جامعة عبد المالك السعدي
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵏ ⵄⴰⴱⴷⵍⵎⴰⵍⵉⴽ ⵙⵙⴰⵄⴷⵉ
Université Abdelmalek Essaadi

Centre des Etudes Doctorales Sciences et Techniques et Sciences Médicales
Formation Doctorale : Sciences et Techniques de l'Ingénieur
Etablissement : ENSA TANGER

Nom et Prénom : Chaoui kenza
Date de la soutenance : 19-12-2024
Directeur de Thèse : Hassan badir

Structure de recherche : *Équipe de Recherche en Ingénierie des Données et des Systèmes (IDS)*

Intitulé de la thèse : VERS UNE PROTECTION DYNAMIQUE ET INTELLIGENTE DES DONNÉES DURANT LEUR CYCLE DE VIE

Résumé

Le développement technologique rapide et l'interconnectivité croissante ont créé un environnement commercial riche en opportunités, mais aussi en défis majeurs en matière de sécurité des données. Avec l'explosion des données sensibles et la sophistication croissante des menaces telles que les attaques informatiques, les ransomwares et le vol de données, la protection des informations est devenue cruciale. Les entreprises sont désormais soumises à des réglementations strictes, comme le RGPD, qui impose des obligations sévères et des sanctions financières en cas de non-conformité.

La sécurité des données repose sur trois piliers fondamentaux : la confidentialité, l'intégrité et la disponibilité (CID). Ces trois éléments sont interdépendants et doivent être garantis tout au long du cycle de vie des données pour assurer leur protection complète. Cependant, de nombreuses solutions existantes se concentrent sur un seul aspect de cette triade, laissant des failles importantes dans la sécurité globale. Chaque étape du cycle de vie des données – de la collecte à la destruction – présente des vulnérabilités potentielles. C'est pourquoi la traçabilité devient un élément clé pour identifier et corriger ces failles tout en garantissant la conformité avec les lois et la transparence des processus. Une traçabilité efficace permet non seulement de limiter les dommages en cas de violation, mais aussi de répondre rapidement aux incidents de sécurité.

Notre approche s'inscrit dans ce cadre en proposant une méthode d'étiquetage des données en deux niveaux, s'appuyant sur des outils avancés et des régulations strictes pour assurer une protection robuste. Le premier niveau repose sur l'algorithme des k plus proches voisins (k-NN) et l'outil ExifTool. ExifTool nous permet d'extraire les métadonnées de façon fiable et automatique, et grâce à k-NN, nous effectuons un étiquetage initial des données en fonction de leur proximité avec des points de référence prédéfinis. Le deuxième niveau utilise une annotation automatique par segments pour l'étiquetage du contenu textuel. Ce processus est enrichi par des bibliothèques légales contenant les régulations de protection des données, notamment la HIPAA, la FERPA, la PIPEDA et le GDPR. Ces lois sont directement intégrées à l'annotation, nous permettant ainsi d'aligner l'étiquetage des données sur les exigences légales spécifiques à chaque juridiction, assurant une classification finale précise.

Ensuite, nous avons intégré l'algorithme MD5checksumK pour vérifier l'intégrité des données à chaque étape du processus. Contrairement aux méthodes de hachage classiques, MD5checksumK offre une détection améliorée des altérations de données tout en évitant les collisions, ce qui le rend plus robuste dans des environnements de données complexes et sensibles. Chaque altération détectée

est ensuite comparée à un dictionnaire de données pour vérifier si elle est autorisée ou non, garantissant ainsi que les modifications légitimes ne sont pas signalées comme des violations.
En complément, nous avons mis en place une méthode d'altération/désaltération pour sécuriser les étiquettes des données sensibles avant leur stockage. Ce processus consiste à altérer temporairement les étiquettes en rendant les données illisibles, de sorte que même en cas de compromission, elles ne révèlent aucune information exploitable.
Lorsqu'un accès aux données est nécessaire, le processus de désaltération rétablit les étiquettes à leur état original, assurant ainsi la confidentialité et l'intégrité des données tout en masquant leur nature sensible en dehors du contexte autorisé.
Pour garantir la gestion et la sécurité continues des données, nous avons conçu un système multi-agent intelligent composé de cinq agents spécialisés, chacun ayant un rôle bien défini :
- L'Agent d'Étiquetage des Données se charge d'étiqueter à la fois les métadonnées et le contenu textuel selon les deux niveaux mentionnés précédemment.
- L'Agent d'Intégrité des Données utilise des algorithmes de hachage, notamment MD5checksumK, pour vérifier et maintenir l'intégrité des données tout au long de leur cycle de vie, en détectant et en signalant toute altération.
- L'Agent d'Altération des Données applique l'altération des données avant leur stockage pour garantir leur sécurité en masquant les informations sensibles, tout en permettant une restauration rapide via désaltération.
- L'Agent de Cohérence des Étiquettes assure la cohérence des étiquettes attribuées, en s'assurant qu'elles ne sont ni altérées ni corrompues au cours du processus de gestion des données.
- L'Agent de Réponse aux Incidents, aussi appelé Agent Intelligent pour la Tolérance aux Pannes, surveille en temps réel toutes les activités du système. Il garantit la disponibilité continue des données grâce à un mécanisme de clonage et de réplication, permettant ainsi une tolérance aux pannes et une réponse rapide en cas d'incident.
Ce système multi-agent permet de renforcer la résilience, la tolérance aux pannes, et la disponibilité des données, tout en assurant une gestion fluide et automatisée de leur sécurité. Grâce à la réplication des agents et au clonage, le système peut réagir instantanément aux menaces, minimisant ainsi les interruptions et assurant une protection permanente.
En résumé, nos travaux de recherche se concentrent sur le développement d'une approche automatisée de sécurisation des données, utilisant des agents intelligents pour garantir la traçabilité, la confidentialité, l'intégrité et la disponibilité des données tout au long de leur cycle de vie. Cette approche répond aux besoins pressants des entreprises dans un environnement technologique de plus en plus complexe et interconnecté, tout en respectant les régulations internationales en matière de protection des données.
Mots clés : Sécurité, cycle de vie, Systèmes multi agents, Traçabilité, Intégrité, Confidentialité, Disponibilité.