulés	T6.1 Simplification	SIMPLE	Scénarios trop stéréotypés
	T6.2 Temporalité	TEMPO	Absence de trajectoires longues
	T6.3 Faible contextualisation	CTX	Manque d'information métier
	T6.4 Absence de signaux faibles	SF	Manque de granularité
	T6.5 Déséquilibre	BAL	Répartition inégale des événements
	T6.6 Labellisation	LABEL	Labels trop simplistes

Un second cycle de codification, inductif cette fois, permet de faire émerger des thèmes nouveaux ou inattendus directement issus du discours des experts (Braun & Clarke, 2006 ; Miles et al., 2014). Des codes supplémentaires sont ajoutés lorsqu'ils reviennent de manière récurrente ou lorsqu'ils apportent un éclairage original sur le phénomène, par exemple, la nécessité de contextualiser les signaux selon le rôle professionnel ou les difficultés d'interprétation des comportements ambigus.

Cette étude adopte ainsi une approche hybride, combinant codage déductif (fondé sur les cadres théoriques existants) et inductif (émergent directement des propos des participants), ce qui permet à la fois de respecter les connaissances établies et de donner une place importante aux perceptions expertes.

4.3 Analyser les données (intra-répondants et inter-répondants)

À la suite de la codification des verbatims, une première analyse est réalisée individuellement pour chacun des répondants. Cette étape intra-répondant permet de faire émerger les thèmes les plus saillants à partir des codes attribués, mais également de comprendre la logique interne de leurs propos. L'objectif n'est pas de quantifier la fréquence d'apparition des thèmes, mais plutôt d'examiner le sens et le contexte des éléments soulevés : comment chaque expert interprète les scénarios présentés, quelles justifications sous-tendaient son raisonnement et comment ses expériences professionnelles influençaient sa perception des comportements potentiellement malveillants.

Dans un second temps, une analyse inter-répondants est effectuée afin de comparer les convergences et les divergences entre les participants. Cette comparaison permet d'identifier les points de consensus, les nuances propres à certains profils d'experts et les apports complémentaires permettant de bonifier l'interprétation. Ce croisement des perspectives contribue à vérifier la robustesse des thèmes dégagés lors de l'analyse individuelle et à déterminer lesquels pouvaient être considérés comme représentatifs de l'ensemble du corpus (Nowell et al., 2017).

Ainsi, cette double approche, d'abord centrée sur chaque répondant, puis élargie à l'ensemble des participants, permet d'obtenir une compréhension approfondie et cohérente des perceptions des experts, tout en assurant une interprétation fidèle et solidement ancrée dans les données recueillies.

4.4 Comparer l'analyse à la littérature

Enfin, une fois les propos des participants consolidés, ceux-ci sont confrontés à la littérature scientifique. Cette étape vise d'abord à situer les résultats obtenus par rapport aux connaissances existantes et à identifier, le cas échéant, des interprétations alternatives ou complémentaires à celles issues de l'analyse (Dhennin, 2011 ; Braun & Clarke, 2006). Cette mise en dialogue entre les données empiriques et les travaux théoriques a également permis de réviser et de renforcer les thèmes dégagés, en

s'appuyant à la fois sur les apports de la recherche et sur l'expérience pratique des experts interrogés. Elle a de plus fourni un éclairage critique sur la pertinence et les limites des jeux de données simulés utilisés dans les travaux sur la menace interne malveillante, notamment ceux de la famille CERT r4.2, souvent mobilisés en recherche mais considérés comme imparfaits par plusieurs experts.

Au terme de cette comparaison, il est possible de formuler des propositions cohérentes et fondées, appuyées à la fois par les résultats des entrevues et par les cadres conceptuels reconnus dans le domaine de la détection des menaces internes malveillantes.

4.5 Considérations éthiques

Tout au long de la collecte des données et de leur analyse, des précautions ont été prises par la chercheuse afin d'assurer le consentement des répondants, la confidentialité de leurs propos et renseignements personnels ainsi que pour minimiser les risques de leur participation à l'étude. Le Comité d'Éthique en Recherche de HEC Montréal a également approuvé l'étude avant que la collecte de données ne débute.

4.5.1 Consentement des répondants

Le consentement des répondants a été obtenu conformément aux exigences éthiques de la recherche. Avant leur participation, chaque expert a reçu un formulaire papier ou électronique présentant les objectifs de l'étude, les modalités de participation, les conditions de confidentialité ainsi que la demande d'enregistrement audio de l'entrevue. Ces éléments ont ensuite été rappelés verbalement au début de la rencontre afin de s'assurer que les participants comprenaient pleinement la nature de leur engagement.

Les répondants ont également été informés qu'ils pouvaient poser des questions en tout temps, refuser de répondre à certaines questions ou mettre fin à leur participation sans justification ni conséquence. Ce rappel visait à garantir un consentement libre, éclairé et continu, conformément aux lignes directrices éthiques.

4.5.2 Confidentialité des répondants et des propos rapportés

La confidentialité des propos rapportés a également été une dimension éthique surveillée tout au long de la collecte et l'analyse des données. Plusieurs mesures ont été mises en place afin de réduire au maximum les risques d'identification des participants. D'abord, bien qu'un anonymat absolu ne puisse être garanti en raison du mode de recrutement, fondé en partie sur le réseau professionnel de la directrice de recherche et sur des recommandations entre experts, tout a été mis en œuvre pour limiter la circulation d'informations permettant d'inférer l'identité des répondants. Par exemple, lorsque des participants suggéraient d'autres personnes à rencontrer, aucune information n'était communiquée quant à l'acceptation ou au refus de ces individus. Cette approche a permis de réduire le risque que des membres d'une même organisation découvrent involontairement la participation d'un collègue.

Ensuite, certains renseignements recueillis dans le cadre des questionnaires ou des entrevues auraient pu, s'ils avaient été conservés tels quels, permettre d'identifier les experts (p. ex., nom complet, organisation d'attache, intitulé de poste très spécifique). Ces informations ont été obtenues uniquement pour soutenir la compréhension contextuelle lors de l'analyse, mais elles n'ont pas été retenues dans les résultats ni utilisées pour tirer des conclusions.

Afin de préserver strictement la confidentialité, tous les éléments permettant d'identifier directement ou indirectement les participants ont été retirés ou reformulés. L'analyse s'est concentrée uniquement sur des caractéristiques professionnelles générales, telles que le champ d'expertise ou la nature des responsabilités, qui ne permettent pas de remonter à une personne en particulier. Cette démarche d'anonymisation rigoureuse visait à protéger les participants tout en préservant l'intégrité analytique des données.

4.5.3 Atténuation des risques

Tout au long du processus de collecte et d'analyse des données, plusieurs mesures ont été mises en place afin de réduire les risques d'enjeux éthiques et d'assurer un traitement responsable des informations recueillies.

Premièrement, les scénarios utilisés lors des entrevues étaient fictifs, bien que construits à partir de situations plausibles inspirées de la littérature sur la menace interne malveillante et des limites observées dans les jeux de données simulés tels que CERT r4.2. Ce choix méthodologique visait à limiter tout risque que les experts dévoilent, volontairement ou non, des éléments confidentiels appartenant à leur organisation. Si un participant abordait une situation sensible ou un exemple trop précis, il était invité à le signaler, et cette portion du verbatim était alors écartée de l'analyse afin de protéger la confidentialité organisationnelle.

Deuxièmement, même si certaines informations générales ont été recueillies pour mieux comprendre le profil des répondants (p. ex., domaine d'expertise, années d'expérience, secteur d'activité), aucune donnée n'a été conservée ou présentée sous une forme permettant d'identifier les participants de manière directe ou indirecte. Les réponses issues des entrevues ont été regroupées et analysées globalement, ce qui confère un anonymat partiel : les propos ne sont pas associés à un individu précis, mais un observateur externe connaissant intimement le milieu de la cybersécurité pourrait, dans de rares cas, inférer des éléments d'identification. Les participants ont été informés de cette limite dès la remise du formulaire de consentement et au début de l'entretien.

Troisièmement, aucune information identificatoire ou sensible pouvant provenir du contexte professionnel des experts (par exemple, le nom de leur entreprise, des incidents réels, des technologies propriétaires ou des politiques internes) n'a été transposée dans ce mémoire. Lorsque des détails techniques étaient nécessaires à la compréhension, ils ont été reformulés ou généralisés afin d'éviter toute reconnaissance organisationnelle. La chercheuse n'avait par ailleurs aucun lien d'emploi ou d'obligation professionnelle envers les participants, ce qui a permis de prévenir les conflits d'intérêts et de garantir une analyse indépendante des données partagées.

Ces mesures ont permis de réduire considérablement les risques de bris de confidentialité ou d'utilisation inappropriée d'informations sensibles, tout en assurant une analyse rigoureuse et éthiquement responsable des perceptions expertes recueillies dans le cadre de l'étude.

4.6 Conclusion

Ce chapitre a pour objectif de présenter et de justifier l'approche méthodologique adoptée pour analyser les perceptions d'experts en cybersécurité concernant la menace interne malveillante et les limites des jeux de données simulés, tels que le CERT r4.2. L'étude repose sur une démarche qualitative, soutenue par des entrevues semi-dirigées permettant d'explorer en profondeur l'expertise et les raisonnements des participants. Dans un premier temps, les données sont recueillies à l'aide de six entrevues semi-dirigées, précédées d'un court questionnaire descriptif visant à situer le profil général des experts rencontrés. L'ensemble des entrevues est ensuite transcrit intégralement afin d'assurer la fidélité des propos et de faciliter l'analyse des nuances, des exemples et des raisonnements mobilisés.

Par la suite, un processus de codification mixte, combinant une approche déductive inspirée de la littérature et une approche inductive issue directement des verbatims, est mis en œuvre pour structurer les données. Une analyse intra-répondant permet d'abord de dégager les thèmes saillants propres à chaque expert, tandis qu'une analyse inter-répondants permet ensuite d'identifier les convergences, divergences et complémentarités entre les participants.

Enfin, les résultats préliminaires sont mis en perspective avec les travaux existants afin de valider, nuancer ou enrichir les thèmes émergents issus des entrevues. Cette triangulation entre données empiriques et littérature contribuera à renforcer la solidité interprétative et la cohérence analytique de la démarche.

Le chapitre suivant présentera les résultats issus de cette analyse et les principaux enseignements tirés des entrevues réalisées auprès des experts.

Chapitre 4. Résultats et Analyse

Ce quatrième chapitre présente les résultats issus de l'ensemble du dispositif de collecte et d'analyse mis en œuvre dans le cadre de cette étude, soit l'application de la méthode d'analyse fonctionnelle des données (FDA) sur le jeu de données CERT r4.2 ainsi que les entrevues menées auprès d'experts en cybersécurité. L'analyse vise principalement à répondre aux deux objectifs empiriques au cœur de ce mémoire :

1. Caractériser la dynamique temporelle des comportements simulés dans le jeu de données CERT r4.2 au moyen de l'analyse fonctionnelle des données (FDA), et évaluer dans quelle mesure ces dynamiques produisent des signaux discriminants pour la détection des menaces internes malveillantes. ;
2. Mettre en perspective les signaux comportementaux mis en évidence par la FDA avec les observations d'experts en cybersécurité, afin d'identifier les convergences, les écarts et les limites du jeu CERT en termes de réalisme et de pertinence pour la recherche sur les menaces internes.

Le chapitre est structuré en deux grandes parties. La première présente les résultats quantitatifs issus de la méthode FDA, incluant la construction des trajectoires fonctionnelles, l'analyse en composantes principales fonctionnelles (FPCA) et la performance des modèles supervisés fondés sur les composantes fonctionnelles. La seconde partie expose les résultats qualitatifs provenant des entrevues avec les experts, qui permettent de contextualiser les tendances observées et d'en évaluer la pertinence dans des environnements organisationnels réels.

1. Approche quantitative

1.1 Description du corpus

L'analyse quantitative repose sur un corpus construit à partir du jeu de données CERT r4.2 après agrégation hebdomadaire des journaux. Le jeu final comporte 1 000 utilisateurs, correspondant à l'ensemble des employés simulés par le jeu de données

CERT au cours de la période observée. La structure temporelle du corpus couvre 73 semaines, allant de la semaine 2009-53 à la semaine 2011-20, soit un peu plus de seize mois d'activité continue. Cette profondeur temporelle est essentielle dans la perspective de la FDA, qui requiert un volume suffisamment long pour reconstruire des trajectoires comportementales lissées et comparer les profils évolutifs des utilisateurs.

La variable cible indiquant l'implication dans un scénario malveillant révèle un déséquilibre marqué des classes, caractéristique des jeux de données liés à la sécurité organisationnelle : 930 utilisateurs (93 %) sont classés bénins, tandis que 70 utilisateurs (7 %) sont identifiés comme malveillants. Ce ratio correspond à une prévalence réaliste des incidents internes dans les organisations, mais constitue un défi classique pour l'apprentissage supervisé, en raison du risque accru de faux négatifs et de biais en faveur de la classe majoritaire. Cette distribution justifie à la fois l'usage de métriques robustes (F1-score, AUC) et l'intégration de méthodes complémentaires dans la phase de classification.

Les analyses s'effectuent sur douze variables comportementales hebdomadaires, couvrant l'essentiel des activités enregistrées dans le jeu de données CERT : connexions et déconnexions, interactions avec les périphériques USB, accès aux fichiers, événements d'authentification, activité courriel (avec pièces jointes) et volume de navigation Web. Ces variables constituent la base des trajectoires fonctionnelles utilisées dans la FDA et reflètent un spectre large d'indicateurs comportementaux employés dans la littérature pour caractériser les signaux faibles liés aux menaces internes malveillantes.

L'examen de la couverture temporelle par utilisateur met en évidence une structure particulièrement hétérogène. Sur les 1 000 utilisateurs, certains ne sont présents que de manière partielle dans le jeu : le nombre de semaines observées varie fortement, allant de 5 semaines seulement à 73 semaines (la période complète). La majorité des employés sont toutefois observés presque entièrement : les premier, deuxième et troisième quartile se situent tous à 72 semaines, tandis que la moyenne atteint 67,24 semaines. Sur le plan proportionnel, cela signifie que la vaste majorité des utilisateurs

dispose d'une couverture supérieure à 98 % de la période étudiée, tandis qu'une minorité présente une présence fragmentaire (minimum : 6,8 % des semaines).

Cette structure comporte plusieurs implications méthodologiques. D'une part, la forte présence des employés sur la quasi-totalité de la période permet la reconstruction de trajectoires fonctionnelles complètes et comparables, un élément fondamental pour l'application de bases de B-splines et de la FPCA. D'autre part, la présence d'utilisateurs faiblement représentés introduit un risque de distorsion statistique, car les courbes construites sur cinq ou six points seulement comportent davantage d'incertitude. Dans le protocole analytique, ces utilisateurs sont conservés afin de ne pas biaiser la distribution globale, mais leur interprétation s'effectue avec prudence, notamment dans les analyses de variabilité interindividuelle.

En somme, le corpus analysé se caractérise par :

- Une cohorte riche de 1 000 utilisateurs,
- Une profondeur temporelle de 73 semaines permettant d'étudier la dynamique évolutive des comportements,
- Un déséquilibre de classes marqué mais propre aux scénarios de menace interne malveillante,
- Une couverture temporelle globalement élevée, favorable à la construction de trajectoires fonctionnelles fiables.

Ces propriétés garantissent la pertinence du jeu pour l'application de la FDA, tout en éclairant les limites structurelles qui seront prises en compte dans l'interprétation des résultats.

Analyse descriptive du jeu de données

L'analyse exploratoire des données constitue une étape essentielle pour comprendre la structure, la répartition et les dynamiques des activités des utilisateurs au sein du jeu CERT. Cette investigation permet de dégager des tendances générales, d'identifier des comportements atypiques et de poser les bases pour la construction de variables

dérivées ainsi que pour l'identification d'indicateurs prédictifs de comportements potentiellement malveillants.

L'examen de la table Device, qui enregistre l'usage des périphériques et les sessions utilisateur par poste de travail, révèle que 265 utilisateurs distincts se connectent à 971 ordinateurs différents, certains individus utilisant plusieurs postes. Le tableau 4.1 présente les statistiques descriptives des principales variables hebdomadaires dérivées de cette table, lesquelles servent à caractériser les volumes d'activité, la dispersion des comportements et les tendances centrales observées.

Tableau 4-1 : Statistiques descriptives des variables comportementales hebdomadaires de la table Device

Variables Device	Nombre total	Moyenne	Écart-type	Minimum	1er quartile (Q1)	Médiane	3e quartile (Q3)	Maximum
nb_connect	203 339	14.2	15.4	1	4	9	15	85
nb_disconnect	202 041	14.2	15.4	0	4	8	15	85
nb_pc_device	21 859	1.53	2.75	1	1	1	1	25
total_events_device	405 380	28.4	30.8	1	8	17	30	170

Dans l'ensemble, les variables (nb_connect) et (nb_disconnect) présentent des moyennes hebdomadaires similaires (14,2 événements), accompagnées d'une forte dispersion, comme en témoignent des écarts-types élevés et des valeurs maximales atteignant 85 événements. Les distributions associées, illustrées aux figures 4.1 et 4.2, sont fortement asymétriques à droite. La majorité des observations se concentre sur de faibles volumes d'activité, tandis qu'un nombre restreint de semaines affiche des niveaux de connexions ou de déconnexions nettement plus élevés. Cette structure suggère l'existence de comportements routiniers dominants, ponctués d'épisodes d'activité intensive susceptibles de correspondre à des situations particulières.

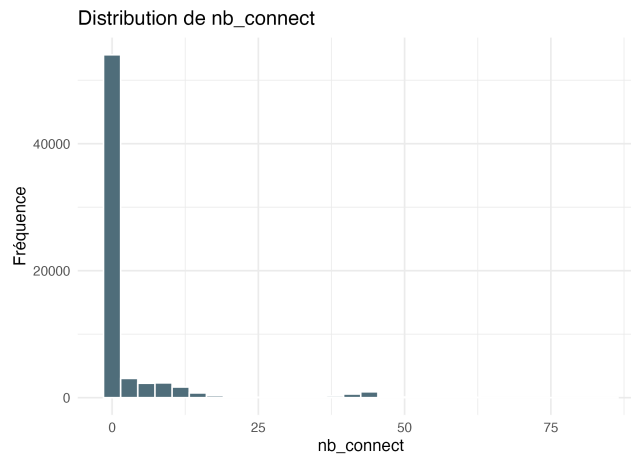


Figure 4-1 : Distribution hebdomadaire du nombre de connexions (nb_connect)

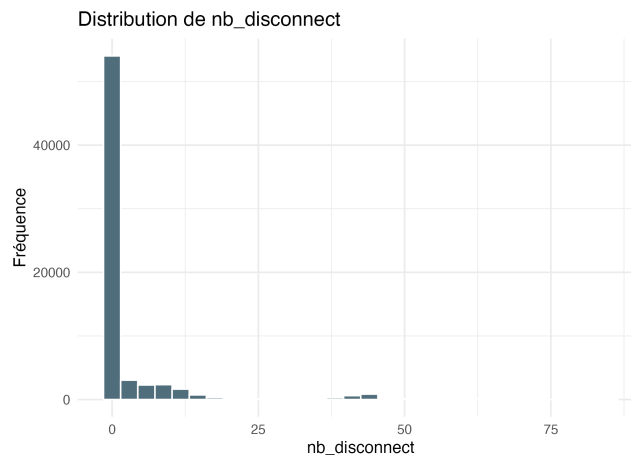


Figure 4-2 : Distribution hebdomadaire du nombre de déconnexions (nb_disconnect)

La variable (nb_pc_device), qui mesure le nombre de postes distincts utilisés par un utilisateur au cours d'une semaine, présente une moyenne faible (1,53) et une médiane égale à 1, indiquant que la majorité des utilisateurs opèrent habituellement sur un seul poste. La figure 4.3 confirme cette concentration marquée. Toutefois, la présence de valeurs extrêmes atteignant jusqu'à 25 postes révèle l'existence de profils atypiques, possiblement associés à des rôles spécifiques tels que des administrateurs ou des utilisateurs disposant de privilèges étendus. Ces comportements se distinguent

nettement de la norme et constituent des signaux potentiellement pertinents pour l'analyse des risques internes.

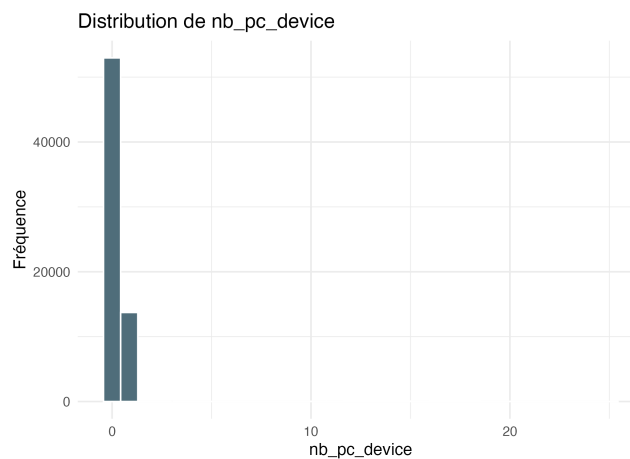


Figure 4-3 : Distribution hebdomadaire du nombre de postes utilisés (nb_pc_device)

La variable agrégée (total_events_device), qui synthétise l'ensemble des événements de connexion et de déconnexion hebdomadaires, présente une distribution fortement asymétrique à droite, comme l'illustre la figure 4.4. La médiane s'établit à 17 événements, tandis que le troisième quartile atteint 30, certaines semaines culminant à 170 événements. Cette asymétrie traduit la coexistence de comportements majoritairement stables et de périodes d'activité exceptionnellement élevées, susceptibles de refléter des contextes opérationnels spécifiques ou des situations atypiques.

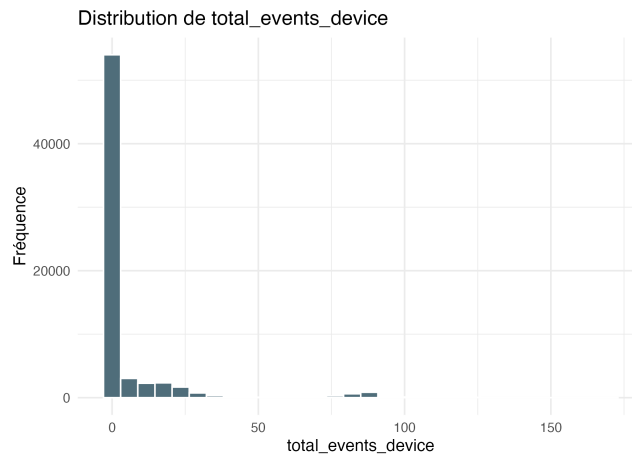


Figure 4-4 : Distribution hebdomadaire du nombre total d'événements Device (*total_events_device*)

Enfin, l'analyse des horaires de connexion met en évidence des régularités temporelles marquées. Comme présenté en annexe 3, la grande majorité des événements Connect se concentre entre 8 h et 17 h, correspondant aux plages horaires habituelles d'un environnement organisationnel standard. Deux pics distincts se dégagent, en début de journée et en début d'après-midi, généralement associés aux ouvertures de session et aux reconnections après des interruptions de travail. À l'inverse, les connexions enregistrées en dehors de ces plages ne représentent qu'environ 2,75 % des événements. Bien que marginales, ces activités hors horaires constituent des signaux faibles particulièrement pertinents, fréquemment associés par les experts à des comportements nécessitant une attention accrue.

La table Logon, complémentaire à la table Device, permet d'analyser plus finement les séquences d'authentification des utilisateurs, en distinguant explicitement les événements de connexion (Logon) et de déconnexion (Logoff). Elle offre ainsi un éclairage plus précis sur les habitudes d'accès, la fréquence des sessions et la cohérence temporelle des comportements, notamment à travers l'identification de situations atypiques telles que des déconnexions sans connexion préalable au cours d'une même période.

Les statistiques descriptives des principales variables comportementales agrégées à l'échelle hebdomadaire sont présentées dans le tableau 4.2. Elles fournissent une vue d'ensemble des volumes d'activité, de la dispersion des comportements et des tendances centrales associées aux événements d'authentification.

Tableau 4-2 : Statistiques descriptives des variables comportementales hebdomadaires de la table Logon

Variables Logon	Nombre total	Moyenne	Écart-type	Minimum	1er quartile (Q1)	Médiane	3e quartile (Q3)	Maximum
nb_logoff	384 268	5.72	2.99	1	5	5	9	35
nb_logon	470 591	7.00	3.79	1	5	5	9	35
total_events_logon	854 859	12.7	6.54	2	10	10	14	70

Globalement, les variables (nb_logon) et (nb_logoff) présentent des distributions comparables, avec des moyennes respectives de 7,0 et 5,72 événements par semaine et des médianes identiques (5 événements). Toutefois, le volume total de connexions demeure systématiquement supérieur à celui des déconnexions, comme en témoigne également la variable agrégée (total_events_logon), dont la moyenne hebdomadaire s'élève à 12,7 événements. Cette asymétrie suggère des pratiques organisationnelles courantes, telles que des sessions laissées ouvertes, des fermetures implicites gérées par le système ou des redémarrages automatiques, mais peut également refléter certaines limites inhérentes à la journalisation des événements.

Les figures 4.5 et 4.6 illustrent les distributions hebdomadaires de (nb_logon) et (nb_logoff). Dans les deux cas, les distributions sont fortement asymétriques à droite : la majorité des observations se concentre autour de faibles volumes d'événements, tandis qu'un nombre restreint de semaines présente des valeurs nettement plus élevées, pouvant atteindre 35 connexions ou déconnexions. Cette structure traduit l'existence de comportements routiniers dominants, ponctués de périodes d'activité plus intensive, susceptibles d'être associées à des contextes spécifiques tels que des changements de poste, des interventions techniques ou des situations exceptionnelles.

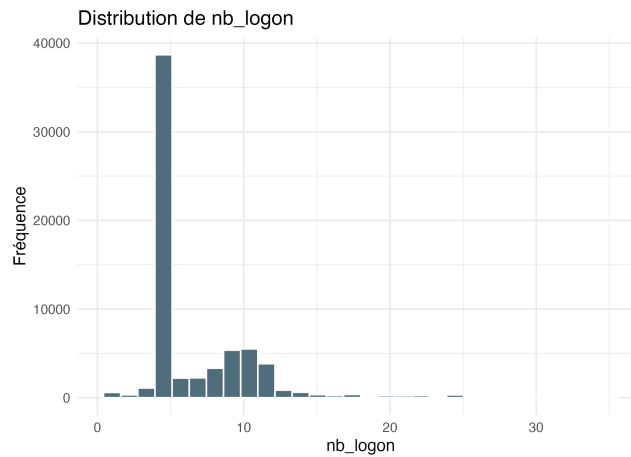


Figure 4-5 : Distribution hebdomadaire du nombre de connexions (nb_logon)

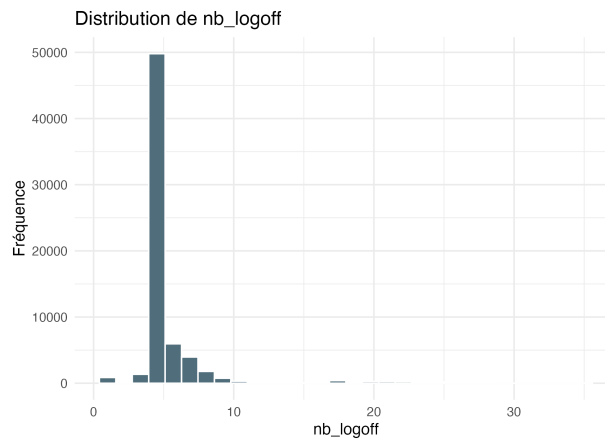


Figure 4-6 : Distribution hebdomadaire du nombre de déconnexions (nb_logoff)

La figure 4.7, représentant la distribution de total_events_logon, confirme cette hétérogénéité. La médiane se situe à 10 événements hebdomadaires, tandis que certaines semaines dépassent largement ce seuil, avec des valeurs maximales atteignant 70 événements. Cette variabilité met en évidence la coexistence de profils d'utilisation stables et de séquences d'authentification particulièrement actives, qui peuvent constituer des signaux d'intérêt pour des analyses plus approfondies.

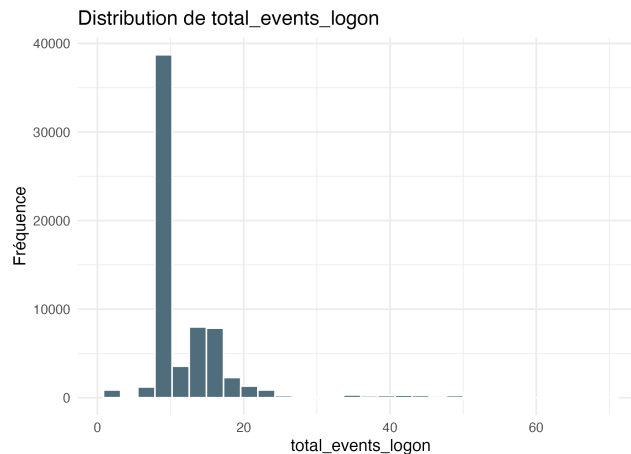


Figure 4-7 : Distribution hebdomadaire du nombre total d'événements d'authentification (total_events_logon)

L'analyse de la fréquence globale des événements Logon et Logoff, présentée en annexe 4, met par ailleurs en évidence une stabilité relative des tendances de Logon et Logoff sur l'ensemble de la période observée. Cette constance constitue un indicateur important de la cohérence interne du jeu de données. À l'inverse, des écarts marqués ou des ruptures dans cette dynamique pourraient signaler des comportements inhabituels, des problèmes de gestion des sessions ou encore des anomalies techniques dans la collecte des journaux.

Ainsi, l'interprétation conjointe des statistiques descriptives hebdomadaires, des distributions empiriques et des tendances globales d'authentification fournit un cadre analytique robuste pour caractériser les habitudes d'accès des utilisateurs. Elle constitue également un socle pertinent pour l'identification ultérieure de séquences atypiques et pour la construction d'indicateurs temporels mobilisés dans les analyses fonctionnelles et les modèles de détection proactive.

L'analyse de la table Email permet de caractériser les habitudes de communication des utilisateurs et d'examiner divers aspects liés à l'envoi de messages électroniques. Cette table renseigne notamment sur le volume d'envois, la présence de pièces jointes, l'utilisation de termes sensibles tels que « confidentiel » ou « secret », ainsi que le nombre de destinataires par envoi. Ces indicateurs contribuent à identifier à la fois les

utilisateurs les plus actifs et ceux dont les pratiques s'écartent des normes organisationnelles, révélant ainsi des comportements potentiellement atypiques.

Les statistiques descriptives des principales variables hebdomadaires dérivées de la table *Email* sont présentées dans le tableau 4.3. Elles offrent une vue d'ensemble des volumes d'envois, de leur dispersion et des tendances centrales observées. La variable (*total_emails*) affiche une moyenne hebdomadaire de 39,1 courriels par utilisateur, avec une médiane de 44, ce qui indique une activité de messagerie soutenue mais relativement homogène pour la majorité des utilisateurs. Néanmoins, la présence de valeurs maximales élevées, pouvant atteindre 180 courriels par semaine, révèle l'existence de profils particulièrement actifs, susceptibles de correspondre à des fonctions impliquant une charge communicationnelle importante.

Tableau 4-3 : Statistiques descriptives des variables comportementales hebdomadaires de la table *Email*

Variables Email	Nombre total	Moyenne	Écart-type	Minimum	1er quartile (Q1)	Médiane	3e quartile (Q3)	Maximum
<i>total_emails</i>	2 629 979	39.12	26.31	1	15	44	55	180
<i>nb_attachments</i>	1 061 449	15.79	17.44	0	1	10	25	150

La figure 4.8 illustre la distribution hebdomadaire du nombre total de courriels envoyés. Cette distribution présente une asymétrie marquée à droite, traduisant la coexistence d'un comportement majoritaire centré autour de volumes modérés et de périodes d'activité exceptionnellement élevées. Ce type de distribution est cohérent avec les pratiques observées en milieu organisationnel, où une minorité d'utilisateurs concentre une part importante des échanges.

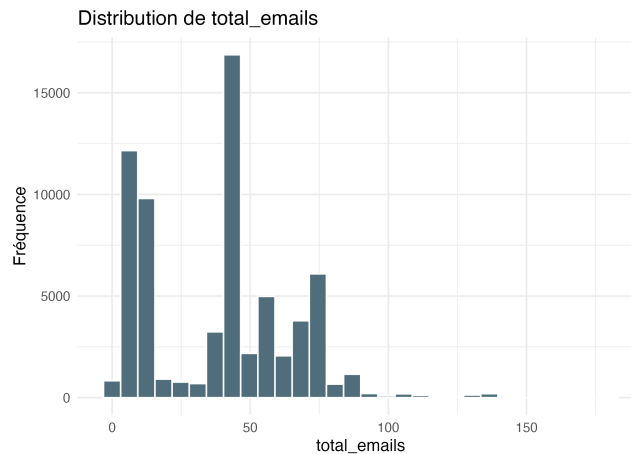


Figure 4-8 : Distribution hebdomadaire du nombre total de courriels envoyés (*total_emails*)

La variable (*nb_attachments*), qui mesure le nombre de pièces jointes envoyées par semaine, présente également une forte variabilité. Bien que la majorité des observations se situe à des niveaux relativement faibles, certaines semaines enregistrent des volumes très élevés, atteignant jusqu'à 150 pièces jointes. La figure 4.9 met en évidence une distribution fortement asymétrique, caractérisée par une concentration des observations autour de faibles valeurs et une longue queue à droite. Cette hétérogénéité reflète des usages différenciés de la messagerie, allant de communications essentiellement textuelles à des échanges fréquents de documents.

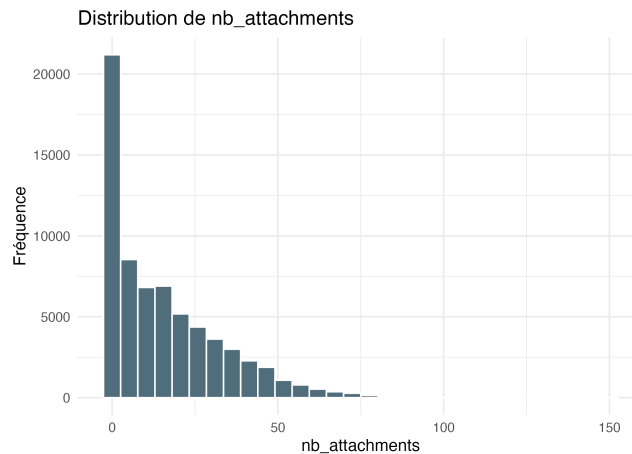


Figure 4-9 : Distribution hebdomadaire du nombre de pièces jointes envoyées (nb_attachments)

Ces distributions suggèrent que, si la majorité des utilisateurs adopte des pratiques de communication relativement stables, une fraction restreinte présente des comportements plus intensifs, susceptibles de constituer des signaux d'intérêt dans une perspective de détection proactive. En particulier, des volumes élevés de courriels ou de pièces jointes, surtout lorsqu'ils sont combinés à d'autres dimensions temporelles ou contextuelles, peuvent signaler des situations nécessitant une analyse plus approfondie.

Dans l'ensemble, l'analyse descriptive de la table Email met en évidence une hétérogénéité marquée des pratiques de communication électronique. Ces résultats constituent une base essentielle pour la construction d'indicateurs comportementaux plus fins et pour l'intégration des signaux liés aux communications dans les trajectoires fonctionnelles et les modèles de détection mobilisés dans les analyses ultérieures.

Pour la table File, l'analyse porte sur plusieurs dimensions essentielles à la compréhension des comportements liés à la manipulation de documents : le volume d'opérations par utilisateur, la fréquence d'accès à des fichiers contenant des mots sensibles et l'activité réalisée en dehors des heures habituelles de travail, en particulier après 20 h. L'examen conjoint de ces indicateurs permet non seulement d'identifier les utilisateurs les plus actifs, mais aussi de repérer les contextes d'usage susceptibles de s'écarter des pratiques organisationnelles attendues.

Tableau 4-4 : Statistiques descriptives des variables comportementales hebdomadaires de la table File

Variable File	Nombre total	Moyenne	Écart-type	Minimum	1er quartile (Q1)	Médiane	3e quartile (Q3)	Maximum
nb_pc_file	17 533	1.28	1.45	1.0	1.0	1.0	1.0	15
total_events_file	445 581	32.43	39.10	1.0	8.0	17.0	35.0	208

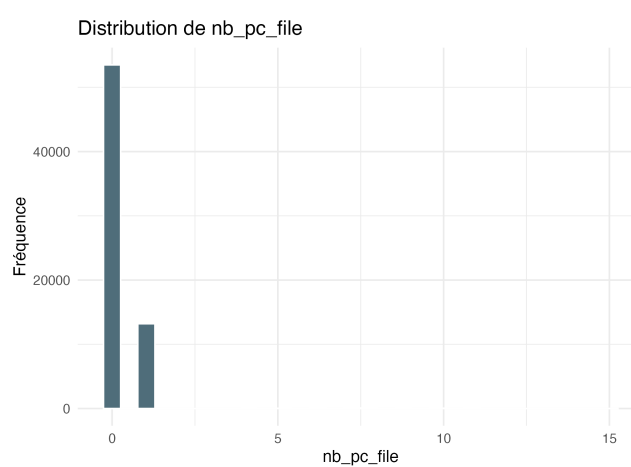


Figure 4-10 : Distribution hebdomadaire du nombre de postes utilisés pour les accès aux fichiers (nb_pc_file)

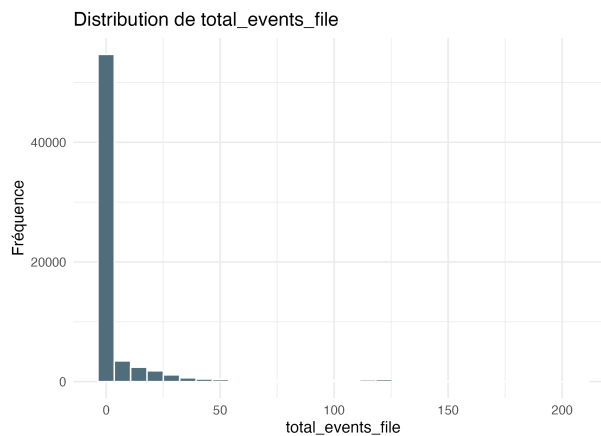


Figure 4-11 : Distribution hebdomadaire du nombre total d'événements liés aux fichiers (*total_events_file*)

L'Annexe 5 offre une illustration particulièrement éclairante de ces dynamiques en présentant les dix utilisateurs ayant effectué le plus grand nombre d'opérations sur des fichiers après 20 h. L'activité nocturne est régulièrement mise en avant dans la littérature comme un signal potentiel de risque, notamment lorsqu'elle s'accompagne d'un volume élevé de manipulations de documents ou d'accès à des fichiers sensibles. Dans le jeu de données CERT, certains utilisateurs se distinguent nettement par une activité intensive en dehors des plages horaires régulières (voir Annexe 5). Bien que de tels comportements puissent parfois refléter des contraintes professionnelles légitimes, échéances serrées, travail hors site ou besoin ponctuel de flexibilité, ils constituent également des éléments importants à surveiller en raison de leur possible association avec des usages atypiques ou potentiellement malveillants.

Ainsi, la combinaison des volumes d'accès, de la nature potentiellement sensible des fichiers manipulés et des horaires de modification contribue à construire des indicateurs pertinents pour la détection précoce de comportements anormaux. Ce travail descriptif établit les premières bases permettant d'identifier les utilisateurs présentant des patterns d'activité nécessitant une attention particulière dans les analyses subséquentes, notamment dans le cadre de la modélisation fonctionnelle et des approches supervisées. Pour la table File, l'analyse porte sur plusieurs dimensions essentielles à la compréhension des comportements liés à la manipulation de

documents : le volume d'opérations par utilisateur, la fréquence d'accès à des fichiers contenant des mots sensibles, et l'activité réalisée en dehors des heures habituelles de travail, notamment après 20 h. L'examen de ces indicateurs permet d'identifier les utilisateurs les plus actifs ainsi que les contextes d'usage susceptibles de s'écarter des pratiques organisationnelles normales.

La table Psychometric est exploitée pour analyser les traits de personnalité des utilisateurs. La distribution des scores selon les cinq dimensions du modèle OCEAN, Ouverture, Conscienciosité, Extraversion, Agréabilité et Neuroticisme, est visualisée, et une matrice de corrélation est produite pour explorer les relations potentielles entre ces traits. Ces variables ne sont toutefois pas intégrées aux modèles de détection développés dans ce mémoire. Leur analyse vise plutôt à fournir un éclairage complémentaire sur les profils comportementaux des utilisateurs et à contextualiser certains schémas observés dans les journaux d'activité. Cette exploration permet ainsi de mieux comprendre, sur le plan descriptif et théorique, comment certaines caractéristiques de personnalité pourraient influencer les interactions avec le système informatique ou être associées à une propension accrue à adopter des comportements atypiques ou à risque, sans pour autant présumer de leur valeur prédictive opérationnelle dans le cadre du présent modèle.

Enfin, la table Http est analysée afin de caractériser l'activité de navigation web des utilisateurs. L'exploration repose sur des requêtes SQL exécutées via DuckDB, permettant d'identifier les volumes de requêtes HTTP, les domaines et URL les plus fréquemment consultés, ainsi que certaines régularités temporelles dans les comportements de navigation. Cette analyse vise à dégager des profils d'usage généraux et à repérer des schémas susceptibles de s'écarter des pratiques organisationnelles attendues.

Le tableau 4.5 présente les statistiques descriptives des principales variables hebdomadaires dérivées de la table Http. La variable (http_total_events), qui agrège l'ensemble des requêtes HTTP effectuées par un utilisateur sur une semaine, affiche une moyenne élevée (422,9 événements) et une dispersion importante (écart-type de 289,9), traduisant une forte hétérogénéité des comportements de navigation. La médiane (475

événements) et le troisième quartile (570 événements) indiquent que la majorité des utilisateurs présentent une activité soutenue, tandis que certaines semaines atteignent des volumes exceptionnellement élevés, avec un maximum observé de 2 000 requêtes.

Tableau 4-5 : Statistiques descriptives des variables comportementales hebdomadaires de la table *http*

Variable File	Nombre total	Moyenne	Écart-type	Minimum	1er quartile (Q1)	Médiane	3e quartile (Q3)	Maximum
http_total_events	28 434 423	422.92	289.94	9.0	145.0	475.0	570.0	2000.0

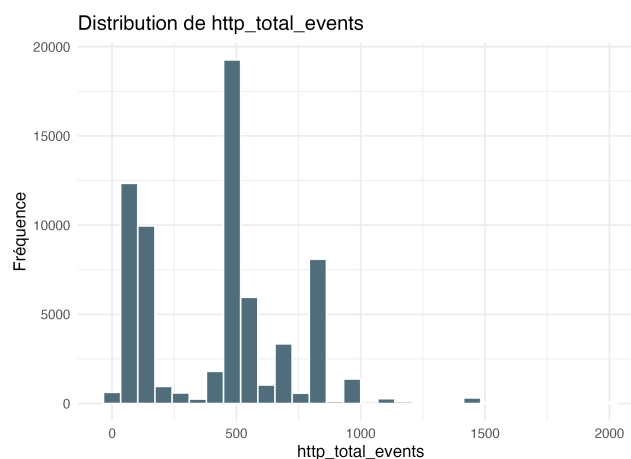


Figure 4-12 : Distribution hebdomadaire du nombre total d'événements HTTP (*http_total_events*)

La distribution illustrée à la figure 4.12 est fortement asymétrique à droite. Elle met en évidence une concentration d'observations autour de valeurs modérées, correspondant à des usages routiniers de la navigation web, ainsi qu'une longue queue associée à des volumes de requêtes particulièrement élevés. Ces valeurs extrêmes peuvent refléter des rôles spécifiques nécessitant une navigation intensive, mais également des

comportements atypiques susceptibles de signaler une exploration automatisée, une recherche ciblée d'information ou des usages non conformes aux pratiques standards.

L'analyse qualitative complémentaire des URL et des domaines consultés met par ailleurs en évidence des requêtes associées à des thématiques sensibles, telles que des forums techniques avancés, des plateformes liées au hacking ou des sites référencés comme appartenant au dark web. De plus, certaines machines apparaissent utilisées par un nombre anormalement élevé d'utilisateurs distincts, ce qui peut indiquer un usage partagé, des comptes mutualisés ou des anomalies dans la journalisation des sessions.

Dans l'ensemble, l'analyse de la table Http permet de dégager des tendances comportementales structurantes, d'identifier des profils de navigation intensifs ou atypiques et d'enrichir le profilage des utilisateurs. Ces éléments constituent une base essentielle pour la construction de variables dérivées intégrant la dimension web, et renforcent la capacité des modèles ultérieurs à capter des signaux faibles liés à des comportements potentiellement malveillants dans un environnement organisationnel simulé.

1.2 Préparation des données et construction des trajectoires fonctionnelles

La préparation des données consiste d'abord à fusionner les observations comportementales hebdomadaires avec la table des incidents, permettant ainsi de construire une mesure cohérente de la malveillance au niveau de l'utilisateur et de la semaine. La cohérence de ce processus repose sur une caractéristique fondamentale du jeu de données CERT : conformément à la documentation officielle, « The logs contain ONLY events that occurred » (CERT). Autrement dit, l'absence d'enregistrement pour une variable donnée n'indique pas un manque de données, mais bien une absence d'activité observée. Cette propriété justifie le choix méthodologique de coder systématiquement les absences d'événements par la valeur zéro plutôt que par des valeurs manquantes.

À partir de ces données fusionnées, les douze variables comportementales sont réorganisées en matrice de type utilisateur $\times$ semaine. Cette restructuration garantit une base temporelle parfaitement homogène pour l'ensemble de la cohorte, de la semaine 2009-53 à la semaine 2011-20, soit soixante-treize points temporels communs. Les semaines où aucun événement n'est enregistré pour un utilisateur sont ajoutées explicitement et remplies de zéros. Cette étape est essentielle, car elle permet d'analyser des trajectoires complètes, continues et comparables, sans que des absences d'observation ne fragmentent ou ne biaisent les profils individuels.

Les séries temporelles ainsi obtenues sont ensuite transformées en trajectoires fonctionnelles au moyen d'un lissage par B-splines. La base retenue comprend douze fonctions de base couvrant l'ensemble de l'intervalle temporel. Le choix de ce nombre résulte d'un compromis méthodologique entre flexibilité et stabilité. D'un côté, un nombre trop faible de B-splines conduit à un sur-lissage, empêchant la détection de variations comportementales pertinentes telles que des hausses progressives d'activité, des ruptures ponctuelles ou des pics soudains. De l'autre, un nombre trop élevé favorise un surajustement à des fluctuations hebdomadaires mineures, rendant la courbe sensible au bruit. Dans un contexte de séries hebdomadaires relativement longues (73 points), douze fonctions de base offrent un niveau de granularité suffisant pour modéliser des comportements complexes tout en préservant la robustesse statistique. Cette configuration est également cohérente avec les recommandations de la littérature en analyse fonctionnelle, qui privilégie un nombre modéré de B-splines pour des données présentant une fréquence et une structure similaires (Annexe 6).

Le lissage par B-splines permet ainsi de convertir chaque série discrète en une courbe continue, atténuant les fluctuations locales tout en conservant les tendances structurelles caractéristiques de chaque utilisateur. L'examen des trajectoires lissées révèle une forte hétérogénéité comportementale. Certains employés affichent des profils réguliers, marqués par une stabilité d'une semaine à l'autre, tandis que d'autres présentent des fluctuations importantes ou des épisodes d'activité inhabituellement élevés sur certains canaux. Pour la variable (total_events_logon), par exemple, la moyenne hebdomadaire se situe autour de 12,9 événements, mais certains utilisateurs maintiennent un niveau constant proche de dix, alors que d'autres dépassent

occasionnellement les 46 événements au cours d'une même semaine. Cette diversité met en évidence des schémas d'utilisation très variés, qui seraient difficilement capturés par des approches classiques basées sur des agrégations ponctuelles.

La construction de ces courbes fonctionnelles constitue ainsi une étape centrale de la méthodologie. Elle repose sur l'hypothèse fondamentale selon laquelle les comportements numériques évoluent de manière continue dans le temps et que leurs propriétés temporelles, tendance, amplitude, dérive ou irrégularité, constituent des informations essentielles pour distinguer des profils normaux et anormaux. En offrant une représentation riche, cohérente et comparable des pratiques numériques des utilisateurs, cette étape prépare le terrain pour les analyses fonctionnelles subséquentes, notamment l'extraction des composantes principales et la modélisation supervisée.

1.3 Analyse en composantes principales fonctionnelles

L'Analyse en Composantes Principales Fonctionnelles (FPCA) constitue une étape centrale de la méthodologie, car elle permet de résumer de manière compacte les trajectoires comportementales issues du lissage par B-splines. Contrairement à l'Analyse en Composantes Principales (ACP) classique, qui s'applique à des données discrètes, la FPCA agit directement sur des courbes continues, capturant non seulement les différences de niveau, mais aussi celles de forme et d'évolution temporelle (Annexe 7). Cette étape méthodologique revêt un rôle particulier dans le cadre de ce mémoire puisqu'elle permet d'évaluer si les trajectoires simulées dans le jeu de données CERT présentent des modes de variation cohérents avec ce que l'on pourrait attendre dans des environnements organisationnels réels.

L'Annexe 7 illustre clairement l'influence des trois premières composantes principales sur les trajectoires hebdomadaires d'activité logon. La première composante reflète principalement des différences globales d'intensité : les courbes situées au-dessus ou en dessous de la trajectoire moyenne témoignent d'utilisateurs présentant des niveaux d'activité plus ou moins soutenus. La deuxième composante modifie la forme générale

de la courbe en opposant le début et la fin de la période d'observation, ce qui met en évidence des tendances temporelles différentielles, comme une diminution progressive ou une accentuation tardive de l'activité. La troisième composante introduit des fluctuations plus localisées, traduisant des variations fines qui seraient difficilement perceptibles à partir de la trajectoire moyenne seule. Ensemble, ces composantes décrivent les principaux modes de variation caractéristiques du comportement des utilisateurs dans le jeu de données CERT.

Une FPCA distincte est appliquée à chacune des douze variables comportementales, permettant d'identifier les axes de variation les plus structurants pour chaque type d'activité. Comme le montre l'Annexe 8, la majorité des variables atteignent un seuil d'environ 95 % de variance expliquée après une ou deux composantes seulement, ce qui indique que les trajectoires temporelles observées sont dominées par quelques motifs principaux. Seules certaines variables, notamment celles associées à des comportements plus irréguliers ou à une variabilité inter-utilisateurs plus prononcée, nécessitent l'ajout de composantes supplémentaires, sans toutefois excéder la limite fixée de cinq composantes.

Le critère de rétention basé sur la proportion cumulée de variance fonctionnelle expliquée garantit ainsi un compromis adéquat entre richesse descriptive et parcimonie. Cette procédure assure une représentation fidèle des différences comportementales tout en maintenant un niveau de complexité compatible avec les exigences des modèles supervisés utilisés par la suite.

Les scores FPCA associés à ces composantes fournissent une représentation concise mais informative de chaque trajectoire. La première composante traduit généralement le niveau moyen d'activité d'un utilisateur sur la période étudiée, tandis que les composantes suivantes capturent les variations plus fines, telles que les tendances globales (croissantes ou décroissantes), les fluctuations saisonnières ou encore les comportements atypiques marqués par des pics isolés ou des ruptures nettes. Ces différentes dimensions permettent de distinguer des profils d'activité stables, rythmiques ou irréguliers, ce qui constitue un apport méthodologique essentiel pour l'analyse comportementale appliquée aux menaces internes malveillantes.

L'un des avantages majeurs de la FPCA dans ce contexte réside dans sa capacité à réduire drastiquement la dimensionnalité des données. Alors que les trajectoires initiales comptent soixante-treize points hebdomadaires, leur représentation fonctionnelle peut être résumée, pour chaque variable, par deux à cinq scores seulement. L'ensemble des scores obtenus pour les douze variables constitue ainsi un vecteur de caractéristiques faiblement dimensionnel mais très expressif, qui alimente ensuite les modèles de classification supervisée. Cette réduction de dimension permet non seulement d'améliorer l'efficacité computationnelle des modèles, mais aussi de limiter le risque de sur-apprentissage, tout en conservant les informations essentielles relatives aux dynamiques comportementales.

L'analyse exploratoire des harmoniques principales met en évidence des variations typiques dans les comportements utilisateurs. Par exemple, certaines variables présentent une première composante dominante reflétant simplement l'intensité globale d'usage (en particulier les événements de connexion et les courriels envoyés), tandis que d'autres montrent des composantes supérieures révélant des motifs temporels plus sophistiqués, tels que des cycles hebdomadaires, des phases d'augmentation progressive ou des fluctuations irrégulières caractéristiques de comportements anormaux (Annexe 9).

Ces observations confirment que la FPCA est bien adaptée à la capture de signaux comportementaux subtils mais importants pour les scénarios de menace interne malveillante. Les scores fonctionnels obtenus constituent dès lors un espace de représentation optimisé que les modèles supervisés exploiteront dans la section suivante, permettant d'évaluer empiriquement leur capacité à distinguer les utilisateurs malveillants des utilisateurs légitimes.

1.4 Modélisation et validation croisée

L'évaluation des modèles prédictifs s'effectue au moyen d'une validation croisée stratifiée à cinq plis, répétée à cinq reprises afin d'obtenir des estimations plus stables des performances. Ce protocole s'avère particulièrement pertinent dans un contexte

marqué par un déséquilibre important entre les utilisateurs malveillants et non malveillants, puisque la stratification garantit que chaque partition conserve un ratio comparable de classes. Cette contrainte assure que les modèles ne sont pas entraînés, par hasard, sur des sous-échantillons contenant trop peu d'observations positives, ce qui pourrait compromettre leur capacité à apprendre des signaux significatifs.

À chaque répétition de la validation croisée, les modèles sont entraînés sur quatre plis et évalués sur le cinquième, de manière que chaque utilisateur apparaisse exactement une fois dans un ensemble de test. Cette rotation systématique réduit l'influence des partitions aléatoires et fournit une estimation de la variabilité inter-folds, permettant de quantifier la robustesse des performances observées. Un tel protocole s'inscrit dans les recommandations méthodologiques pour les jeux de données de petite taille et fortement déséquilibrés, où un simple train/test split serait insuffisant pour caractériser la stabilité du modèle.

Les modèles produisent des probabilités d'appartenance à la classe « malveillant ». Ces probabilités sont ensuite converties en décisions binaires à l'aide du seuil optimal de Youden, déterminé à partir de la courbe ROC de chaque fold. Ce choix de seuil est justifié par sa capacité à maximiser simultanément la sensibilité (rappel) et la spécificité, ce qui est crucial dans un cadre de détection d'incidents rares où les faux négatifs doivent être limités sans pour autant générer un volume excessif d'alertes. L'ajustement du seuil à chaque itération permet également de tenir compte des fluctuations intrinsèques à chaque sous-échantillon, plutôt que d'imposer une valeur fixe potentiellement non optimale (Annexe 10).

Les performances des modèles sont évaluées à l'aide de métriques standard dans le domaine de la détection des menaces internes malveillantes, notamment :

- L'AUC (Area Under the Curve), qui mesure la capacité du modèle à distinguer les utilisateurs malveillants des non malveillants indépendamment du seuil choisi ;
- La précision, qui renseigne sur la proportion d'alertes correctes parmi les prédictions positives ;

- Le rappel, qui évalue la capacité du modèle à identifier réellement tous les utilisateurs malveillants ;
- Le F1-score, qui synthétise précision et rappel de manière équilibrée.

L'utilisation conjointe de ces mesures offre une vision nuancée de la performance, particulièrement nécessaire lorsqu'une classe est rare. Par exemple, l'AUC fournit une indication globale de la capacité discriminante du modèle, tandis que le F1-score permet d'évaluer un compromis opérationnel entre faux positifs et faux négatifs. Les résultats consolidés sur les vingt-cinq configurations (5 folds $\times$ 5 répétitions) permettent ainsi de dégager une estimation robuste des performances moyennes, mais également d'analyser leur dispersion, un aspect essentiel dans des scénarios de sécurité où la stabilité du modèle prime autant que sa performance nominale.

En complément, l'analyse des modèles porte une attention particulière au risque de sur-apprentissage. L'utilisation d'une validation croisée répétée, combinée à une réduction de dimension préalable par FPCA, limite efficacement ce risque, mais l'examen des divergences entre les performances en apprentissage et en test demeure indispensable. Les modèles dont les performances varient fortement entre les folds ou entre les répétitions sont considérés comme moins fiables opérationnellement, même si leurs scores moyens semblent élevés.

Dans son ensemble, cette stratégie d'évaluation offre une appréciation rigoureuse, stable et interprétable de la qualité des modèles, tout en respectant les contraintes propres aux données de sécurité, où la rareté des incidents impose des précautions méthodologiques particulières.

1.5 Résultats de la régression logistique

La régression logistique entraînée sur les scores issus de la FPCA présente des performances solides et cohérentes, confirmant la valeur ajoutée de la représentation fonctionnelle pour la détection des comportements malveillants. Le tableau 4.4 ci-dessous synthétise les performances moyennes obtenues sur l'ensemble des 25

itérations de validation croisée, et permet de situer la régression logistique par rapport à la forêt aléatoire.

Tableau 4-6 : Comparaison des performances des modèles (FPCA) pour la détection des menaces internes malveillantes

Méthode	AUC	Rappel	Précision	F1
Régression logistique	0.942	0.937	0.523	0.661
Forêt aléatoire	0.997	0.997	0.832	0.895

Sur ces 25 itérations, la régression logistique atteint une AUC moyenne d'environ 0,94, avec un écart-type de 0,04. Cette valeur élevée, combinée à une faible variabilité, témoigne d'une capacité discriminante stable d'un fold à l'autre et illustre la pertinence des composantes principales fonctionnelles comme vecteurs de signaux comportementaux.

L'un des résultats les plus marquants concerne le rappel, qui s'établit en moyenne à 0,94. Autrement dit, la grande majorité des utilisateurs malveillants sont correctement identifiés par le modèle, ce qui est particulièrement crucial dans le contexte de la sécurité organisationnelle, où les faux négatifs représentent un risque majeur. Cette sensibilité élevée est obtenue sans recourir à des techniques de sur-échantillonnage ou à des ajustements artificiels des données, ce qui souligne la force discriminante intrinsèque des caractéristiques fonctionnelles dérivées de la FPCA.

La précision, en revanche, demeure modérée, autour de 0,52 en moyenne. Ce niveau reflète la présence de faux positifs, un phénomène attendu dans les scénarios où la classe positive est rare. La logistique fonctionnelle tend à élargir le périmètre des alertes afin de ne pas manquer les cas malveillants, ce qui se traduit par une augmentation du nombre de prédictions positives erronées. Ce compromis, rappel élevé, précision plus

faible, est courant dans les systèmes de détection et doit être interprété à l'aune des impératifs opérationnels : une organisation peut tolérer un certain volume d'alertes supplémentaires si cela permet de réduire substantiellement les incidents non détectés.

Le F1-score, qui représente le compromis harmonique entre précision et rappel, atteint en moyenne 0,66, ce qui constitue une amélioration significative par rapport aux modèles non fonctionnels évalués lors des phases exploratoires. Ces modèles, fondés sur les données brutes agrégées ou sur des statistiques ponctuelles, présentent des F1-scores nettement inférieurs, illustrant leur incapacité à capturer la dynamique temporelle complexe caractéristique des comportements malveillants dans le jeu de données CERT. L'intégration de la dimension fonctionnelle renforce donc clairement la capacité du modèle à appréhender des signaux faibles répartis dans le temps.

L'analyse de la variabilité des performances entre les différentes partitions de validation montre que les résultats du modèle restent relativement stables : les écarts-types calculés pour l'AUC, le rappel et le F1-score demeurent faibles, ce qui traduit une robustesse structurelle de la régression logistique. Cette stabilité est essentielle pour envisager une utilisation opérationnelle, car elle suggère que les performances ne dépendent pas de variations fortuites dans la composition des échantillons d'apprentissage et de test (Annexe 11).

Sur le plan méthodologique, il est intéressant de souligner que la forte sensibilité du modèle s'explique en partie par le seuil de décision optimal déterminé à chaque fold par le critère de Youden. Cette stratégie, fondée sur la courbe ROC, permet de trouver l'équilibre le plus favorable entre sensibilité et spécificité pour chaque sous-échantillon, plutôt que d'appliquer un seuil uniforme potentiellement sous-optimal.

Enfin, l'examen des coefficients de la régression logistique permet d'identifier quelles composantes fonctionnelles contribuent le plus à la discrimination entre utilisateurs malveillants et non malveillants. Certaines premières composantes, associées au niveau moyen d'activité, apparaissent comme des déterminants importants, mais plusieurs composantes supérieures, capturant des motifs temporels plus fins, jouent

également un rôle significatif. Cela montre que les comportements malveillants ne se distinguent pas uniquement par leur intensité globale, mais aussi par leur structure temporelle, leurs irrégularités et leurs variations abruptes (Annexe 12).

Dans l'ensemble, les résultats de la régression logistique confirment l'apport substantiel de l'approche FDA : la transformation des séries temporelles en trajectoires continues, suivie de leur réduction en composantes principales, permet d'extraire des représentations informatives par lesquelles le modèle atteint des performances à la fois élevées, stables et adaptées à la détection de comportements déviants dans un environnement où les incidents sont rares.

1.6 Résultats de la forêt aléatoire

La forêt aléatoire appliquée aux scores issus de la FPCA présente des performances exceptionnellement élevées, nettement supérieures à celles observées avec la régression logistique, comme le montre le tableau 4.5. Sur l'ensemble des 25 itérations de validation croisée, le modèle atteint une AUC moyenne de 0,997, accompagnée d'une variance extrêmement faible. Cette stabilité témoigne d'une capacité discriminante remarquable et suggère que la forêt aléatoire capture efficacement les motifs complexes présents dans les trajectoires comportementales.

Le rappel, qui mesure la proportion d'utilisateurs malveillants correctement identifiés, atteint 0,997, indiquant que le modèle détecte presque systématiquement les comportements malveillants. La précision, établie à 0,83, est également très élevée, montrant que la majorité des alertes générées correspond à des utilisateurs réellement malveillants. Le F1-score, qui combine précision et rappel, s'élève en moyenne à 0,90, témoignant d'un compromis harmonieux entre sensibilité et spécificité et plaçant ce modèle comme l'un des plus performants dans le cadre étudié.

Ces résultats s'expliquent en grande partie par la capacité des forêts aléatoires à modéliser des relations non linéaires, des interactions complexes entre les composantes principales et des patterns irréguliers dans les trajectoires temporelles.

Contrairement à la régression logistique, limitée aux effets additifs et linéaires, la forêt aléatoire explore un espace décisionnel multidimensionnel et fortement structuré, ce qui lui permet de capturer des signaux subtils présents dans les courbes fonctionnelles.

L'analyse des importances de variables révèle d'ailleurs que certaines composantes supérieures, correspondant à des fluctuations locales ou à des motifs atypiques, contribuent fortement à la décision. Cette observation montre que les comportements malveillants simulés ne se distinguent pas seulement par des tendances générales (captées par les premières composantes), mais aussi par des irrégularités spécifiques dans le temps, ce que la forêt aléatoire parvient à détecter efficacement (Annexe 13).

La variabilité extrêmement faible des performances entre les folds confirme la robustesse du modèle. À l'inverse de plusieurs approches supervisées sensibles à la composition des échantillons d'apprentissage, la forêt aléatoire reste stable, ce qui en fait un candidat favorable à une utilisation opérationnelle.

Cependant, ces performances doivent être interprétées avec prudence. Les comportements malveillants simulés par le jeu de données CERT sont souvent plus structurés, plus homogènes et plus facilement détectables que ceux observés dans des environnements réels. De nombreuses études soulignent que les jeux CERT ont tendance à accentuer les signaux caractéristiques des menaces internes malveillantes, ce qui peut conduire à une surestimation des performances des modèles. La quasi-perfection obtenue par la forêt aléatoire peut donc refléter autant les propriétés du jeu de données que les qualités intrinsèques du modèle.

Enfin, il convient de noter que la forêt aléatoire implique un coût computationnel plus élevé que la régression logistique, notamment en raison du nombre d'arbres et de l'importance des interactions modélisées. Toutefois, la réduction de dimension opérée par la FPCA permet de limiter cette complexité, rendant le modèle utilisable dans un contexte opérationnel, bien que son interprétabilité demeure plus limitée qu'un modèle paramétrique.

Dans l'ensemble, les résultats de la forêt aléatoire montrent que l'approche FDA permet de dégager des représentations extrêmement discriminantes des comportements

utilisateurs, et que les modèles non linéaires sont capables d'en exploiter pleinement la richesse. Bien que les performances doivent être replacées dans le contexte artificiel du jeu de données CERT, elles indiquent clairement que l'analyse fonctionnelle combinée à des techniques d'apprentissage avancées constitue une approche prometteuse pour la détection proactive des menaces internes malveillantes.

2. Approche qualitative

2.1 Profil des participants et crédibilité du panel

Le panel d'experts réunit des professionnels aux profils complémentaires couvrant l'ensemble du cycle de vie des incidents : direction de la sécurité, administration réseau, réponse aux incidents, tests d'intrusion, gouvernance TI et recherche académique. Tous disposent d'une expérience substantielle, entre 5 et plus de 20 années en cybersécurité, incluant la manipulation de données de journalisation et, pour plusieurs, la gestion de cas réels de menaces internes malveillantes. Cette diversité de rôles, allant de la prévention à la détection, en passant par la réponse et la gouvernance, renforce la crédibilité analytique du panel. Elle permet en effet de croiser des perspectives techniques, organisationnelles et scientifiques, offrant ainsi une compréhension riche et multidimensionnelle du phénomène étudié.

Les six entrevues semi-dirigées menées auprès de ces experts visent à documenter :

- (1) les comportements jugés caractéristiques d'une menace interne malveillante,
- (2) les signaux effectivement mobilisés par les praticiens pour la détection, et
- (3) les limites perçues des jeux de données simulés, notamment CERT.

Ces entretiens constituent la base du processus de codification, réalisé en deux temps. D'abord, une grille de codification déductive est élaborée à partir de la littérature scientifique sur la menace interne malveillante, incluant les comportements malveillants, les signaux faibles, les profils à risque et les limites des mécanismes de détection. Ensuite, un codage inductif est appliqué afin de faire émerger les thèmes

nouveaux ou inattendus directement issus des discours des experts. Cette approche hybride permet de structurer l'analyse tout en respectant la richesse et la complexité des représentations expertes.

2.2 Codage des données

Le processus d'analyse s'appuie sur la démarche de codification présentée dans la méthodologie (Section 4.2). Le premier cycle de codification repose sur une grille déductive élaborée à partir de la littérature scientifique (tableau 3.2) : chaque unité de sens extraite des transcriptions a été associée au code thématique correspondant, assurant ainsi une continuité étroite entre le cadre conceptuel et l'analyse empirique.

Le recours à des scénarios durant les entrevues a également contribué de manière significative à la qualité des données collectées. Conçus pour reproduire des situations réalistes inspirées de la littérature et des comportements simulés dans le jeu de données CERT, ces scénarios visaient à éviter toute référence à des incidents réels ou à des enjeux sensibles de confidentialité ou de politique interne. Cette approche a permis de présenter aux experts des situations crédibles et suffisamment génériques pour favoriser un échange ouvert et sécurisant. Les participants ont d'ailleurs estimé ces scénarios pertinents et représentatifs de leur expérience professionnelle, ce qui a facilité la compréhension des questions, soutenu la réflexion et fourni une base commune pour approfondir l'analyse des signaux comportementaux.

Grâce à ce dispositif, les entrevues ont généré des réponses riches et nuancées, permettant aux experts de décrire en détail les mécanismes sous-jacents aux comportements malveillants, ainsi que les facteurs organisationnels, techniques et psychologiques influençant la détection ou la non-détection de tels comportements. Cette richesse analytique a ouvert la voie au second cycle de codification, inductif cette fois, qui a fait émerger des éléments non anticipés par la grille initiale : contextualisation des signaux selon le rôle occupé, dynamique progressive menant de comportements ambigus à des actes malveillants, et difficultés d'interprétation liées à la zone grise entre erreur, négligence et intention.

Enfin, une interprétation transversale a permis de regrouper l'ensemble des codes en thèmes centraux, articulant les propos des experts autour des enjeux centraux du

mémoire, la détection proactive des menaces internes et le réalisme des comportements simulés dans le jeu de données CERT. Cette approche hybride, combinant codage déductif, codage inductif et analyse fondée sur des scénarios, assure un équilibre solide entre fondements théoriques et apports empiriques, tout en renforçant la validité, la profondeur et la robustesse de l'analyse qualitative.

2.3 Thème : nature et complexité de la menace interne malveillante

Les experts décrivent la menace interne malveillante comme un phénomène qui se construit graduellement, plutôt que comme un événement soudain et isolé. Plusieurs incidents débutent par des gestes ambigus, accès répétitifs « par curiosité », usage d'outils non autorisés ou recours à une clé USB personnelle, avant de se transformer en comportements explicitement malveillants. Comme l'explique un expert : « On voit souvent des incidents internes qui commencent par des gestes anodins. Ce n'est pas toujours de la malveillance explicite au départ. » (Expert A).

Cette ambiguïté est renforcée par la difficulté à distinguer l'erreur de l'intentionnalité : « La frontière entre une erreur et une intention malveillante est parfois très mince. » (Expert B). Cette zone grise, omniprésente dans les organisations, complique la détection et souligne l'importance d'examiner les trajectoires comportementales complètes plutôt que des événements isolés. La menace interne apparaît ainsi comme un phénomène sociotechnique où interagissent comportements individuels, contexte organisationnel et infrastructures de sécurité.

2.4 Thème : comportements malveillants et profils d'utilisateurs à risque

Les entrevues révèlent plusieurs catégories de comportements malveillants prioritaires : exfiltration de données, abus de privilèges, sabotage, fraude et mauvaise gestion d'identifiants. Leur apparition est rarement brutale ; elle s'inscrit dans une dynamique progressive marquée par des accès répétitifs, hors rôle ou hors horaire. Comme le résume un expert : « Avant un sabotage ou une fraude, on observe presque toujours une phase de repérage. L'utilisateur teste les limites. » (Expert D).

L'exfiltration demeure l'un des comportements les plus fréquemment observés dans la pratique : « L'exfiltration reste de loin le comportement le plus fréquent : clé USB, email personnel, stockage externe... » (Expert C).

Du point de vue des profils, les experts soulignent que tout utilisateur peut potentiellement devenir malveillant, mais certains groupes sont plus critiques : employés réguliers ayant accès à des données sensibles, administrateurs systèmes, personnels RH ou financiers, nouveaux employés ou contractuels. Comme le rappellent deux experts :

« On sous-estime les employés réguliers... ils manipulent souvent les données les plus sensibles. » (Expert A) ;

« Les administrateurs sont les plus dangereux... parce qu'ils peuvent tout faire sans laisser de traces. » (Expert E).

Le risque découle donc de la combinaison du rôle, des responsabilités, du contexte organisationnel et non du seul niveau de privilège technique.

2.5 Thème : signaux faibles et limites des outils de détection actuels

L'analyse met également en lumière le rôle central des signaux faibles dans la détection proactive. Ces indices isolés, connexions tardives, accès répétés, périphériques externes, modifications d'habilitations, transferts vers des services externes, prennent tout leur sens lorsqu'ils sont corrélés dans le temps. Comme le souligne un expert : « Un accès hors horaire n'est pas suffisant. C'est quand on combine l'horaire, la fréquence, le type de fichiers et les périphériques que ça devient vraiment suspect. » (Expert C).

Un autre renchérit : « Les incidents internes... ce sont des petits comportements répétés dans le temps qui finissent par former un schéma. » (Expert B).

Malgré cela, les outils actuels peinent à exploiter ces signaux. Les experts dénoncent un manque de contextualisation métier, un taux élevé de faux positifs et une forte dépendance à l'expertise humaine :

« Les SIEM sont bons pour les menaces externes, mais pour l'interne, on manque de contexte... les alertes sont trop bruyantes. » (Expert D) ;

« Les outils actuels ne comprennent pas le métier. Ils voient des logs, pas des intentions. » (Expert F).

Ces limites illustrent la nécessité d’approches plus intégrées, combinant données techniques, contexte organisationnel et trajectoires comportementales.

2.6 Thème : perception des jeux de données simulés et mise en perspective du CERT

Les experts convergent pour affirmer que les jeux de données simulés, et en particulier CERT, sont utiles pour l’expérimentation méthodologique, mais insuffisants pour représenter la complexité réelle des menaces internes. Comme l’exprime un participant : « Les datasets comme CERT sont utiles pour faire des démonstrations, mais ils ne reflètent absolument pas la réalité d’une organisation. » (Expert E).

Les critiques portent sur plusieurs aspects : scénarios trop stéréotypés, signaux faibles sous-représentés, granularité temporelle limitée, labels binaires, répartition artificielle des types d’événements. Un expert précise : « Ce qui manque le plus, c’est le contexte... un comportement dépend du rôle, de l’historique, des responsabilités. Les jeux simulés ignorent cette dimension. » (Expert A).

Ces constats expliquent pourquoi les modèles atteignent des performances très élevées dans le jeu de données CERT : les signaux y sont plus nets, moins bruités et plus homogènes que dans les environnements réels. La mise en perspective qualitative montre ainsi que le jeu de données CERT constitue un banc d’essai technique, mais un reflet imparfait de la complexité sociotechnique des organisations.

Chapitre 5. Discussion

Ce chapitre propose une mise en perspective intégrée des résultats obtenus en mobilisant à la fois l'analyse quantitative fondée sur l'approche fonctionnelle des données (FDA) et l'analyse qualitative issue des entrevues menées auprès d'experts en cybersécurité. L'objectif est de comprendre comment ces deux volets se complètent, se renforcent ou se contredisent, afin d'évaluer la capacité du jeu de données CERT à représenter des dynamiques réalistes de menace interne malveillante et à soutenir le développement de méthodes avancées de détection.

D'un côté, l'analyse quantitative a mis en évidence la forte capacité discriminante des trajectoires fonctionnelles dérivées des comportements utilisateurs. Le recours au lissage par B-splines, à la FPCA et à des modèles supervisés non linéaires, en particulier la forêt aléatoire, a permis d'obtenir des performances très élevées, parfois proches de la perfection, surtout dans la distinction entre utilisateurs malveillants et non malveillants. Cette performance reflète en partie la cohérence interne et la structure expérimentale contrôlée du jeu de données CERT, qui offre un environnement homogène, bien documenté et exempt des imprécisions courantes des journaux réels. Ces résultats suggèrent ainsi que le jeu de données CERT contient des signaux temporels suffisamment structurés pour que les méthodes fonctionnelles capturent efficacement les motifs associés aux comportements simulés.

De l'autre côté, l'analyse qualitative nuance et contextualise ces performances. Les entrevues réalisées à partir de scénarios réalistes ont mis en lumière la complexité des comportements malveillants réels, qui se déploient souvent de manière progressive, ambiguë et hautement contextualisée. Les experts ont souligné l'importance des signaux faibles, de la corrélation temporelle entre actions, du rôle occupé par l'utilisateur, du contexte organisationnel et des dynamiques psychologiques ou relationnelles, dimensions largement absentes ou simplifiées dans le jeu de données CERT. Ces constats permettent de mieux comprendre pourquoi les modèles obtiennent des performances aussi élevées : les comportements simulés y sont plus homogènes, plus visibles et moins ambigus que dans les organisations réelles.

Ainsi, l'apport combiné des analyses quantitative et qualitative permet d'identifier les principaux écarts entre les trajectoires simulées et les dynamiques réellement

observées. Plus encore, il révèle les conditions dans lesquelles les méthodes fonctionnelles sont particulièrement efficaces, mais aussi les limites structurelles des jeux de données simulés pour représenter le caractère multiforme, évolutif et contextuel des menaces internes malveillantes.

La discussion s'articule autour de deux volets complémentaires :

- Le premier examine les résultats quantitatifs à la lumière des caractéristiques du jeu de données CERT : structure temporelle, distribution des comportements, artifices de simulation et implications pour la performance des modèles fonctionnels.
- Le second volet analyse les résultats qualitatifs en profondeur, en montrant comment les perspectives des experts enrichissent, nuancent ou remettent en question les observations issues de la FDA, et en quoi elles éclairent les limites du jeu de données CERT pour la recherche et la pratique.

1. Discussion des résultats quantitatifs

L'analyse quantitative fondée sur l'approche fonctionnelle révèle une capacité discriminante particulièrement élevée pour la détection des comportements malveillants simulés dans le jeu de données CERT. Le lissage par B-splines, la réduction dimensionnelle via la FPCA et l'utilisation de modèles supervisés ont permis d'extraire des signaux temporels distinctifs. L'efficacité de ces traitements se reflète dans les performances très élevées obtenues avec la forêt aléatoire ($AUC = 0,997$; $F1 = 0,90$), accompagnées d'une variabilité presque nulle entre les partitions de validation croisée. Ces résultats illustrent la puissance de la représentation fonctionnelle pour capturer des motifs dynamiques, une propriété déjà mise en avant dans des travaux antérieurs en analyse comportementale (Ramsay & Silverman, 2005 ; Kokoszka & Reimherr, 2017). Cependant, la quasi-perfection observée appelle une interprétation prudente, car elle reflète largement les propriétés structurelles du jeu de données CERT plutôt qu'une performance nécessairement reproductible en contexte réel. Plusieurs études ont souligné que les jeux de données simulés, incluant les versions du CERT, tendent à

produire des comportements malveillants très contrastés et plus facilement détectables que dans la pratique (Eberle & Holder, 2009 ; Legg et al., 2015).

Premièrement, les scénarios malveillants du CERT présentent une signature comportementale homogène et peu ambiguë. Cette homogénéité réduit la variance intra-classe et amplifie les contrastes avec la classe non malveillante, ce qui facilite l'apprentissage supervisé. Plusieurs auteurs (Magklaras & Furnell, 2002 ; Greitzer et al., 2012) insistent pourtant sur le fait que, dans les environnements réels, les menaces internes émergent souvent de comportements graduels, indirects et ambigus, une complexité absente des scénarios du CERT.

Deuxièmement, la temporalité des comportements simulés est artificiellement nette. Comme l'ont remarqué Brown et al. (2017), les scénarios du CERT présentent des fenêtres d'activité concentrées et répétitives qui ne reflètent pas la variabilité journalière et hebdomadaire observée dans les logs réels. Cette structuration artificielle rend les motifs temporels particulièrement accessibles à la FPCA, qui excelle dans la capture de variations locales lorsqu'elles sont fortement corrélées à un label.

Troisièmement, le jeu de données CERT manque de bruit contextuel. Les travaux de Brdiczka et al. (2012) et de Nurse et al. (2014) soulignent que l'interprétation d'un même signal (connexion tardive, copie de fichiers, changement de répertoire) dépend fortement du rôle, de la charge de travail, des pratiques d'équipe et du contexte organisationnel. Dans le jeu de données CERT, l'absence de ces dimensions produit des trajectoires épurées et artificiellement discriminantes.

Quatrièmement, le jeu de données CERT repose sur un échantillonnage temporel hebdomadaire qui simplifie fortement la structure des séries. Contrairement aux logs réels générés en continu et marqués par des fluctuations fines, du bruit opérationnel et des irrégularités humaines, le jeu de données CERT produit des trajectoires régulières, peu bruitées et parfaitement alignées avec les attentes des modèles fonctionnels.

À ces éléments s'ajoutent des limites liées à l'échantillonnage temporel. Contrairement aux journaux réels, générés en continu, hétérogènes et souvent bruités (Salem & Stolfo, 2010), le jeu de données CERT repose sur une simulation à granularité hebdomadaire dont les périodes d'activité sont déterminées par la logique de génération des scénarios. Cette simplification temporelle diminue la complexité statistique et rend les modèles plus performants qu'ils ne le seraient face à des données réelles.

La FPCA joue un rôle central dans cette dynamique. La première composante reflète principalement l'intensité globale, dimension déjà connue comme discriminante dans plusieurs études sur la menace interne malveillante (Eberle, 2010). Les composantes supérieures, quant à elles, capturent des motifs temporels plus fins, mais, comme le rappellent Legg et al. (2015), ces motifs sont plus réguliers et répétitifs dans le jeu de données CERT que dans les logs organisationnels authentiques.

Ainsi, bien que l'approche fonctionnelle s'avère méthodologiquement pertinente pour l'analyse de données comportementales, les performances observées doivent être interprétées à la lumière des limites inhérentes aux jeux simulés : signatures stéréotypées, faible ambiguïté, absence de bruit organisationnel, distribution irréaliste des événements et granularité temporelle simplifiée. Ces éléments sont largement documentés dans la littérature, notamment dans les travaux évaluant la représentativité des données du CERT pour la recherche appliquée.

En somme, les résultats quantitatifs confirment la capacité de la FDA et des modèles supervisés à exploiter efficacement des dynamiques temporelles, mais révèlent aussi que la structure même du jeu de données CERT amplifie artificiellement ces performances. Comme l'indiquent plusieurs études (Greitzer et al., 2019 ; Althebyan & Panda, 2007), la transposition à des environnements opérationnels exigerait des données plus riches, contextualisées et bruitées, dimensions explorées en complément dans l'analyse qualitative.

2. Discussion des résultats qualitatifs

Les résultats qualitatifs enrichissent et nuancent de manière substantielle l'interprétation des observations quantitatives. Grâce à une démarche de codification hybride et à un dispositif d'entrevue fondé sur des scénarios réalistes, les experts fournissent un éclairage essentiel sur les mécanismes comportementaux associés à la menace interne malveillante ainsi que sur les limites structurelles des jeux de données simulés tels que le jeu de données CERT. Cette triangulation méthodologique s'inscrit dans les recommandations de la littérature, qui souligne l'importance de combiner approches quantitatives et qualitatives pour saisir la complexité multidimensionnelle des menaces internes (Greitzer & Hohimer, 2011 ; Nurse et al., 2014).

Un premier apport majeur concerne la dynamique progressive et ambiguë de la menace interne malveillante. Alors que le jeu de données CERT modélise des comportements malveillants nets, délimités et concentrés dans des fenêtres temporelles précises, les experts décrivent des trajectoires beaucoup plus nuancées, souvent initiées par des gestes ambigus qui ne deviennent malveillants qu'en fonction du contexte, des pressions psychologiques ou des interactions organisationnelles. Cette conception graduelle et ambivalente rejoint les travaux de Shaw et Sellers (2015), qui insistent sur l'existence d'une vaste « zone grise » entre négligence, erreur, non-conformité et intention malveillante. L'absence de cette ambiguïté dans le jeu de données CERT renforce artificiellement la visibilité des comportements problématiques, ce qui facilite leur détection algorithmique et contribue à expliquer les performances élevées observées dans l'analyse quantitative.

Un second apport déterminant concerne le rôle central du contexte dans l'interprétation des signaux. Les experts rappellent que l'analyse d'un comportement doit tenir compte du rôle professionnel, des responsabilités, de la charge de travail, des habitudes d'équipe ou encore de l'historique individuel. Ces facteurs constituent des éléments clés dans les modèles conceptuels de menace interne malveillante (Cappelli, Moore & Trzeciak, 2012 ; Legg et al., 2015). Or le jeu de données CERT ne contient aucune de ces dimensions : les utilisateurs y sont interchangeable, dépourvus d'identité fonctionnelle et de trajectoire personnelle. Cette absence de contextualisation limite fortement la portée interprétative des modèles issus de la FDA et compromet la transférabilité des résultats à des situations réelles.

Les entrevues mettent également en lumière la complexité des signaux faibles et des corrélations temporelles mobilisées par les analystes dans les pratiques réelles de détection. Dans les organisations, les comportements sont évalués non seulement en fonction de leur occurrence isolée, mais aussi en fonction de leur séquence, de leur durée, de leur intensité relative et de leur synchronisation, dimensions que la littérature associe à la détection de patterns narratifs ou séquentiels caractéristiques de la malveillance (Legg et al., 2015 ; Eberz et al., 2020). À l'inverse, les signaux simulés dans le jeu de données CERT demeurent simples, isolés et faiblement dépendants du contexte, ce qui les rend particulièrement accessibles aux modèles supervisés, en particulier lorsqu'ils reposent sur des représentations fonctionnelles. Cette différence

structurelle permet de comprendre pourquoi les modèles parviennent à identifier les utilisateurs malveillants dans le jeu de données CERT avec une telle efficacité, alors que la détection dans les environnements réels demeure un défi considérable.

Ces constats ouvrent néanmoins une perspective importante pour l'évolution des jeux de données simulés. Plusieurs experts suggèrent que certaines limites du jeu de données CERT pourraient être atténuées par l'intégration de techniques d'intelligence artificielle générative. La littérature récente explore déjà l'utilisation de modèles probabilistes, de réseaux génératifs avancés et d'architectures séquentielles avancées, notamment de type Transformer ou diffusion, pour produire des logs synthétiques capturant plus fidèlement la variabilité humaine, les dépendances temporelles complexes et les fluctuations organisationnelles caractéristiques des milieux réels (Brown et al., 2022). De telles approches permettraient de simuler des comportements comprenant du bruit organisationnel, des transitions progressives entre normalité et malveillance, ainsi que des patterns temporels plus fins et plus crédibles que ceux produits par des scénarios écrits manuellement. Elles soulèvent néanmoins des défis méthodologiques importants, notamment le risque d'introduire de nouveaux biais, la difficulté de maintenir une cohérence organisationnelle globale et la nécessité de faire valider la plausibilité des séquences générées par des experts humains (Tuor et al., 2017).

Enfin, les experts convergent vers plusieurs critiques récurrentes du jeu de données CERT : simplification excessive des scénarios, absence de signaux faibles, faible granularité temporelle, labellisation strictement binaire, déséquilibre des comportements et diversité limitée des profils utilisateurs. Ces limites rejoignent les constats de la littérature, qui souligne que les jeux de données simulés peinent à reproduire la complexité sociotechnique des environnements réels (Eberz et al., 2020). Ces observations expliquent en partie pourquoi les modèles quantitatifs obtiennent des performances exceptionnellement élevées : ils apprennent sur un environnement artificiellement régulier, propre et faiblement ambigu, éloigné des conditions opérationnelles où la malveillance est rare, contextuelle et profondément humaine. Ainsi, l'analyse qualitative ne contredit pas les résultats obtenus par la FDA ; elle en précise les conditions de validité, en identifiant les éléments structurels qui amplifient artificiellement la performance des modèles dans le jeu de données CERT. Elle montre

que si le jeu de données constitue un excellent terrain d'expérimentation méthodologique, il demeure un reflet partiel et simplifié de la réalité organisationnelle. L'apport qualitatif apparaît donc indispensable pour interpréter correctement les résultats quantitatifs et pour orienter la conception de jeux de données et de modèles plus représentatifs des dynamiques réelles associées aux menaces internes.

Chapitre 6. Conclusion

Ce mémoire avait pour objectif d'évaluer dans quelle mesure le jeu de données CERT r4.2 permet de représenter de façon réaliste les dynamiques associées aux menaces internes malveillantes, ainsi que d'examiner la pertinence des approches d'analyse fonctionnelle des données (FDA) pour la détection proactive de comportements suspects. En s'appuyant sur une méthodologie mixte intégrant (1) une analyse quantitative approfondie des trajectoires comportementales et des performances de modèles supervisés (S.O.1), et (2) une analyse qualitative fondée sur des entrevues d'experts en cybersécurité (S.O.2), ce travail a permis d'identifier certains apports méthodologiques du jeu CERT, mais aussi les limites structurelles susceptibles d'affecter sa représentativité organisationnelle.

L'intégration des deux volets d'analyse a contribué à dégager une compréhension plus intégrée de la manière dont le jeu de données CERT soutient, ou contraint, la détection proactive. D'un côté, les résultats quantitatifs expliquent pourquoi certains modèles atteignent des performances très élevées dans le contexte étudié ; de l'autre, les constats qualitatifs montrent en quoi ces performances sont difficilement transférables aux environnements réels. Cette double perspective contribue à répondre aux objectifs S.O.3 et S.O.4, en évaluant les forces et limites du jeu de données CERT et en identifiant les améliorations nécessaires pour en accroître la valeur en pratique.

1. Retour sur la question de recherche et de l'approche méthodologique

La question centrale de ce mémoire était :

Comment le jeu de données CERT peut-il être amélioré afin de mieux soutenir la détection proactive des menaces internes malveillantes ?

Pour y répondre, le mémoire a suivi une démarche en plusieurs étapes complémentaires, chacune associée à un sous-objectif précis (S.O.1 à S.O.4), permettant de croiser les résultats quantitatifs, les constats qualitatifs et l'analyse critique de la littérature.

La première étape a consisté à préparer, lisser et transformer les données du CERT r4.2 en trajectoires temporelles hebdomadaires à l'aide de l'analyse fonctionnelle des données. Les résultats suggèrent que, dans le contexte spécifique du jeu CERT, les approches fondées sur l'analyse fonctionnelle des données combinées à des modèles supervisés, en particulier la forêt aléatoire, atteignent des performances élevées. Ces performances s'expliquent en partie par la capacité de la FDA à capter les dynamiques temporelles globales des comportements utilisateurs et à produire des représentations compactes facilitant l'apprentissage supervisé. Elles indiquent que certaines méthodes apparaissent efficaces dans le contexte du jeu de données CERT, lorsque les trajectoires comportementales présentent une structure régulière et des contrastes marqués entre comportements bénins et malveillants.

La seconde étape, correspondant au S.O.2, a porté sur la réalisation d'entrevues qualitatives auprès de professionnels de la cybersécurité. Cette phase avait pour objectif de comparer les comportements malveillants réels observés en organisation avec ceux simulés dans le jeu de données CERT. Les experts ont mis en évidence plusieurs écarts fondamentaux : la malveillance réelle est graduelle, contextuelle et marquée par l'ambiguïté, alors que le jeu de données CERT présente des trajectoires fortement scénarisées, peu ambiguës et dénuées de contexte métier. Cette comparaison suggère certaines implications importantes pour l'interprétation des résultats quantitatifs. D'une part, elle montre une limite importante du jeu CERT, en suggérant que les performances élevées observées pourraient être en partie liées à des scénarios simplifiés et ne peuvent être directement transposées à des environnements organisationnels réels sans validation supplémentaire. D'autre part, elle confère une valeur explicative aux résultats obtenus, en permettant d'identifier précisément les conditions sous lesquelles les modèles sont performants. Cette étape qualitative joue ainsi un rôle important dans la compréhension des écarts entre performance algorithmique et plausibilité organisationnelle, et justifie la nécessité d'enrichir les jeux de données utilisés pour la recherche en détection des menaces internes malveillantes. La troisième étape, associée au S.O.3, a consisté en une analyse transversale de la représentativité, des forces et des limites du jeu de données CERT, éclairée à la fois par les résultats empiriques et par la littérature scientifique. Cette analyse critique suggère que le jeu de données CERT peut constituer un outil méthodologiquement pertinent pour

le développement conceptuel et méthodologique, mais qu'il ne représente qu'imparfaitement les dynamiques humaines, organisationnelles et relationnelles présentes dans les incidents réels de menace interne malveillante. En particulier, l'absence de bruit organisationnel, la granularité temporelle hebdomadaire, la faible diversité comportementale et la définition scénarisée des incidents limitent sa capacité à soutenir une détection réellement proactive dans certains contextes.

Enfin, la quatrième étape, en lien direct avec S.O.4, a permis de formuler des recommandations pour améliorer le jeu de données CERT. Afin d'en offrir une synthèse structurée et directement exploitable, le tableau 6.1 présente de manière intégrée les principales limites identifiées ainsi que les recommandations correspondantes.

Ces pistes d'amélioration portent notamment sur l'ajout d'éléments contextuels, l'enrichissement des dynamiques comportementales, l'augmentation de la granularité temporelle et l'introduction de mécanismes de simulation plus réalistes. Le tableau 6.1 détaille chacune de ces recommandations en les reliant explicitement aux limites observées dans CERT r4.2.

Tableau 6-1 : Recommandations structurées pour améliorer la représentativité du jeu de données CERT

Axe d'amélioration	Recommandations proposées	Limites observées dans CERT r4.2
Amélioration du contenu des données	Introduire une granularité journalière ou intra-journalière afin de refléter la variabilité réelle des journaux d'activité.	Données hebdomadaires trop grossières, limitant l'observation des dynamiques fines.
	Ajouter des attributs métier tels que la fonction, l'équipe, les responsabilités, l'ancienneté et les horaires de travail.	Absence de rôles professionnels, d'horaires et de contexte organisationnel.

	Intégrer des indicateurs de charge de travail, d'absences et de déplacements professionnels.	Environnement trop homogène et déconnecté des contraintes organisationnelles.
Amélioration de la simulation des comportements	Simuler des séquences comportementales ambiguës, graduelles et intermittentes, incluant erreurs, négligence et variations comportementales.	Comportements malveillants trop visibles et fortement contrastés.
	Introduire une variabilité interindividuelle et inter-scénarios plus marquée.	Utilisateurs homogènes et stratégies malveillantes similaires.
	Développer des scénarios stochastiques ou inspirés de cas réels plutôt que des scénarios fixes et narratifs.	Scénarios déterministes et peu variés.
	Simuler des transitions progressives entre normalité, zone grise et malveillance explicite.	Passage abrupt à la malveillance.
	Ajouter des interactions entre utilisateurs (collaboration, échanges de courriels, tensions hiérarchiques).	Absence de dynamique relationnelle entre utilisateurs.
Amélioration des conditions d'analyse et d'expérimentation	Utiliser des modèles probabilistes, séquentiels ou génératifs pour produire des	Simulations déterministes peu adaptées à l'étude des signaux faibles.

	trajectoires comportementales plus réalistes.	
	Introduire du bruit organisationnel réaliste (fluctuations de charge, erreurs système, interruptions).	Environnement trop propre et absence de perturbations.
	Combiner l'analyse fonctionnelle des données avec des approches multimodales (graphes comportementaux, modèles relationnels).	Limitation des analyses unidimensionnelles centrées uniquement sur la temporalité.

Ainsi, les analyses quantitatives et qualitatives convergent vers une conclusion centrale : le jeu de données CERT, dans sa forme actuelle, constitue un outil d'expérimentation technique utile, mais un reflet partiel de la menace interne malveillante telle qu'elle se manifeste en contexte organisationnel réel. Sa capacité à soutenir la détection proactive dépendra de sa transformation en un jeu de données plus riche, plus contextuel et plus varié. Les résultats suggèrent que l'ajout de dimensions organisationnelles et comportementales pourrait contribuer à mieux soutenir la détection proactive des menaces internes malveillantes, en permettant de mieux représenter la variabilité, l'ambiguïté et la dynamique graduelle qui caractérisent ces phénomènes en contexte opérationnels.

2. Implications pour les professionnels

Les résultats de ce mémoire offrent plusieurs orientations pratiques pour les équipes de sécurité et les organisations.

Premièrement, les performances élevées observées sur le jeu de données CERT ne doivent pas être interprétées comme indicatives de l'efficacité des modèles en contexte

réel. Les signaux comportementaux rencontrés dans les organisations sont plus diffus, moins visibles et profondément dépendants du rôle, de l'historique et du contexte professionnel. Toute implantation opérationnelle d'outils fondés sur la FDA ou sur des modèles supervisés doit donc s'accompagner d'une validation sur des données internes ou sur des environnements hybrides.

Deuxièmement, l'interprétation des signaux techniques ne peut être dissociée du contexte organisationnel. Une connexion tardive, une copie de fichier ou une hausse soudaine d'activité n'ont de sens que recontextualisées dans la trajectoire professionnelle de l'utilisateur. Les résultats qualitatifs montrent que les professionnels doivent intégrer des dimensions humaines, relationnelles et organisationnelles dans leurs processus de surveillance.

Enfin, la détection proactive des menaces internes malveillantes s'appuie sur la compréhension fine des comportements ambigus, cette zone grise où les gestes fautifs, négligents ou à risque coexistent. Les outils analytiques doivent être considérés comme un soutien au jugement humain plutôt que comme des mécanismes prescriptifs, afin d'éviter les interprétations erronées et la surcharge d'alertes.

3. Implications pour la recherche

Sur le plan scientifique, ce mémoire met en évidence la pertinence du jeu de données CERT r4.2 comme terrain d'expérimentation méthodologique pour l'évaluation et la comparaison de méthodes avancées de détection des menaces internes malveillantes. Plusieurs travaux antérieurs avaient déjà souligné le potentiel du jeu de données CERT comme banc d'essai méthodologique et ont appelé à un développement accru d'approches capables de tirer parti de sa structure temporelle (Glasser & Lindauer, 2013 ; Eberle & Holder, 2009). Les résultats empiriques obtenus, dans cette recherche, répondent directement à ces appels : l'analyse fonctionnelle des données (FDA) et la FPCA indiquent leur capacité à extraire des motifs longitudinaux complexes et à soutenir la détection de comportements malveillants simulés. Ainsi, ce mémoire tend à apporter une contribution méthodologique qui prolonge de manière concrète les orientations proposées par ces travaux antérieurs. Ces résultats permettent d'éclairer certaines conditions méthodologiques sous lesquelles la FDA apparaît performante, en montrant

qu'elle dépend fortement du contraste comportemental inhérent aux données synthétiques.

Toutefois, un courant tout aussi significatif de la littérature met en garde contre une utilisation non critique du jeu CERT et insiste sur ses limites en matière de réalisme, de variabilité comportementale et de complexité sociotechnique (Cappelli et al., 2012 ; Greitzer & Frincke, 2010 ; Homoliak et al., 2019). Ces auteurs appellent explicitement à enrichir les environnements d'évaluation ou à bâtir des jeux de données plus représentatifs des dynamiques organisationnelles. Ce mémoire s'inscrit dans ces appels formulés dans la littérature en montrant empiriquement que les performances élevées observées sur le jeu de données CERT semblent en grande partie liées à ses propriétés synthétiques, scénarios peu ambigus, signaux fortement contrastés, absence de bruit organisationnel et faible variabilité interindividuelle, et non sur une capacité intrinsèque des modèles à généraliser à des environnements réels.

Dans cette perspective, plusieurs travaux insistent sur la nécessité de développer des jeux de données plus riches et mieux contextualisés, notamment en combinant données simulées, données réelles anonymisées et séquences générées par des modèles probabilistes ou génératifs. De tels environnements hybrides permettraient de reproduire une variabilité comportementale plus proche des organisations réelles, incluant du bruit opérationnel, des transitions progressives entre normalité et malveillance et des ambiguïtés intentionnelles caractéristiques des comportements humains. L'intégration de ces éléments apparaît nécessaire pour dépasser les limites actuelles du CERT et offrir aux chercheurs des bases d'évaluation plus robustes et plus représentatives.

L'intégration d'une analyse qualitative constitue par ailleurs une contribution distinctive à la recherche. Plusieurs auteurs appelaient à inclure davantage la perspective humaine et organisationnelle dans l'évaluation des approches de détection (Shaw & Fischer, 2005 ; Greitzer et al., 2014), mais peu d'études avaient articulé une comparaison explicite entre les dynamiques simulées de CERT et celles observées sur le terrain. En révélant les écarts majeurs liés à l'ambiguïté comportementale, au contexte métier ou encore aux trajectoires progressives qui caractérisent les menaces internes malveillantes, ce mémoire fournit un éclairage supplémentaire sur les conditions qui expliquent la différence entre performances théoriques et applicabilité réelle des modèles.

Les résultats convergent également avec les recommandations récentes invitant à développer des jeux de données enrichis, hybrides ou génératifs, capables de reproduire des comportements plus réalistes, incluant fluctuations, transitions graduelles, signalements ambigus et interactions sociales (Nasir et al., 2021 ; Pantelidis et al., 2021 ; Li et al., 2019). Dans ce contexte, les recommandations proposées dans ce mémoire, granularité temporelle plus fine, contextualisation métier, diversification comportementale, variabilité interindividuelle, intégration de signaux faibles, recours à des modèles génératifs, constituent une réponse pouvant contribuer à répondre à ces appels. Elles offrent aux chercheurs des orientations structurées pour faire évoluer le jeu de données CERT dans un sens souhaité par la communauté scientifique.

Ces recommandations pourraient soutenir la recherche de plusieurs manières :

- Améliorer la qualité et la pertinence des environnements expérimentaux. En rendant les données plus réalistes et plus variées, elles facilitent l'évaluation de la robustesse des modèles et évitent de tirer des conclusions fondées sur des performances artificiellement élevées.
- Fournir un cadre opérationnel pour concevoir les futures simulations. Les chercheurs disposent ainsi de critères précis pour développer ou enrichir des jeux de données, ce qui peut accélérer la création d'ensembles hybrides ou générés par des modèles probabilistes.
- Contribuer à renforcer le lien entre les approches computationnelles et la réalité organisationnelle, un besoin largement identifié dans la littérature mais encore rarement adressé de manière empirique.

Enfin, ce mémoire souligne la valeur des approches méthodologiques mixtes dans l'étude des menaces internes malveillantes. En articulant résultats quantitatifs et apports qualitatifs, il montre qu'une compréhension rigoureuse de la détection nécessite d'intégrer les dimensions humaines, technique et organisationnelle. Cette contribution aide à dépasser une vision strictement algorithmique de la menace interne malveillante et rappelle que la validité scientifique des modèles dépend étroitement de la représentativité des jeux de données utilisés.

4. Limites

Comme toute recherche, ce mémoire comporte un ensemble de limites méthodologiques et empiriques qui doivent être reconnues pour interpréter adéquatement les résultats.

La première limite découle de la dépendance au jeu de données CERT r4.2. Sa nature synthétique et scénarisée influence directement les résultats, en raison notamment de la forte visibilité des comportements malveillants et de l'absence de bruit organisationnel. Cette particularité explique en partie les performances très élevées obtenues par les modèles fonctionnels et limite la transférabilité des conclusions vers des environnements réels, caractérisés par une ambiguïté beaucoup plus marquée.

Une deuxième limite tient au nombre restreint d'experts interrogés dans le cadre de l'étude qualitative. Bien que leur diversité professionnelle ait permis de dégager des tendances robustes, un échantillon plus large aurait enrichi la compréhension des mécanismes comportementaux, des signaux faibles mobilisés en pratique et des variations propres à différents milieux organisationnels.

Une troisième limite touche directement la nature simulée des scénarios utilisés dans la phase qualitative. Faute d'accès à de véritables journaux organisationnels, dont la diffusion est extrêmement restreinte pour des raisons de confidentialité, de sécurité et de conformité réglementaire, il a été nécessaire de recourir à des scénarios construits à partir de la littérature et d'exemples plausibles. Bien que ces scénarios soient inspirés de situations réalistes, ils ne reproduisent pas la granularité, la complexité contextuelle, le volume et l'hétérogénéité des comportements observés dans les logs réels. Les jugements des experts portent donc sur des situations reconstruites plutôt que sur des traces brutes, ce qui réduit la capacité de l'étude à refléter fidèlement la dynamique des menaces internes malveillantes telles qu'elles se manifestent dans les organisations.

Une quatrième limite concerne l'impossibilité de valider empiriquement les modèles développés sur des données organisationnelles réelles, en raison des contraintes associées à la confidentialité, à la sécurité de l'information et à l'accès aux journaux internes. Cette limite est courante dans les travaux portant sur les menaces internes, mais elle empêche néanmoins d'évaluer la performance réelle des approches fonctionnelles sur des données authentiques.

Enfin, bien que la FPCA et les modèles supervisés se soient montrés particulièrement performants sur CERT, leur efficacité réelle est susceptible de dépendre de la qualité, de la richesse et de la granularité des journaux disponibles. Dans des environnements réels, les signaux peuvent être beaucoup plus fragmentés, sporadiques ou bruités, ce qui pourrait réduire substantiellement la capacité des modèles à identifier des comportements suspects.

5. Avenue de recherche

À partir des constats empiriques et théoriques formulés dans ce mémoire, plusieurs pistes de recherche se dessinent pour améliorer la représentativité des jeux de données dédiés à la détection proactive des menaces internes et, plus largement, pour renforcer la validité méthodologique des études dans ce domaine.

Avenue de recherche 1

Une première avenue consiste à développer des jeux de données hybrides combinant des données simulées, des données générées par des modèles d'intelligence artificielle et des données réelles anonymisées. Une telle approche permettrait de dépasser les contraintes inhérentes aux journaux organisationnels, souvent inaccessibles pour des raisons de confidentialité, tout en intégrant une variabilité comportementale plus proche des environnements réels. Les résultats de ce mémoire ont montré que la structure trop homogène et les scénarios peu ambigus du CERT limitent fortement la capacité des modèles à se généraliser ; l'hybridation des sources permettrait de pallier ces limites en introduisant du bruit, de la diversité interindividuelle et des fluctuations contextuelles essentielles à la construction de trajectoires réalistes.

Cette avenue s'inscrit directement dans les recommandations proposées dans ce mémoire, en offrant un cadre dans lequel chacune d'elles pourrait être concrètement mise en œuvre. En effet, l'intégration de données réelles permettrait de tester l'effet d'une granularité temporelle plus fine, tandis que l'ajout de métadonnées organisationnelles anonymisées faciliterait l'enrichissement contextuel recommandé, notamment en ce qui concerne les rôles professionnels, les horaires ou les responsabilités. De même, la variabilité interindividuelle et la diversification

comportementale identifiées comme nécessaires pourraient être obtenues en entraînant des modèles probabilistes ou génératifs sur des exemples issus du terrain. L'intégration de données provenant de sources diverses offrirait également la possibilité d'introduire des signaux faibles, des ambiguïtés intentionnelles et des fluctuations organisationnelles qui ne peuvent être reproduites à partir de données strictement simulées. Enfin, le recours à des modèles génératifs, également recommandé, trouverait dans un jeu hybride un terrain d'application idéal pour produire des trajectoires plus réalistes, en reproduisant des transitions progressives et des comportements plus nuancés.

Ainsi, cette première avenue ne constitue pas seulement une piste théorique pour enrichir les données disponibles : elle représente un moyen concret pour les chercheurs de mobiliser, tester et articuler l'ensemble des recommandations formulées dans ce mémoire, en vue de construire des environnements expérimentaux plus fidèles aux dynamiques organisationnelles réelles et mieux adaptés à l'évaluation de méthodes avancées de détection.

Avenue de recherche 2

Une deuxième piste de recherche porte sur l'intégration de modèles génératifs, en particulier les architectures séquentielles avancées, les modèles probabilistes ou les modèles de diffusion. Ces techniques pourraient être mobilisées pour générer des logs synthétiques reproduisant plus fidèlement les dynamiques temporelles, les transitions graduelles entre comportements normaux et malveillants et les signaux faibles identifiés par les experts comme centraux dans la détection des menaces internes. Les résultats qualitatifs ont en effet montré que la malveillance réelle se manifeste rarement sous la forme d'épisodes abrupts, mais plutôt comme une progression contextuelle marquée par des hésitations, des fluctuations et des ambiguïtés intentionnelles. Les modèles génératifs sont particulièrement bien adaptés pour intégrer ces nuances dans les simulations, en s'éloignant des scénarios rigides et déterministes qui caractérisent encore les versions actuelles du jeu CERT.

Dans cette perspective, plusieurs travaux récents ont démontré le potentiel des modèles génératifs pour produire des données comportementales ou des séquences temporelles réalistes. Les GAN appliqués aux comportements utilisateurs et aux environnements de

cybersécurité (Rigaki & Garcia, 2018 ; Bogoiavlenski et al., 2021) ont montré leur capacité à reproduire ou altérer des séquences selon des dynamiques progressives, confirmant leur utilité pour simuler des comportements nuancés. De même, les modèles variationnels et probabilistes utilisés pour la modélisation comportementale (Chen et al., 2020) ainsi que les modèles de diffusion appliqués aux séries temporelles (Rasul et al., 2021) offrent des perspectives particulièrement prometteuses pour injecter du bruit organisationnel, des transitions graduelles et des ambiguïtés intentionnelles dans les simulations.

L'utilisation de ces approches permettrait donc de s'éloigner des scénarios rigides et déterministes qui caractérisent encore les versions actuelles du CERT et d'approcher davantage la complexité dynamique observée en milieu organisationnel. Elle ouvrirait également la voie à la génération de jeux de données mieux adaptés à l'analyse fonctionnelle et aux modèles supervisés, en créant des trajectoires plus réalistes, hétérogènes et progressives, reflétant les nuances comportementales propres aux menaces internes malveillantes.

Avenue de recherche 3

Une troisième avenue réside dans l'exploration de la contextualisation avancée des signaux comportementaux, un élément central pour rapprocher les simulations de la réalité organisationnelle. Les analyses menées dans ce mémoire ont montré que l'absence de rôles professionnels, d'horaires de travail, de charges variables ou d'événements organisationnels dans CERT conduit à une représentation trop décontextualisée des comportements simulés. Or, comme l'ont souligné les experts interrogés, la compréhension des signaux repose largement sur leur inscription dans un contexte professionnel précis.

Pour opérationnaliser cette contextualisation, plusieurs options peuvent être envisagées. Une première consiste à associer à chaque utilisateur un profil métier détaillé, incluant la fonction, l'ancienneté, la position hiérarchique, les responsabilités quotidiennes et les contraintes propres à son rôle. Ces métadonnées permettraient de simuler des comportements différenciés selon les individus, par exemple une variabilité normale plus élevée chez certains postes (administrateurs, développeurs, analystes) ou des périodes d'activité corrélées aux calendriers professionnels.

Une deuxième option porte sur l'intégration d'une journalisation fine des comportements d'accès, reflétant plus fidèlement les pratiques observées dans les organisations. Cela inclurait la fréquence d'accès aux fichiers, la nature des ressources consultées, les écarts par rapport aux habitudes quotidiennes ou encore la survenue de comportements atypiques, tels que la consultation répétée de documents sensibles que l'utilisateur n'ouvre habituellement pas. Ces dimensions sont au cœur des analyses comportementales utilisées en pratique, notamment dans les solutions de type UEBA.

Une troisième option vise à introduire des routines temporelles réalistes, telles que des horaires individualisés, des cycles de charge de travail, des périodes de pointe, des congés ou des déplacements professionnels. De tels éléments permettraient de modéliser des déviations temporelles réellement suspectes plutôt que de simples variations par rapport à une norme uniforme, renforçant ainsi les capacités des modèles à détecter des anomalies propres à chaque individu.

Enfin, il serait pertinent d'intégrer des événements organisationnels dynamiques, comme des réorganisations internes, des changements d'équipe, des tensions hiérarchiques, des audits ou des lancements de projet. Associés à des indicateurs de motivation ou de frustration organisationnelle – tels que ceux suivis par les gestionnaires ou détectés par les outils UEBA (insatisfaction, réaffectations, conflits internes, baisse de performance) – ces éléments permettraient de simuler les fluctuations psychosociales qui précèdent souvent l'apparition de comportements à risque. Leur modélisation renforcerait considérablement la capacité des jeux de données simulés à reproduire les conditions réelles dans lesquelles émergent les menaces internes malveillantes.

En réunissant ces différentes dimensions, les simulations futures pourraient non seulement refléter la complexité des environnements où émergent réellement les menaces internes, mais également soutenir le développement de modèles centrés sur les trajectoires individuelles. Une telle contextualisation permettrait de détecter des anomalies propres à chaque utilisateur, plutôt que de s'appuyer sur une norme globale, et constituerait ainsi un fondement essentiel pour une détection proactive véritablement sensible aux variations comportementales fines.

Avenue de recherche 4

Enfin, un champ de recherche prometteur concerne la combinaison de l'analyse fonctionnelle des données avec d'autres approches multimodales, notamment les graphes comportementaux et les modèles relationnels. Les résultats de ce mémoire ont montré que la FDA capture efficacement les dynamiques temporelles, mais qu'elle ne saisit pas entièrement les dimensions relationnelles ou structurelles qui caractérisent les interactions en milieu organisationnel. Des travaux récents en détection de menaces internes soulignent d'ailleurs l'importance de modèles fondés sur les graphes pour représenter explicitement les relations entre utilisateurs, ressources et actions, permettant ainsi de détecter des motifs organisationnels complexes et difficilement perceptibles dans les données linéaires (Yuan, Sun & Liu, 2023).

La fusion de représentations fonctionnelles avec des structures de graphes, intégrant par exemple les réseaux de collaboration, les accès partagés à des ressources ou les interactions entre sous-systèmes, offrirait une compréhension plus intégrée de la menace interne malveillante. Une telle approche permettrait de modéliser simultanément la temporalité, la structure sociale et les comportements techniques, afin de détecter des motifs complexes impossibles à saisir par une seule modalité.

Dans l'ensemble, ces avenues de recherche convergent vers un objectif commun : développer des environnements de simulation et des cadres analytiques capables de refléter la complexité, l'ambiguïté et la variabilité des comportements humains en organisation. Ce sont ces caractéristiques, absentes des jeux de données actuels comme CERT, qui conditionnent la capacité réelle des modèles à soutenir une détection proactive pertinente, robuste et transférable.

Bibliographie

- Abu-Nimeh, S., Korman, S., & Nappa, A. (2007). The CERT Insider Threat Dataset: A tool for cybersecurity research. Proceedings of the International Workshop on Cyber Security and Privacy.
- Baracaldo, N., Chen, B., Ludwig, H., & Safavi, S. (2021). Detecting anomalous insider behavior in enterprise environments. IBM Research Report.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2).
- Breiman, L. (2001). Random forests. *Machine Learning*.
- Brdiczka, O., Liu, J., Price, B., Shen, J., Paterson, B., & Bartley, N. (2012). Proactive insider threat detection through graph learning and psychological context. *IEEE Security & Privacy Workshops*.
- Brown, G., Gasser, L., & Kuchar, O. (2017). Scenario generation in insider threat modeling: Challenges and opportunities. *Journal of Cybersecurity*, 3(2), 121–137.
- Cappelli, D. M., Moore, A. P., & Trzeciak, R. F. (2012). The CERT guide to insider threats: How to prevent, detect, and respond to information technology crimes. Addison-Wesley.
- Cappelli, D. M., Moore, A. P., & Trzeciak, R. F. (2012). Insider threat patterns. *IEEE Security & Privacy*.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD Conference*.
- Cybersecurity and Infrastructure Security Agency (CISA). (2023). Defining insider threat. U.S. Department of Homeland Security (<https://www.cisa.gov>).
- Cybersecurity and Infrastructure Security Agency (CISA). (2023). Insider threat risk mitigation guide. U.S. Department of Homeland Security (<https://www.cisa.gov/insider-threat>).
- CMU Software Engineering Institute. (2021). Insider threat dataset documentation (CERT r4.2). Carnegie Mellon University.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297 (<https://doi.org/10.1007/BF00994018>).
- DARPA. (2012). Anomaly Detection at Multiple Scales (ADAMS).
- Dai, X., & Müller, H.-G. (2018). Principal component analysis for functional data. *Journal of Multivariate Analysis*.
- Eberle, W. (2010). Discovering abnormal behavior in graph-based data (Doctoral dissertation, University of Tennessee).
- Eberle, W., & Holder, L. (2009). Detecting insider threat through graph-based anomaly detection. *IEEE Security & Privacy Workshops*.
- Eberz, S., Rasmussen, K., Lenders, V., & Martinovic, I. (2020). Evaluating behavioral biometrics for continuous authentication. *ACM Computing Surveys*, 53(2).
- ENISA. (2022). Insider threats in the EU: Trends and best practices.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters with noise. *Proceedings of the KDD Conference*.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer.
- Gamachchi, H., Sun, Y., & Boztas, S. (2017). A graph-based framework for malicious insider threat detection. *Hawaii International Conference on System Sciences (HICSS)*.
- Gartner. (2015). User and entity behavior analytics (UEBA): Market guide. Gartner Research.
- Gertheiss, J., Goldsmith, J., & Greven, S. (2024). Functional data analysis: An overview and recent developments. *Journal of the Royal Statistical Society: Series A*, 187(1), 1–45. <https://doi.org/10.1093/jrsssa/qnad120>
- Glasser, J., & Lindauer, B. (2013). Bridging the gap: A pragmatic approach to generating insider threat data. *2013 IEEE Security and Privacy Workshops*, 98–104. <https://doi.org/10.1109/SPW.2013.35>
- Greitzer, F. L., & Frincke, D. A. (2010). Combining traditional cyber security audit data with psychosocial data: Towards predictive modeling for insider threat mitigation. In *Insider threats in cyber security* (pp. 85–113). Springer.
- Greitzer, F. L., & Hohimer, R. E. (2011). Modeling human behavior to anticipate insider attacks. *Information Systems Frontiers*.

- Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2014). Identifying at-risk employees: A behavioral model for predicting insider threat. *Proceedings of the 47th Hawaii International Conference on System Sciences*, 2392–2401. <https://doi.org/10.1109/HICSS.2014.301>
- Guest, G., MacQueen, K. M., & Namey, E. (2012). *Applied thematic analysis*. Sage.
- Harilal, A., Tauritz, D., & Shen, Y. (2017). Toward automated feature engineering for insider threat detection. *IEEE Security & Privacy Workshops*.
- Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2019). Insight into insider threat detection: A survey. *IEEE Access*.
- Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.
- IBM Security. (2023). *Cost of a data breach report*. IBM Corporation.
- Jain, R. (2025). Cutting-edge developments in science, engineering and emerging technologies: A comprehensive review. *SSRN Electronic Journal (Working Paper)*.
- Kokoszka, P., & Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman & Hall/CRC.
- Le, T., Phan, T., & Nguyen, H. (2019). Insider threat detection using graph neural networks. *IEEE Big Data Conference*.
- Legg, P., Buckley, O., Goldsmith, M., & Creese, S. (2015). A survey of insider threat detection. *Journal of Information Security and Applications*.
- Li, H., et al. (2019). Insider threat detection with deep neural networks. *IEEE Access*.
- Liu, B., Kuhn, M., et al. (2018). Anomaly detection for insider threat. *IEEE Transactions on Information Forensics and Security*.
- Magklaras, G., & Furnell, S. (2020). Insider threat prediction: Past, present, and future. *Computers & Security*, 92.
- MITRE. (2019). *Insider threat framework*. MITRE Corporation.
- Nasir, R., Afzal, M., Latif, R., & Iqbal, W. (2021). Behavioral-based insider threat detection using deep learning. *IEEE Access*.
- Nurse, J. R. C., Creese, S., Goldsmith, M., & Lamberts, K. (2014). Understanding insider threat. *IEEE Security & Privacy*.

- Nurse, J. R. C., Buckley, O., Legg, P., Goldsmith, M., & Creese, S. (2020). COVID-19 and the insider threat. *Computers & Security*.
- Paré, G. (2004). Investigating IT practices. *Information & Management*.
- Ponemon Institute. (2020). Cost of insider threats report. Proofpoint.
- Rahman, M. H., Al Naeem, M. A., & Abubakar, A. (2022). Threats from unintentional internal actors. *IEEE Access*.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer.
- Shaw, E., & Fischer, L. (2005). Ten years of insider threat study. *Behavioral Sciences & the Law*.
- Shaw, E., & Sellers, L. (2015). Application of the critical-path method to detect insider threats. *Studies in Intelligence*.
- Sinclair, S., & Smith, S. (2008). Preventative directions for insider threat mitigation via access control. In *Insider threats in cyber security* (pp. 141–167). Springer.
- Theoharidou, M., Kokolakis, S., Karyda, M., & Kiountouzis, E. (2005). The insider threat to information systems. *Computers & Security*, 24(6).
- Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep learning for unsupervised insider threat detection. *IEEE MILCOM*.
- Ullah, A., & Finch, S. (2013). Functional data classification using PCA. *Computational Statistics & Data Analysis*.
- Voss, L., McCulley, D., & Park, J. (2015). Analyzing insider threats using the CERT dataset. *Proceedings of the International Conference on Cybersecurity and Privacy*.
- Yuan, F., Cao, Y., Shang, Y., Liu, Y., & Tan, J. (2018). Insider threat detection with deep neural networks. Springer.
- Yuan, X., Li, C., & Liu, X. (2018). Deep learning-based insider threat detection. *IEEE Access*.
- Yuan, X., Sun, L., & Liu, X. (2023). Graph neural networks for insider threat detection. *IEEE Transactions on Cybernetics*.

Annexes

Annexe 1 : Protocole d'entrevue

Merci de participer à cette entrevue. Elle s'inscrit dans le cadre de mon mémoire de maîtrise à HEC Montréal, portant sur les jeux de données simulés utilisés pour la détection proactive des menaces internes malveillantes.

Objectifs :

- Mieux comprendre les types de comportements jugés préoccupants
- Mieux comprendre les limites et défis associés à l'utilisation des jeux de données simulées
- Mieux comprendre les ensembles d'actions, mécanismes techniques pour repérer de manière proactive ou réactive les comportements suspects.

Remarque : On s'intéresse uniquement à la gestion d'incidents impliquant des employés permanents.

Nous nous intéresserons particulièrement à la gestion des incidents impliquant des employés permanents (pour les besoins de cette recherche, les employés temporaires, contractuels ou consultants sont exclus) et de leurs comportements potentiellement malveillants (c'est-à-dire des individus qui agissent délibérément avec l'intention de nuire à leur employeur, contrairement aux menaces internes malhabiles qui relèvent d'erreurs, de négligence ou d'accidents).

L'entretien durera environ 60 minutes.

Vos réponses resteront anonymes et confidentielles. Vous pouvez à tout moment refuser de répondre à une question ou mettre fin à l'entrevue.

Voici le formulaire de consentement à signer

Je démarre l'enregistrement

Questions générales

Objectif – Section 02 et Section 03 : déterminer la nature de l’expertise du répondant pour évaluer les scénarios de la troisième section

1. Contexte professionnel du participant

- Quel est votre rôle actuel en cybersécurité et le nombre d’années d’expérience ?
- Êtes-vous impliqué(e) dans la gestion d’incidents malveillants (surveillance de type de comportements alarmants) ?
- Avez-vous déjà travaillé avec des jeux de données simulés (au sens large) pour la détection des incidents malveillants ? Pour d’autres activités liées à la gestion des incidents malveillants (identifier, prévenir, répondre ou se remettre des incidents malveillants) ?

2. Nature et formes des menaces internes malveillantes

- Quels sont les types de comportements malveillants les plus fréquents selon votre expérience ? (Ex : vol de données confidentielles, utilisation abusive des ressources, fraude financière, etc.)
- Combien de cas de ces types de menaces internes malveillantes avez-vous observés au cours de votre carrière (une dizaine, une centaine, un millier, et si possible de quel type) ?

Objectif – Section 05 : comprendre les points forts et les limites des jeux de données

4. Données simulées et jeu de données

Si vous deviez concevoir un jeu de données simulé réaliste, parmi les éléments suivants, lesquels incluriez-vous :

1. Types de comportements à simuler (cochez tout ce qui s’applique) :

- ☐ Vol de données confidentielles
- ☐ Sabotage de systèmes
- ☐ Fraude financière

- ☐ Usage abusif des ressources
- ☐ Autre (précisez) : _____

2. Types de signaux à inclure (cochez tout ce qui s'applique) :

- ☐ Connexions hors horaires
- ☐ Accès inhabituels à des fichiers sensibles
- ☐ Utilisation de périphériques externes (USB, disques durs)
- ☐ Changements de rôle ou de permissions
- ☐ Impressions massives de documents
- ☐ Activité sur les intranets ou réseaux sociaux internes
- ☐ Autre : _____

3. Rôles ou profils à représenter (cochez tout ce qui s'applique) :

- ☐ Employé régulier
- ☐ Administrateur système
- ☐ Employé du service financier
- ☐ Gestionnaire de projet
- ☐ Stagiaire ou nouvel employé
- ☐ Autre : _____

Selon vous, quels seraient les enjeux des jeux de données simulés en pratique ? (Cochez tout ce qui s'applique)

- ☐ Données trop simplifiées ou stéréotypées
- ☐ Manque de réalisme dans les comportements simulés
- ☐ Faible granularité temporelle
- ☐ Absence de contexte métier ou organisationnel (ex. : rôle, service, mission)
- ☐ Manque de signaux faibles ou ambigus présents dans les situations réelles
- ☐ Données peu variées (ex. : seules les connexions réseau sont simulées, pas les accès physiques ou les courriels)

- ☐ Scénarios trop courts ou déconnectés de dynamiques à long terme
- ☐ Métadonnées absentes ou incomplètes (ex. : localisation, horaire précis, identité floue)
- ☐ Autre (veuillez préciser) : _____

Objectif – Section 06 : comprendre les points forts et les limites des outils de détection

5. Outils et logique de détection des comportements malveillants (5 min)

- Quels types d’outils utilisez-vous actuellement pour détecter les comportements malveillants internes (Ex. : systèmes automatisés fournisseur, systèmes maison, alertes manuelles, etc.) ?
- Êtes-vous satisfait de ces solutions ? Justifiez
- Selon vous, quels sont les limites ou enjeux actuels des systèmes actuels de détection automatisée de la menace malveillante (trop simples, trop rigides, ou au contraire trop complexes à expliquer) ?

6. Clôture

- Avez-vous des suggestions ou réflexions à partager ?
- Souhaitez-vous recevoir un résumé des résultats ?

Annexe 2 : Scénarios de menaces internes présentés aux participants

Cette annexe présente les scénarios de menaces internes malveillantes utilisés dans la section 04 du protocole d’entrevue.

 Scénario 1	1	Menaces internes malveillantes
En mars 2023, Daniel Tremblay, ingénieur logiciel chez NovaTech, travaille dans l'équipe de développement d'un produit stratégique dans le domaine des solutions intelligentes pour l'industrie manufacturière. Daniel considère que le climat de travail chez NovaTech est toxique. Son gestionnaire est perçu comme injuste, ne fournissant que des commentaires négatifs à son égard et envers son équipe. De plus, les priorités du produit changent fréquemment, souvent sans justification claire, ce qui alimente le sentiment de frustration et de désengagement chez Daniel. Progressivement, il devient plus réservé et fâché. Un jour, il décide qu'il en a eu assez. Il commence alors à modifier subtilement ses habitudes de travail. Au cours des semaines suivantes, il accède de façon répétée à des répertoires contenant des documents confidentiels portant sur l'architecture logicielle, les rapports de performance, ainsi que des analyses de marché. Bien qu'il ait le droit d'accéder à ces répertoires, il l'a rarement fait auparavant parce qu'ils sont peu utiles pour ses tâches habituelles. Discrètement, Daniel commence à transférer certains fichiers vers son adresse personnelle Gmail, en les renommant avec des titres génériques ou en changeant leur extension de fichier (par exemple .docx en .jpg) pour éviter de déclencher les règles de détection classiques. À plusieurs reprises, il connecte une clé USB non autorisée en fin de journée, souvent après les heures de bureau, pour copier des volumes importants de données. Ces actions sont fragmentées dans le temps, rendant leur détection plus difficile. Trois jours avant sa démission officielle, Daniel envoie par courriel plusieurs fichiers confidentiels regroupés, incluant des présentations internes sur la feuille de route produit, des projections financières et des plans de partenariat. Les systèmes de surveillance ne déclenchent aucune alerte, car chaque action isolée demeure dans les seuils de tolérance configurés par l'équipe de sécurité de NovaTech. Une semaine après son départ, un concurrent direct de NovaTech lance une campagne marketing très ciblée, reprenant plusieurs éléments clés de la stratégie de NovaTech et visant exactement les mêmes segments clients. Cette coïncidence soulève des soupçons : la direction suspecte une fuite de données stratégiques et déclenche une enquête interne.		

 Scénario 2	2	Menaces internes malveillantes
Depuis six ans, Nicolas B. travaille comme administrateur système chez CoreSys, une entreprise spécialisée dans les solutions d'hébergement cloud. Réputé pour sa grande compétence technique, il est aussi connu pour son caractère difficile et ses relations tendues avec plusieurs collègues. Ses gestionnaires décident alors de réduire ses responsabilités ainsi que la nécessité de collaborer avec les autres membres de son groupe. Officiellement, Nicolas accepte, sans protester, d'effectuer des tâches plus routinières, mais ses collègues remarquent un changement de comportement. Nicolas devient très silencieux, isolé, et semble amer. En privé, il exprime le sentiment d'avoir été trahi et humilié par la direction, et considère que son expertise n'est pas reconnue à sa juste valeur. Dans les semaines qui suivent, Nicolas se connecte régulièrement la nuit et les fins de semaine, sans demande officielle ni ticket de maintenance associé. Il remarque qu'il conserve encore plusieurs privilèges d'administration, qui n'ont pas été révoqués après sa réaffectation. Discrètement, il commence à modifier les scripts d'automatisation des sauvegardes. Ainsi, certains fichiers critiques sont exclus, d'autres sont corrompus volontairement. À plusieurs reprises, il désactive brièvement certaines règles de pare-feu, exposant temporairement les systèmes internes, puis les réactive sans laisser de trace explicite. Chaque action isolée pourrait être interprétée comme une erreur ou une configuration temporaire. Tout bascule lorsqu'une défaillance majeure affecte plusieurs environnements clients et les sauvegardes automatisées se révèlent inutilisables. En analysant les journaux, les analystes découvrent une séquence de modifications non documentées faites par le compte de Nicolas. Mais à ce moment-là, il est trop tard : Nicolas a donné sa démission quelques jours plus tôt et a quitté l'entreprise sans qu'aucune enquête n'ait été déclenchée à temps.		



Scénario 3

3

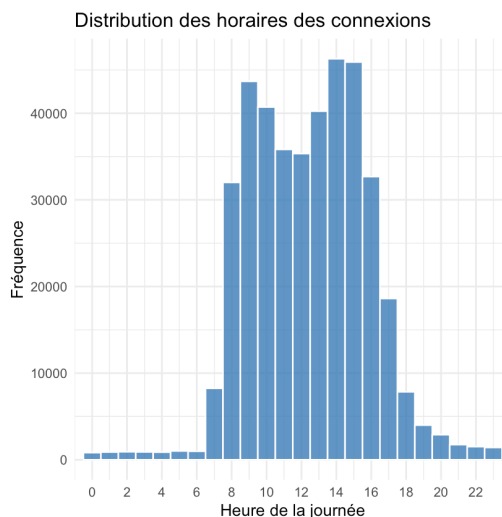
Menaces internes malveillantes

Sophie travaille depuis près de dix ans au service des comptes fournisseurs d'une organisation à but non lucratif. Reconnue pour sa rigueur et sa fiabilité, elle bénéficie de la confiance du directeur financier et est la seule employée à disposer à la fois des droits de création et de validation des paiements, une situation clairement en contradiction avec le principe de séparation des tâches.

À la suite d'un divorce coûteux, Sophie se retrouve en difficulté financière. Refusant de modifier son train de vie, elle voit dans son accès privilégié une opportunité de "reprendre le contrôle" de sa vie. Elle commence alors à créer de fausses factures au nom d'un fournisseur inactif, soigneusement choisies pour rester sous les seuils de validation secondaire. Les paiements sont versés sur un compte bancaire personnel, ouvert sous le nom d'une entreprise fictive. Elle approuve elle-même chaque transaction, en les espaçant dans le temps pour éviter les soupçons.

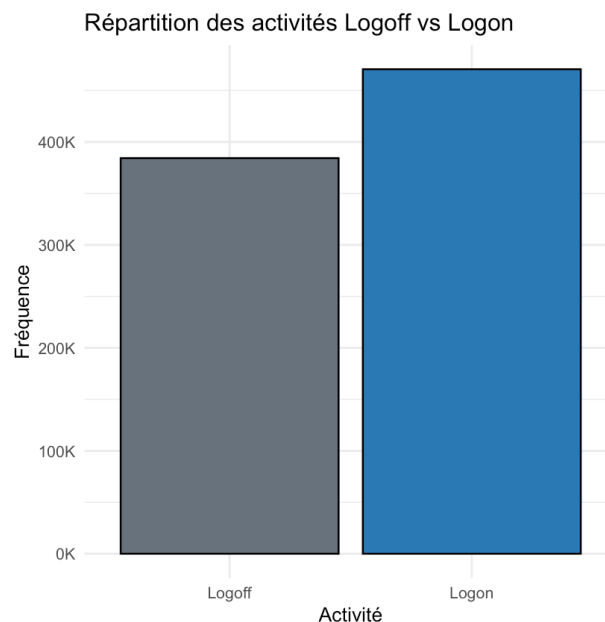
Pendant près de deux ans, personne ne détecte la fraude. Les paiements apparaissent réguliers et les montants restent modestes. Au total, plus de 80 000\$ sont détournés avant que l'arnaque ne soit découverte par hasard lors d'un audit externe indépendant.

Annexe 3 : Distribution horaire des connexions (table Device)



Description : Cette figure illustre la répartition des événements de connexion selon l'heure de la journée. La majorité des activités se concentrent entre 9h et 17h, tandis qu'environ 2,75 % des connexions surviennent en dehors des heures normales.

Annexe 4 : Distribution des activités Logon et Logoff sur toute la période (table Logon)



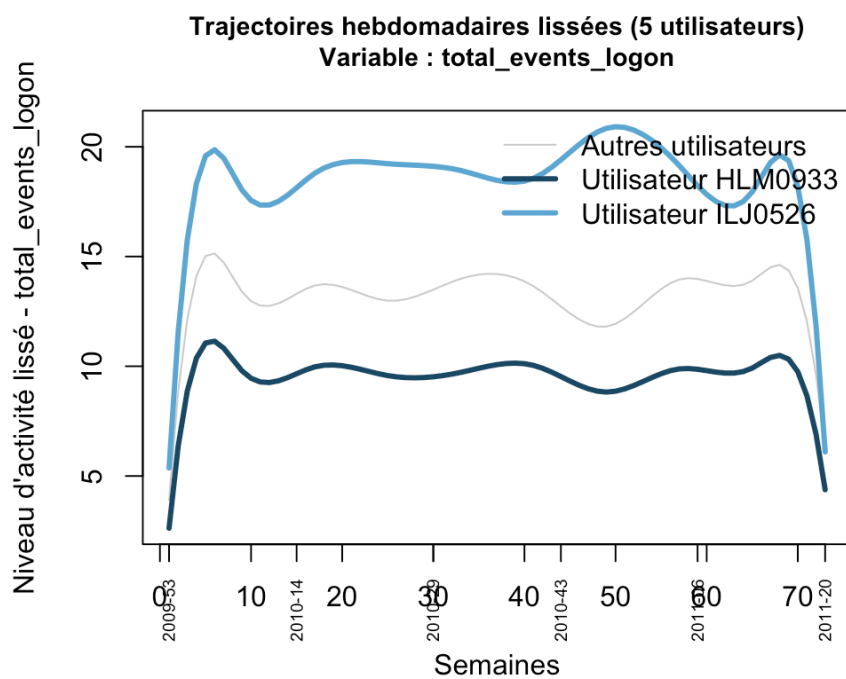
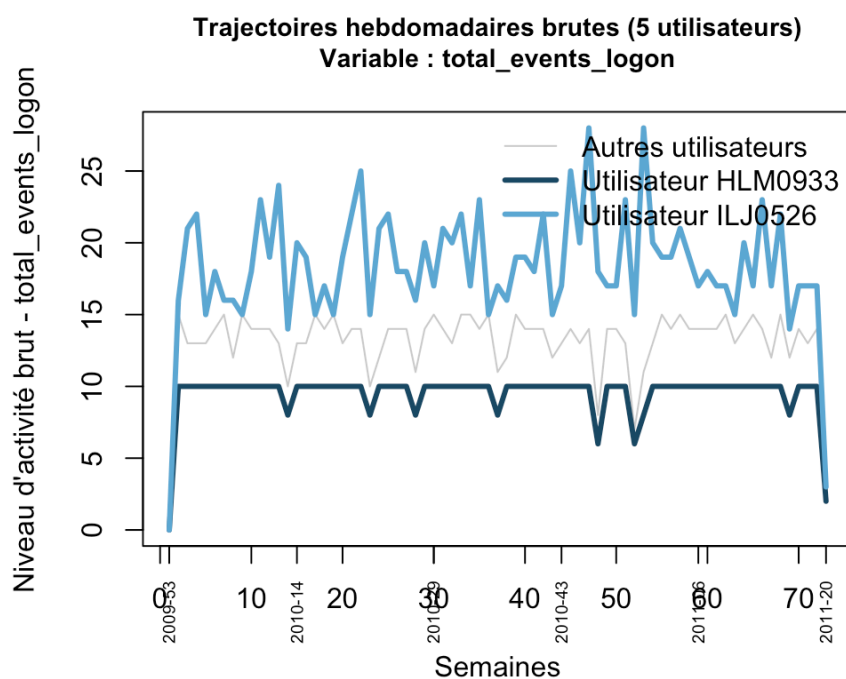
Description : Cette figure illustre la fréquence des événements *Logon* et *Logoff* dans les journaux d'authentification du jeu de données CERT r4.2. On observe un volume légèrement plus élevé de connexions (*Logon*) par rapport aux déconnexions (*Logoff*). Cet écart peut refléter des sessions laissées ouvertes, des fermetures implicites de session ou, dans certains cas, de possibles incohérences dans les journaux. Cette distribution offre un premier aperçu des habitudes d'authentification des utilisateurs et constitue un élément utile pour l'analyse des comportements atypiques ou incohérents dans les séquences d'accès.

Annexe 5 : Top 10 des utilisateurs actifs après 20h en fonction du nombre de fichiers envoyés (table File)



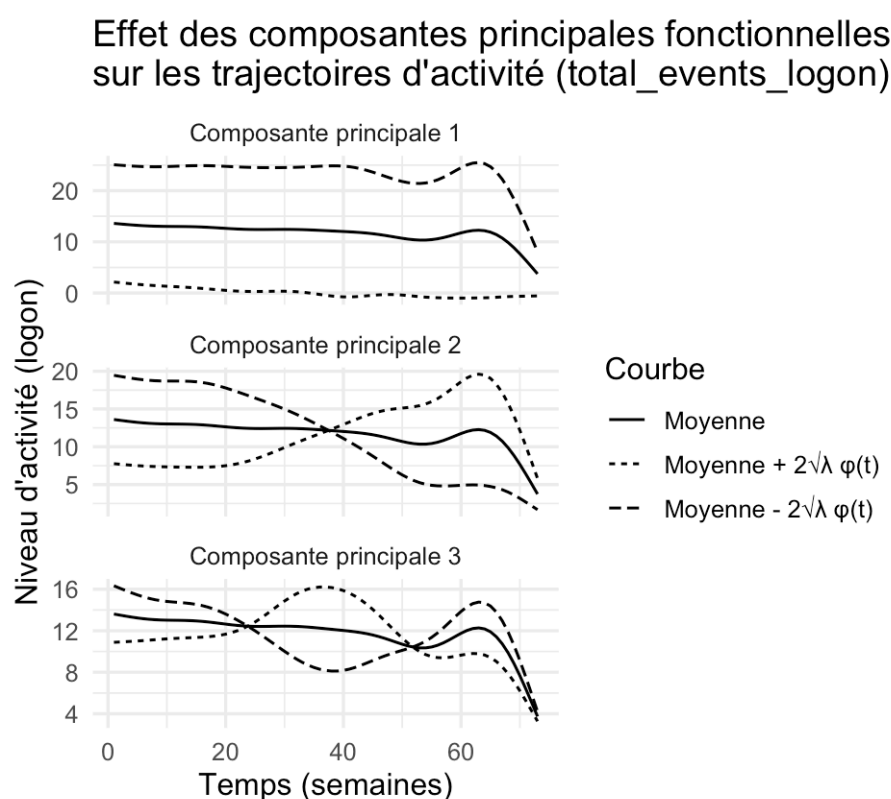
Description : Cette figure présente les dix utilisateurs ayant effectué le plus grand nombre d'opérations sur des fichiers après 20h, selon les journaux de la table File. L'activité nocturne constitue un indicateur couramment utilisé dans la littérature sur la détection des menaces internes malveillantes, car elle peut signaler des usages atypiques ou non conformes aux pratiques organisationnelles habituelles. Ici, certains utilisateurs (notamment AJF0370, BAL0044 et EIS0041) se démarquent par un volume particulièrement élevé de manipulations de fichiers en dehors des heures régulières, ce qui peut correspondre à des comportements inhabituels nécessitant une attention particulière dans le cadre des analyses subséquentes.

Annexe 6 : Trajectoires hebdomadaires brutes et lissées pour la variable total_events_logon



Description : Cette annexe illustre les trajectoires hebdomadaires de la variable (total_events_logon) pour cinq utilisateurs. La première figure montre les séries brutes, caractérisées par une forte variabilité et du bruit. La seconde figure présente les courbes lissées par B-splines, qui révèlent plus clairement les tendances temporelles et facilitent la comparaison entre utilisateurs. Deux utilisateurs sont mis en évidence afin d'illustrer la différence entre comportements individuels dans les données brutes et dans leurs versions lissées.

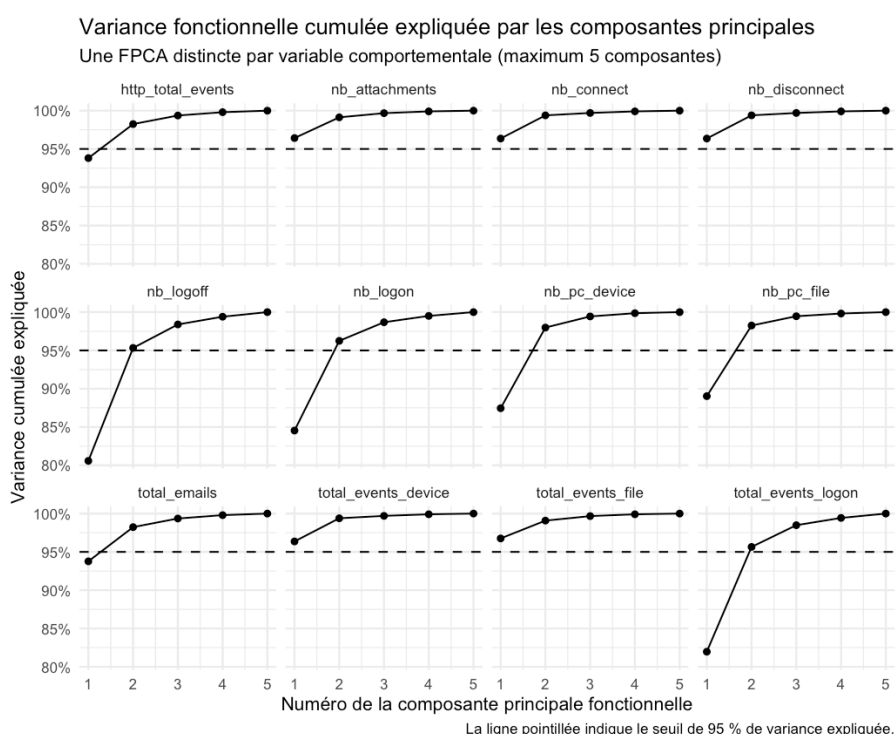
Annexe 7 : Effet des composantes principales fonctionnelles sur les trajectoires hebdomadaires (ex : activité logon)



Description : Le graphique présente l'effet des trois premières composantes principales fonctionnelles (FPCA) sur les trajectoires hebdomadaires d'activité logon. La ligne continue correspond à la trajectoire moyenne, tandis que les lignes pointillées montrent comment les composantes modifient cette trajectoire ($\pm 2\sqrt{\lambda} \phi(t)$). La première

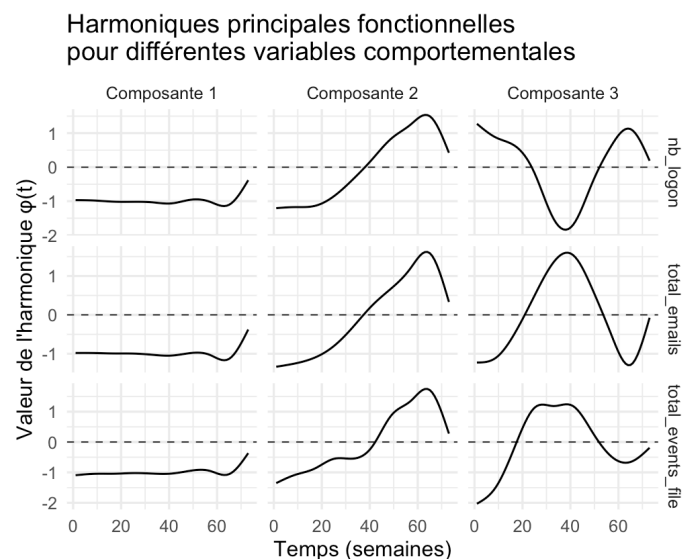
composante reflète des différences globales de niveau, la deuxième oppose le début et la fin de la période observée, et la troisième capture des fluctuations locales. Ensemble, ces composantes décrivent les principaux modes de variation des comportements dans CERT.

Annexe 8 : Variance fonctionnelle cumulée expliquée par les composantes principales pour les douze variables comportementales



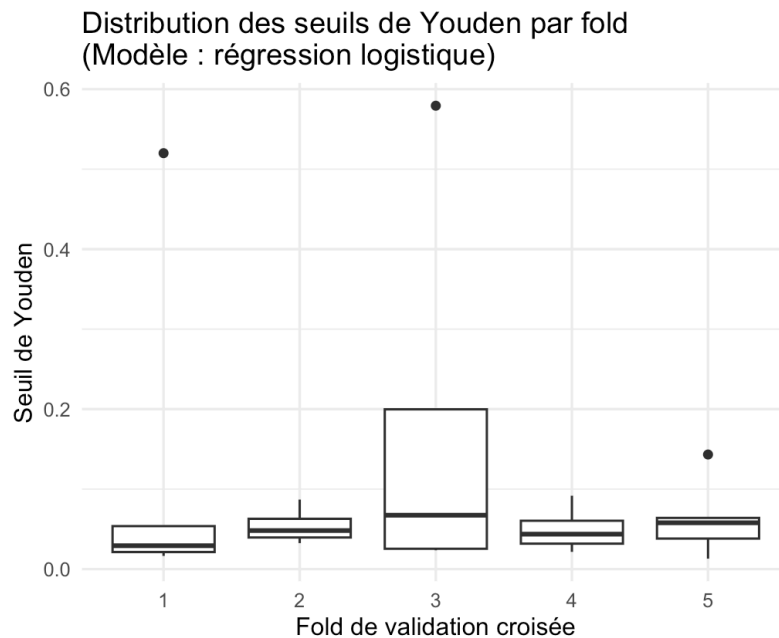
Description : Ce graphique montre, pour chacune des douze variables comportementales, la proportion cumulée de variance expliquée par les composantes principales issues de la FPCA. La plupart des variables atteignent le seuil de 95 % de variance expliquée en une ou deux composantes, indiquant que leurs trajectoires temporelles sont dominées par quelques modes de variation principaux. Cela confirme l'efficacité de la FPCA pour résumer les comportements simulés dans CERT tout en limitant la complexité des modèles supervisés.

Annexe 9 : Harmoniques principales fonctionnelles pour trois variables comportementales (FPCA)



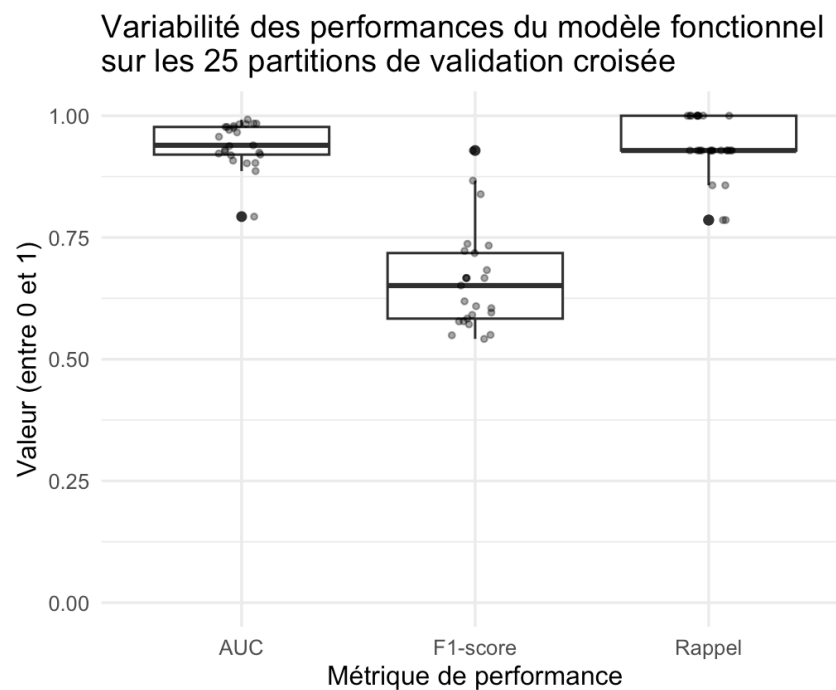
Description : Ce graphique illustre les trois premières harmoniques principales fonctionnelles (composantes 1 à 3) obtenues par FPCA pour trois variables comportementales : (nb_logon), (total_emails) et (total_events_file). Chaque courbe $\phi(t)$ représente un axe de variation majeur des trajectoires hebdomadaires des utilisateurs. La première composante reflète principalement le niveau global d'activité, tandis que les composantes 2 et 3 capturent des motifs temporels plus complexes, tels que des augmentations progressives, des fluctuations cycliques ou des variations anormales de comportement. Ces patterns permettent d'identifier les dynamiques sous-jacentes propres à chaque type d'activité et de repérer des comportements potentiellement atypiques.

Annexe 10 : Distribution des seuils de Youden par fold de validation croisée (régression logistique)



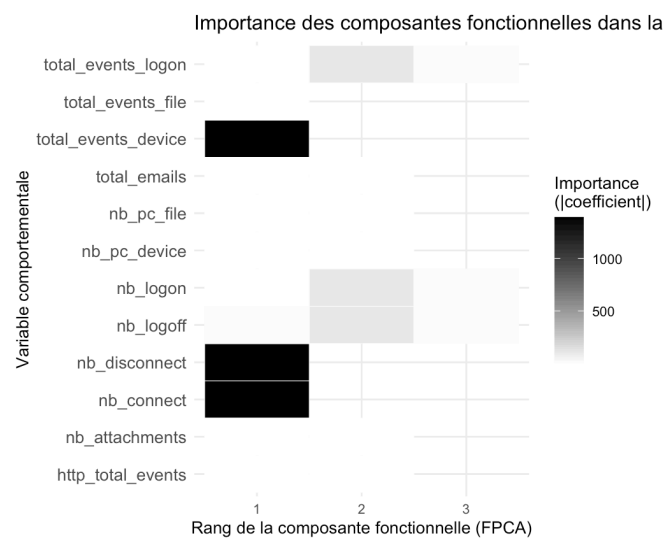
Description : Ce graphique en boîtes montre la variabilité du seuil optimal de Youden obtenu pour chaque fold de la validation croisée à 5 plis. On observe que la plupart des seuils restent bas et relativement stables d'un fold à l'autre, avec quelques valeurs plus élevées qui traduisent des fluctuations propres à certains sous-échantillons.

Annexe 11 : Variabilité des performances du modèle fonctionnel sur l'ensemble des partitions de validation croisée



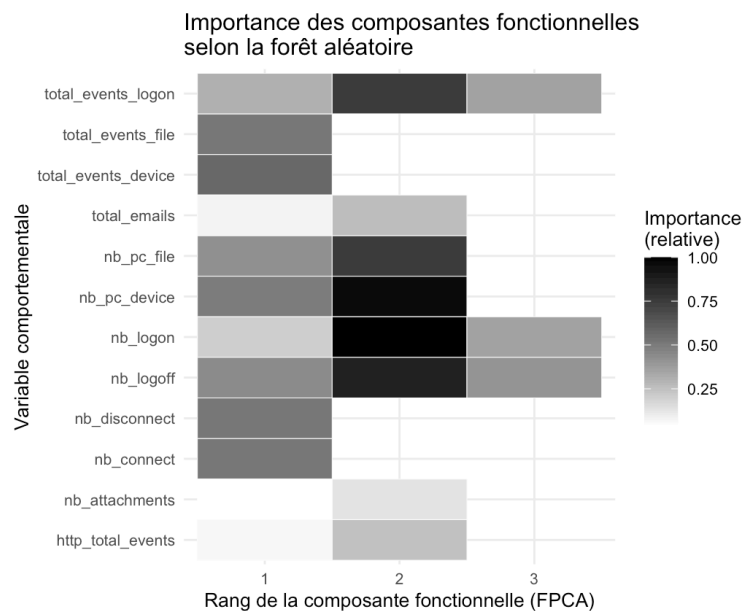
Description : Ce graphique illustre la variabilité des performances du modèle fonctionnel à travers les 25 partitions de validation croisée. Les boxplots montrent que l'AUC et le rappel demeurent très stables, révélant une capacité discriminante robuste et peu sensible aux variations entre les échantillons d'entraînement et de test. Le F1-score présente une dispersion légèrement plus élevée, reflétant l'équilibre plus délicat entre précision et rappel. Dans l'ensemble, la faible variabilité observée confirme la solidité et la fiabilité du modèle fonctionnel basé sur la FPCA, renforçant son potentiel d'application dans des contextes opérationnels de détection.

Annexe 12 : Importance relative des composantes principales fonctionnelles dans le modèle de régression logistique



Description : Ce graphique présente, pour chaque variable comportementale, l'importance des composantes principales fonctionnelles utilisées dans le modèle de régression logistique. L'intensité de la couleur reflète la magnitude absolue du coefficient associé à chaque composante : plus la teinte est sombre, plus la composante contribue fortement à la discrimination entre utilisateurs malveillants et non malveillants. On observe que plusieurs variables, notamment (nb_connect), (nb_disconnect) et (total_events_device), présentent des composantes fortement influentes, confirmant que la structure temporelle de ces comportements joue un rôle clé dans la détection des menaces internes malveillantes.

Annexe 13 : Importance relative des composantes principales fonctionnelles dans la forêt aléatoire



Description : Ce graphique présente l'importance attribuée par la forêt aléatoire aux différentes composantes fonctionnelles (FPCA) dérivées des variables comportementales. On observe que, pour plusieurs variables, les composantes de rang supérieur (FPCA2 et FPCA3) jouent un rôle déterminant dans la prédiction, ce qui confirme que le modèle ne se limite pas aux tendances générales d'activité. Les motifs temporels plus fins, tels que les fluctuations locales ou les irrégularités comportementales, contribuent fortement à la distinction entre utilisateurs malveillants et non malveillants.

Annexe 14 : Déclaration sur l'utilisation de l'intelligence artificielle générative (IAG)

Je déclare avoir pris une entente avec mes directeurs de recherche dans le cadre de ce travail quant aux types d'utilisation faite de l'intelligence artificielle générative (IAG). Usages permis de l'IAG.

Je m'engage à respecter les principes d'intégrité académique et à utiliser l'IAG uniquement dans les limites suivantes :

- **Compréhension des concepts**

L'IA peut être utilisée pour m'aider à mieux comprendre un modèle, un article de recherche ou une méthodologie scientifique. Cependant, l'analyse critique et l'interprétation des résultats doivent être personnelles et réalisées par moi-même.

- **Collecte de la littérature**

L'IA peut être utilisée comme outil d'assistance pour identifier des articles et des ouvrages pertinents en lien avec ma problématique. Toutefois, la sélection finale et l'analyse de ces sources doivent être effectuées de manière critique par moi-même.

- **Correction linguistique**

L'IA peut être employée pour la correction des fautes d'orthographe, de grammaire et de syntaxe. Cependant, je veille à ce que mon argumentation et mes idées restent personnelles et conformes aux exigences académiques.

- **Correction de code informatique**

Dans le cas de l'utilisation de code informatique dans mon mémoire, l'IA peut être sollicitée pour identifier et corriger des erreurs techniques. Toutefois, la conception, l'implémentation et l'explication du code doivent refléter mon propre travail et ma compréhension.

- **Recherche de données**

L'IA peut m'aider à trouver des sources de données pertinentes pour ma problématique de recherche. Néanmoins, la vérification de la fiabilité et de la validité de ces données m'incombe, tout comme leur analyse.

Usages exclus de l'IA

Je déclare ne pas avoir utilisé l'IA pour :

- Générer automatiquement des portions de mon mémoire sans analyse critique de ma part.
- Produire du contenu sans vérification et reformulation personnelles.
- Réaliser une analyse ou une discussion artificiellement construite sans intervention réelle de ma part.