République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A/Mira de Béjaia
Faculté des Sciences Exactes
Département d'Informatique

MÉMOIRE DE MASTER RECHERCHE EN INFORMATIQUE

Option

Réseaux et Sécurité

Thème

ConfidCloud+ : Une approche de confidentialité dans un environnement multi-cloud.

Présenté par : Mlle. Aidi Nassima & Mlle. Ait ikhlef Imene

Président	**Dr S. Aissani**	**Maître de conférence. A**	**U. A/Mira Béjaia.**
Examinatrice	**Dr D. Zamouche**	**Maître assistant. B**	**U. A/Mira Béjaia.**
Encadrante	**Dr N. Yessad**	**Maître de conférence. B**	**U. A/Mira Béjaia.**

Béjaia, Juin 2023.

⁕ *Remerciements* ⁕

Nous tenons à exprimer notre profonde gratitude et nos sincères remerciements à toutes les personnes qui nous ont soutenus tout au long de notre parcours et qui ont contribué à la réalisation de ce mémoire.

Tout d'abord, nous souhaitons exprimer notre reconnaissance à ALLAH, qui nous a guidés et nous a accordé la force et la persévérance nécessaires pour mener à bien ce travail. Sa grâce infinie nous a permis d'accomplir cette étape importante de notre vie académique.

Nous tenons également à exprimer notre gratitude à nos parents, AIT IKHLEF SAID & MOUSSAOUI FATIMA et AIDI SAID & AZRAINE ZAHRA, pour leur amour, leur soutien indéfectible et leurs encouragements constants. Leur confiance en nous a été un moteur essentiel dans la réalisation de ce mémoire. Leurs sacrifices, leurs valeurs et leur éducation ont été des piliers fondamentaux dans nos vies.

Nous souhaitons adresser nos remerciements les plus chaleureux à notre encadrante, NAWAL YESSAD, pour sa guidance précieuse, son expertise et son dévouement tout au long de ce projet. Ses conseils éclairés, sa disponibilité et sa patience ont grandement enrichi notre travail. Nous sommes reconnaissants pour l'opportunité qui nous a été donnée de bénéficier de son encadrement.

Nous souhaitons exprimer notre reconnaissance à toutes les personnes qui ont été proches de nous pendant cette période. Votre soutien moral, vos encouragements, vos précieux conseils et votre présence ont été d'une importance capitale pour nous. Vos mots d'encouragement et vos discussions fructueuses ont nourri notre réflexion et ont contribué à l'amélioration de ce mémoire.

Enfin, nous voudrions adresser nos remerciements à tous les enseignants et nos camarades de classe qui ont contribué de différentes manières à notre apprentissage et à notre développement académique.

Nos sincères remerciements vont également à toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce mémoire et qui n'ont pas été mentionnées précédemment.

Que toutes ces personnes soient assurées de notre reconnaissance éternelle et de notre profonde gratitude pour leur soutien inestimable.

Mlle. NASSIMA AIDI Mlle. AIT IKHLEF IMENE

Table des matières

Table des figures

Liste des tableaux

Notations et symboles

$NIST$	National Institute of Standards and Technology.
IT/TI	Information Technology.
IA/AI	Artificiel Intelligent.
CSA	Cloud Security Alliance.
DNN	Deep Neural Network.
ML	Machine Learning.
KNN	k-nearest neighbors.
VM	Virtual Machine .
$KDD99$	Knowledge Discovery in Databases.
IDS	Intrusion Detection System.
$DDOS$	Destrebuted denial of service.
IBE	identity based encryption.
ECC	Elliptical Curve Cryptography.
CNN	Convolutional Neural Network.
$SOAR$	security orchestration, automation and response.
MFA	multi-factor authentication.
$5G$	5-ème génération.
AWS	Amazon Web Services.
Vs	verses.
LOB	Line of Business.
RH	Ressources humaines.
DES	Data Encryption Standard.
AES	Advanced Encryption Standard.
RSA	Rivest-Shamir-Adleman.
DAC	Discretionary Access Control.
HRU	Rôle d'Habilitation Utilisateur.
MAC	Mandatory Access Control.
$RBAC$	Role-Based Access Control.
API	Application Programming Interface.
UI	User Interface.
$HTTP$	Hypertext Transfer Protocol.
SYN	Synchronous.
UDP	User Datagram Protocol.
TCP	Transmission Control Protocol.
$ICMP$	Internet Control Message Protocol.
IP	Internet Protocol.
SQL	Structured Query Language.
$MITM$	Man in the middle.
BDD	Base de données.

Introduction générale

Au cours des dernières années, la quantité de données produites par les individus, les entreprises et les organisations a entraîné une explosion ; pour faire face à cette surcharge d'informations, le cloud computing est devenu une solution essentielle. Il permet de stocker les données appartenant tant aux utilisateurs qu'aux entreprises ; ainsi que le cloud offre d'autres nombreux avantages en matière de flexibilité, d'évolutivité et de coût,etc. Néanmoins, étant donné la sensibilité de ces données, il est important d'assurer leurs sécurités contre les atteintes à leur confidentialité par des cybercriminels.

La confidentialité dans le cloud est un aspect fondamental dans le monde de l'information, dont la protection des données sensibles contre tout accès non autorisé représente un grand défi, car les conséquences de telles intrusions peuvent être désastreuses dans la majorité des cas ; Et malgré l'utilisation de diverses méthodes, mais les résultats ne sont pas assez satisfaisants,dont les cybercriminels arrivent au fil du temps à trouver des nouvelles vulnérabilités afin de compromettre la confidentialité de ces données.

Notre approche CondifCloud+ vise à garantir un niveau élevé de confidentialité des données stockées dans le cloud en utilisant diverses techniques ; afin de réussir nous allons analyser les vulnérabilités et les risques potentiels auxquels les données sont exposées, ainsi qu'à identifier les meilleures pratiques en matière de protection des informations confidentielles.

Dans le but d'atteindre nos objectifs, nous avons suivi une approche méthodologique basée sur une recherche dans la littérature, en examinant des différents articles, enquêtes,etc. qui traite et propose des outils pour assurer la confidentialité des données stockées ; notre analyse porte essentiellement sur les protocoles de chiffrement, les mécanismes de contrôle d'accès ainsi que les techniques de classification des données en trois niveaux. Par ailleurs, nous avons prévu d'effectuer une analyse de sécurité, des calculs de complexité,etc. En vue d'évaluer les capacités de l'approche ConfidCloud+.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre a pour objectif de définir les concepts fondamentaux du cloud computing, ainsi que d'introduire certaines notions essentielles pour une meilleure compréhension. Le deuxième chapitre, aborde la sécurité au niveau de cloud, en présentant les services de sécurité nécessaires et les diverses attaques et menaces existantes, ainsi que les revues de littérature les plus couramment utilisées dans le cadre de notre recherche. Dans le troisième chapitre nous avons présentées notre solution ConfidCloud +, en expliquant son fonctionnement, ainsi que les techniques mises en oeuvre pour mettre en place cette approche visant à garantir la confidentialité des données. Le chapitre quatre, présente les résultats de l'évaluation des capacités de l'approche ConfidCloud+ atteintes.

Nous clôturons par une conclusion générale en résumant les points principaux de l'approche ConfidCloud+, et les étapes suivies afin de réaliser ce travail.

Chapitre 1

Généralités et conceptes de base

1.1 Introduction

Dans le monde Informatique, une technologie qui gagne progressivement du terrain est le cloud computing. Cette méthode permet aux utilisateurs d'accéder à distance aux services Informatiques, mais sans le fardeau de la gestion de l'infrastructure à partir de laquelle il s'exécute. Ce faisant, le cloud computing permet aux utilisateurs de partager des ressources telles que des données, des applications, des services et des réseaux. Cette méthode présente des avantages distincts, parmi lesquels la flexibilité, la disponibilité et la rentabilité. Le cloud computing peut entraîner à la fois des difficultés de gestion des données et des dangers pour la sécurité. Pour mieux comprendre, ce chapitre d'introduction se penchera sur les différents modèles de services de cloud computing, les types de déploiement, les avantages et les obstacles. De plus, les développements actuels seront examinés et des conseils d'experts sur la manière d'optimiser les avantages tout en minimisant les risques seront donnés.

1.2 Définition du cloud computing

Le Cloud computing ou l'Informatique en nuage en français représente un modèle unique, il est considéré parmi les meilleures techniques de gestion des ressources Informatique, il permit aux utilisateurs de partager et d'accéder aux ressources Informatique telles que les services, les données, et les applications à distance, à travers le réseau à partir de n'importe quel appareil connecté à l'Internet [41].

1.3 Caractéristiques du cloud computing

Voici les principales caractéristiques qui différencie le cloud des autres technologies, en d'autres mots ce sont ces points qui font d'un cloud un cloud :

- La mise en commun des ressources :

« Pool the infrastructure, virtual platforms, and applications » [53] C'est Le fournisseur qui collecte et rassemble les ressources dans un pool, dont des parties peuvent être allouées à différents consommateurs (généralement sur la base de politiques), Les exemples de ressources incluent le stockage, le traitement, la mémoire,etc. [19][10].

- Le Libre-service à la demande :

« Consume services when you want » [53] Les consommateurs fournissent des ressources à partir du pool en utilisant le libre-service à la demande tel que le temps du serveur et le stockage. Ils gèrent eux-mêmes les ressources sans avoir à traiter avec des administrateurs humains [19].

- L'accès étendu au réseau :

« Consume services from anywhere » [53] L'accès WAN signifie que toutes les ressources sont disponibles sur le réseau sans accès physique direct. Le réseau ne fait pas nécessairement partie du service, les ressources sont disponibles via : téléphones mobiles, ordinateurs portables,etc. [19][10].

- Le service mesuré :

« Pay for the service you consume as you consume it » [53] Les services de comptage mesurent les services rendus pour s'assurer que les consommateurs n'utilisent que les services qui leur sont attribués, et si nécessaire, les facturent. C'est de là que vient le terme d'Informatique utilitaire. Les ressources Informatiques peuvent désormais être consommées comme l'eau et l'électricité, et les clients ne paient que ce qu'ils consomment [19].

- La flexibilité :

« Share pooled resources to enable horizontal scalability » [53] La flexibilité, ou élasticité rapide, permet aux consommateurs d'augmenter ou de réduire les ressources qu'ils consomment à partir du pool (provisionnement et déprovisionnement), généralement de manière entièrement automatique. Cela leur permet de mieux faire correspondre la consommation de ressources à la demande (par exemple, ajouter des serveurs virtuels lorsque la demande augmente, puis les arrêter) [19].

1.4 Les modèles de services cloud

Tout peut être considéré comme un service mais voici les principaux services du cloud computing :

1.4.1 Platform-as-a-Service

PaaS : (Platforme-as-a-Service) il offre une plateforme en ligne qui est accessible aux différents utilisateurs telles que les bases de données, les plateformes d'application (par exemple, un endroit où exécuter du code Python, PHP ou autre) cette plateforme peut être utilisée pour le stockage de fichiers et la collaboration, ou même le traitement d'applications propriétaires (telles que l'apprentissage automatique ou le traitement des données volumineuses)[19].

1.4.2 Infrastructure-as-a-Service

IaaS : (Infrastructure-as-a-Service) ; c'est la couche inférieure des services cloud computing elle offre des ressources basiques de l'infrastructure Informatique fondamentale, le réseau et le stockage, et les utilisateurs sont obligés de faire la gestion de leurs serveurs Cloud et de faire l'interconnexion entre ces serveurs ; il est utilisé pour éviter de perdre beaucoup de temps et d'argent dans la gestion du matériel et du logiciel.

1.4.3 Software-as-a-Service

SaaS : (Software-as-a-Service) : C'est un modèle de développement logiciel, le déploiement le management,etc. Il est confier à une tierce partie, c'est pourquoi on a plus besoin de se soucier des problèmes techniques et juste utiliser les logiciels comme on le souhaite, dont tout le monde a accès aux services SaaS, comme Dropbox, Facebook, Instagram, drive,etc. Le SaaS permet de réduire les coûts et facilite le développement d'applications. En résumé c'est une application complète gérée par les fournisseurs, et les utilisateurs peuvent y accéder via un navigateur.

1.4.4 Data-as-a-Service

DaaS :(Data-as-a-Service) : Avec l'évolution technologique et la modernisation de la science des données des nouvelles catégories de services cloud computing ont vu le jour dont le Data-as-a-Service représente essentiellement un sous-type du Software-as a- Service, il est principalement basé sur la caractéristique de la demande sur mesure qui offre un accès sur demande au produit de données [60].

1.4.5 Desktop-as-a-Service

DaaS (Desktop-as-a-Service) : C'est un autre sous-type du Software-as-a-Service il délivre des applications virtuelles et des services de bureau sur internet, il permet aussi d'utiliser un service de bureau à distance pour accéder aux ordinateurs de bureau à distance. Le serveur citrix est un exemple de DaaS qu'on peut utiliser par le biais d'un navigateur web [60].

1.4.6 Function-as-a-Service

FaaS (Function-as-a-Service) : Parmi les nouvelles catégories qui ont vu le jour ; figure le FaaS, ici l'utilisateur n'a plus besoin de se focaliser sur l'infrastructure mais il se base surtout sur le code à exécuter, et c'est le FaaS qui s'en charge de l'exécution, c'est lui qui fournit l'environnement d'exécution et celui qui s'en charge de sa gestion.

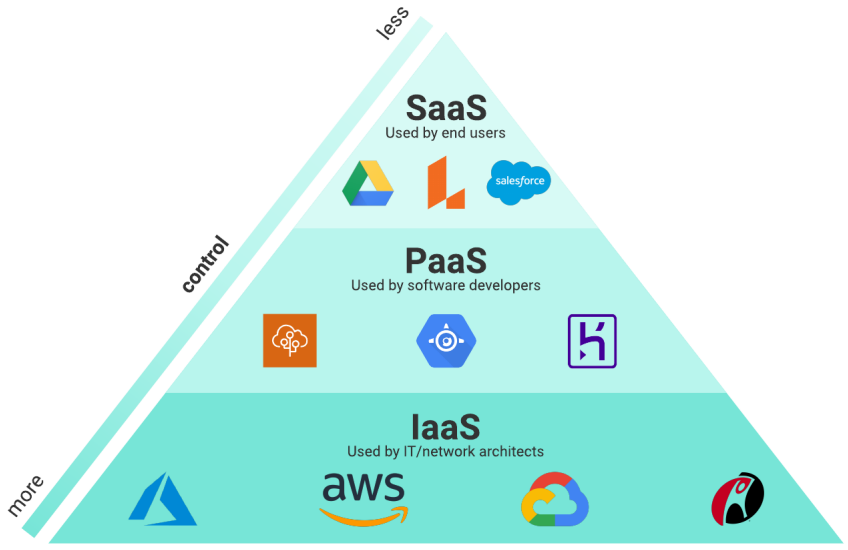

FIGURE 1.1 – Les trois modéles essentiels de services cloud computing [63].

1.5 Comparaison entre les modèles de services essentiels du cloud

Le IaaS, PaaS et SaaS représentent les niveaux de services du cloud computing, si nous utilisons que l'infrastructure c'est le IaaS, si l'environnement d'exécution et de travail est prêt c'est le PaaS, et si on utilise directement les logiciels c'est le SaaS ; ce qui signifie que c'est le type d'application qui détermine le niveau de service à utiliser.

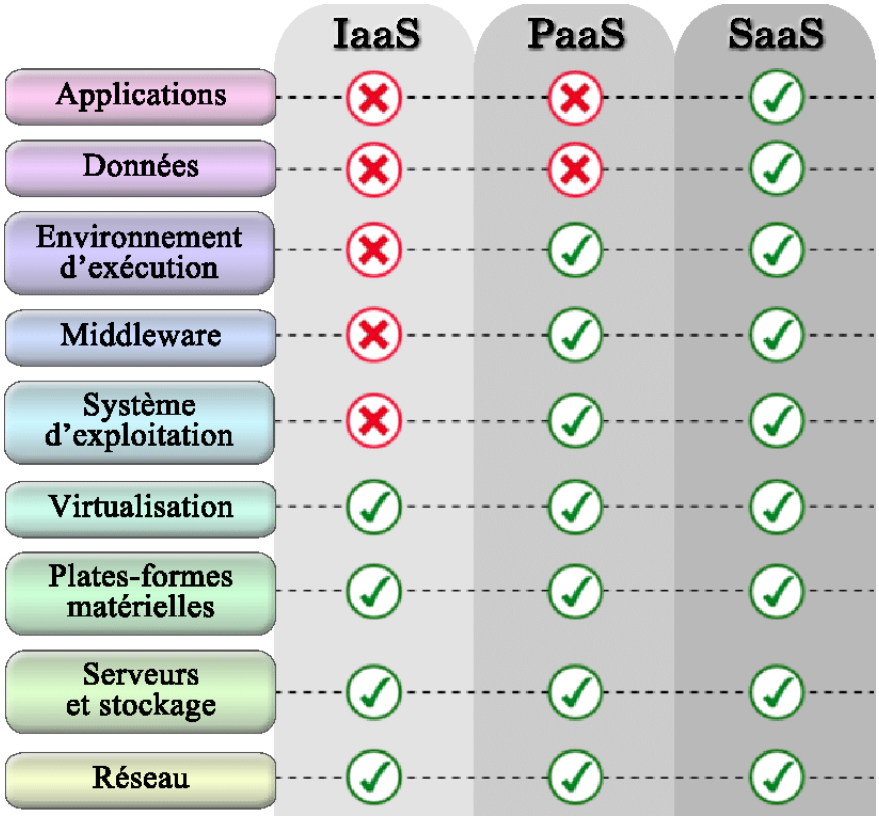

FIGURE 1.2 – Comparaison entre les modéles de services essentiels cloud [54].

1.6 Les modèles de déploiement du cloud

1.6.1 Public

Le cloud public est le modèle le plus commun et le plus utilisé parmi les autres modèle de cloud computing, grâce à son déploiement rapide et facile [62] ; les services offerts par ce cloud sont accessibles par tout le monde, et les ressources sont mises en commun entre plusieurs clients et facturées en fonction de leur utilisation, ainsi que les entreprises peuvent utiliser les fonctionnalités infonuagiques d'autres fournisseurs ou proposer leurs propre services à des utilisateurs en dehors de l'entreprise [27].

1.6.2 Privé

Ce modèle de cloud est appelé 'privé', car en base il est réservé aux utilisateurs exclusives contrairement au cloud public, les ressources offertes par le cloud privé sont utilisées seulement par une entreprise ou une organisation, beaucoup plus par les organismes gouvernementaux et les institutions financières, ces dernières peuvent gérer elles-mêmes le cloud privé ou engagé un tiers pour le faire, au même temps le cloud privé peut être protégé de l'accès non autorisé par des tiers, grâce à son niveau de sécurité élevé [62].

1.6.3 Hybrid

Le terme ”hybride” dans ce modèle de cloud fait référence à une combinaison entre le cloud privé et le cloud public, dont la gestion entre eux se fait d'une manière cohérente,au niveau de ce cloud les charges de travail et les données sont réparties entre les cloud privés et publics afin d'accroître la flexibilité et l'efficacité de l'infrastructure. Donc, l'utilisation de cloud public ou privé dépend des besoins de l'entreprise [27].

1.6.4 Communautaire

Le cloud communautaire est un modèle de cloud qui est utilisé afin de satisfaire les besoins particuliers d'une communauté d'organismes ayant des intérêts communs, et qui partagent les même ressources Informatiques telles que les serveurs, [27] les réseaux et les applications, ainsi que le partage des données et des ressources se fait d'une façon sécurisée et rentable grâce à la gestion et au stockage locale des données.

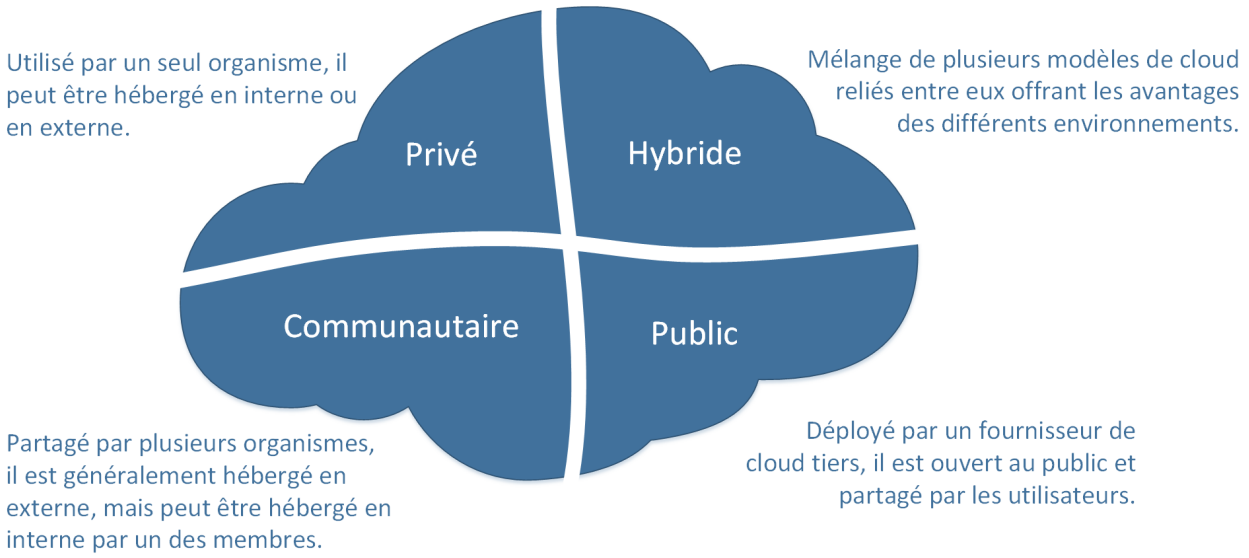

FIGURE 1.3 – Les modèles de déploiement du cloud computing selon NIST [61].

1.6.5 Multi-cloud

Le cloud multiple signifie l'utilisation de plusieurs plateformes cloud de même type (deux ou plus), et donc de plusieurs fournisseurs au même temps, pour stocker, partager et fournir des services et des ressources, grâce à cette technique les entreprises peuvent se profiter des avantages de tous les fournisseurs de services infonuagiques [27][62].

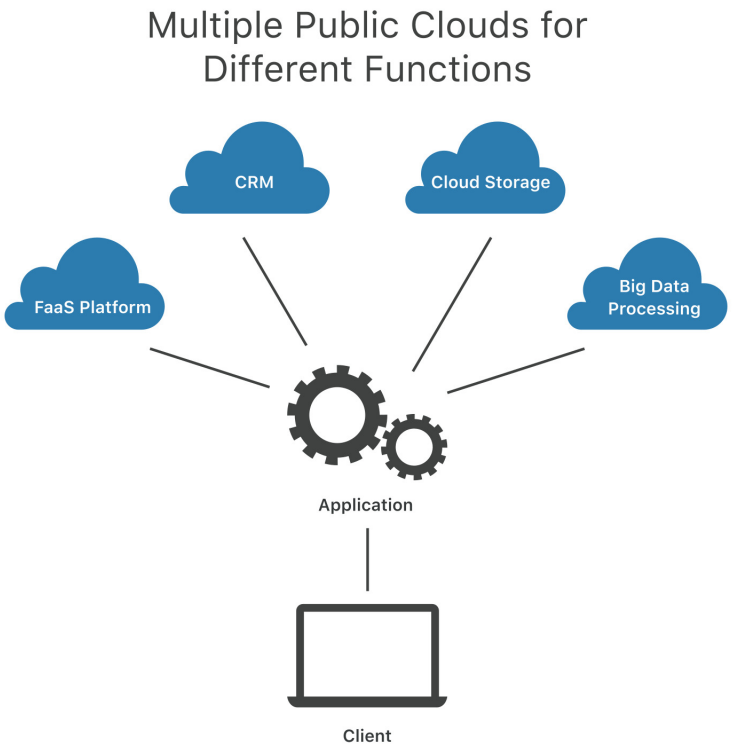

FIGURE 1.4 – Schéma d'une architecture multi-cloud [13].

1.7 Les tendances du cloud computing

Des nouvelles technologies ne cessent de croître rapidement et de disparaître rapidement ci-dessous quelques aspects des récentes technologies du cloud computing :

1.7.1 Cloud edge

L'Informatique de périphérie est la couche réseau qui englobe les terminaux et leurs utilisateurs, par exemple en fournissant une puissance de calcul locale aux capteurs, compteurs ou autres appareils accessibles par le réseau. La périphérie du cloud permet aux appareils distants de traiter les données, par exemple, les données où les serveurs locaux sont proches du réseau favorisent l'avantage de la technologie de transmettre des données extrêmement importantes pour le traitement, réduisant ainsi la latence, permettant des opérations plus efficaces et des temps de réponse plus rapides [37] [6].

1.7.2 Green cloud

Le green computing est la pratique de recherche consistant à utiliser les ressources de manière respectueuse de l'environnement, pour réduire l'impact de l'Informatique sur l'environnement, réduire l'utilisation contraire à l'éthique des ressources et des matières dangereuses, maximiser l'efficacité énergétique pendant le cycle de vie d'un produit ; cela favorisera la biodégradation et le recyclage des produits redondants réduit la perte d'argent et réduit l'émission de dioxyde de carbone. Le green cloud est utilisé pour gérer efficacement l'utilisation de l'infrastructure et réduire la consommation d'énergie [39].

1.7.3 Eternity cloud

Ethernity cloud est une technologie Informatique confidentielle décentralisée en cours de développement. Il repose sur trois piliers importants que sont la confidentialité, l'intégrité et la disponibilité continue. Il est basé sur la technologie blockchain. L'objectif principal est un environnement décentralisé qui permet d'exécuter des logiciels cloud en tant qu'applications cloud décentralisées. Le noeud n'a rien à voir avec l'emplacement, l'autoréplication, la réplication constante sur Internet sans interaction de l'utilisateur, la protection des données contre les activités abusives des fournisseurs de cloud, la garantie d'équité, de décentralisation et d'une véritable confidentialité, la possibilité de fonctionner sur un réseau décentralisé Exécuter des activités de cloud computing tout en gardant une parfaite maîtrise de la confidentialité, de l'anonymat et de la répétabilité du processus, ainsi qu'un faible coût de fonctionnement, afin qu'il soit accessible à tous [36].

1.8 La relation entre le cloud et les technologies émergeantes

Le cloud peut être adapter et rattacher à presque n'importe quel technologie (IT) sois c'est lui qui alimente ou bien c'est les autres qui l'alimente créant ainsi une relation directe entre elles, citant :

1.8.1 Internet of things

L'internet des objets désigne tout objet connecté à internet, et toute technologie opérationnelle basé sur l'internet comme les maisons connectées, les télévisions connectées,etc. Nous les utilisons généralement pour le suivi numérique, le marketing et les applications connectées de santé, et c'est là que le cloud intervient la plupart des appareils sont connectés au cloud pour les tâches de traitement et de stockage de données cela facilite l'accès à distance aux données [19].

1.8.2 Intelligence artificielle

Le marché de l'IA est l'un des piliers les plus utilisés dans les entreprises, afin d'augmenter le niveau de concurrence entre les entreprises, pour être plus complet, elles se tournent vers l'environnement cloud pour mieux gérer les données générées par l'IA ; l'apprentissage automatique repose sur des quantités massives de données, ces données deviennent à leur tour évolutives et accessibles instantanément via le cloud, ce qui signifie un contrôle permanent sur les données, des coûts réduits,etc.

1.8.3 5G

La 5G, une technologie qui accélère le flux d'informations collectées via des capteurs, et le faite de la relier au cloud fournira une combinaison parfaite de vitesse de transmission de données, et de stockage optimal, avec le potentiel de remodeler des systèmes distribués et complexes.

L'essor de la 5G, combiné au cloud, permettra aux entreprises d'obtenir plus facilement les ressources dont elles ont besoin pour démarrer et gérer. L'automatisation 5G a le pouvoir d'organisation, de coordination de systèmes dans plusieurs services de réseau, qu'il s'agisse de téléphonie, de télévision, de serveurs d'entreprise,etc. Quel que soit leur emplacement [2].

La figure suivante represente l'utilisation actuelle et future des technologies liées au cloud d'aprés CSA :

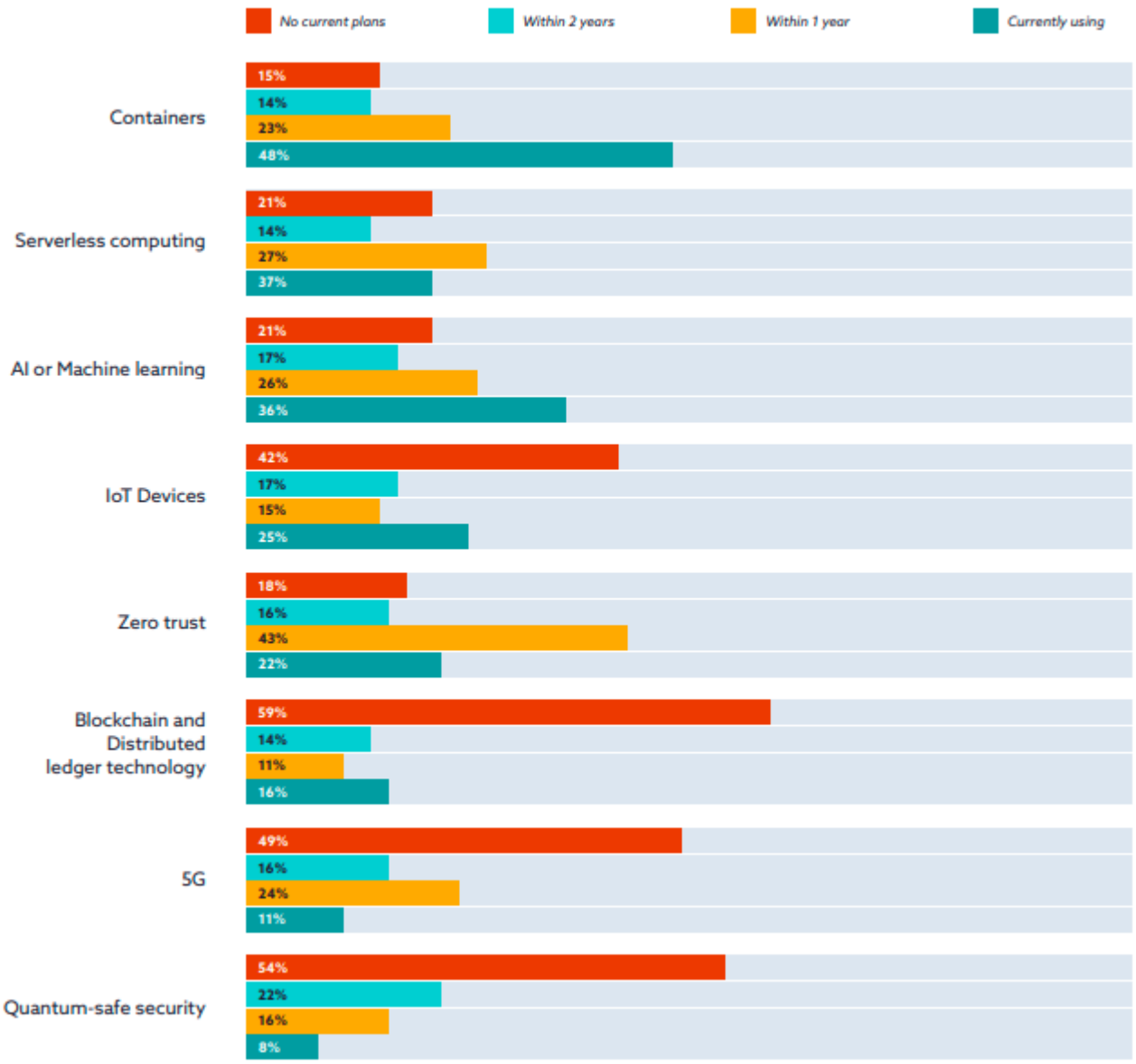

FIGURE 1.5 – Utilisation actuelle et future des technologies liées au cloud d'aprés CSA [9].

1.9 L'impact du cloud computing

Le cloud computing se développe très rapidement, et bien que la trajectoire de croissance du cloud ait été largement anticipée, son impact sur les départements Informatiques et leurs fonctions, structures et stratégies reste relativement flou. De profondes mutations affectent toutes les phases de l'Informatique d'entreprise : planification, approvisionnement, déploiement, exploitation et gouvernance des services Informatiques. L'une des manifestations les plus visibles du changement induit par le cloud est l'émergence de « secteurs d'activité » ou LOB (le domaine dans lequel les RH, les ventes et les autres personnels sont des utilisateurs de services Informatiques), qui sont tous des consommateurs directs de services cloud, ceux qui influencent ce qui est fait pour l'entre-

prise, et la personne qui prend les décisions Informatiques. Le cloud est une source d'innovations technologiques inattendues, dont la plupart sortent du cadre traditionnel de ce que l'on entend actuellement par « service Informatique » [17]. La figure 1.6 illustre l'impact du cloud computing dans l'environnement IT.

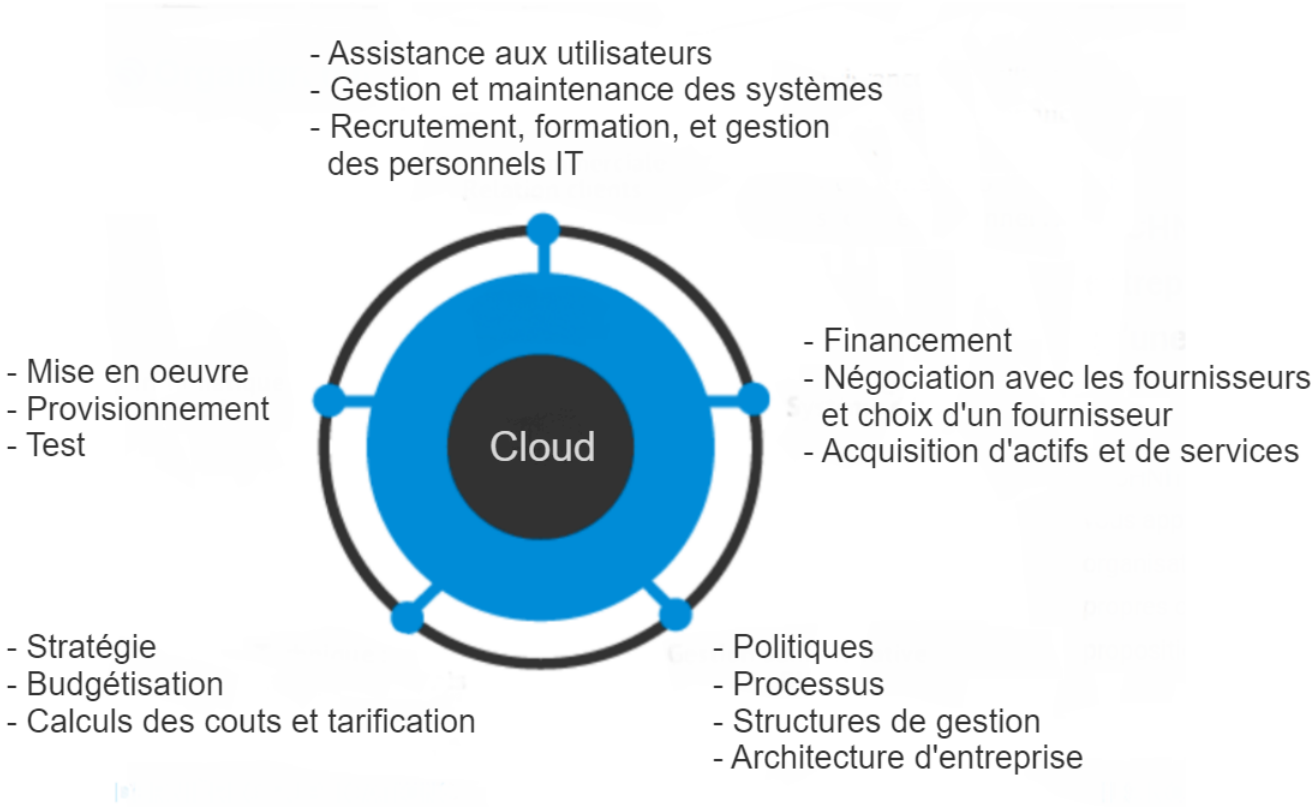

FIGURE 1.6 – L'impact du cloud computing dans l'environnement IT [17].

1.10 Les avantages du cloud computing

Le cloud computing a dénombrable avantages, voici quelques-uns citées ci-dessous :

- Flexibilité et évolutivité : le cloud computing permet aux entreprises de choisir les services qui leurs conviennent parmi celles proposées par les fournisseurs cloud, en conformément à leurs exigences, besoins et de leur budget. Ainsi, il peut toujours s'adapter en fonction de l'évolution des besoins de l'entreprise.

- Accès facile et à distance : Les utilisateurs de cloud peuvent accéder aux services, ressources Informatiques et applications où qu'ils se trouvent dans le monde, à condition de se connecter sur Internet .

- Sécurité des données : Grace aux politiques de sécurité actuelles utiliser dans le cloud, les informations et les données stockées sur les serveurs cloud sont de plus en plus confidentielles et sécurisées.

- Réduction des coûts : cela permet aux entreprises de réduire les coûts d'infrastructure, de maintenance et de soutien des TI, tout en accroissant l'efficacité opérationnelle. Amélioration de l'efficacité opérationnelle : afin d'améliorer la productivité et la rentabilité le cloud optimise les processus et les ressources des entreprises en identifiant des méthodes pour accomplir des tâches plus rapidement, avec des ressources réduites et sans sacrifier la qualité [27].

1.11 Les limites du cloud computing

Malgré les avantages que peut apporter le cloud ; cela ne peut être dépourvus de lacunes :

-Aucun accès hors ligne : L'accès aux ressources cloud dépend totalement d'une connexion internet, au cas d'une interruption de la connexion internet les utilisateurs ne peuvent pas récupérer leurs données.

-Le contrôle : Les ressources Informatique et les données situées au niveau de cloud sont stockées sur des serveurs externes, pour cela les utilisateurs ont moins de contrôle sur leurs propres données.

-La performance en cloud : Des déférents facteurs contrôlent la performance dans le cloud tel que la capacité de stockage qui peut impacter la performance de cloud si y on a suffisamment d'espace pour rependre aux besoins des clients, aussi la latence qui peut affecter la qualité des services.

-Gouvernance Informatique : Suite à l'absence de contrôle sur la passation du marchés, le déclassement et la gestion des opérations d'infrastructure, l'attention portée à l'auto-fabrication dans le cloud peut rendre la gouvernance Informatique difficile ; cela peut nuire à la gestion efficace des risques et de la sécurité Informatique [20].

1.12 Conclusion

Ce chapitre a permis de poser les bases pour la compréhension sur le cloud computing, posant les termes généraux relatifs au cloud dont les objectifs enjeux et défis,etc. Liés au cloud computing, aussi examiné les différents modèles de déploiement et de services offerts par le cloud computing. Dans le chapitre suivant, nous explorerons plus en détail les différents aspects de la confidentialité des données sensibles dans le cloud computing.

Chapitre 2

Étude sur la sécurité dans le cloud computing

2.1 Introduction

La confidentialité des données sensibles stockées et traitées dans le cloud est une préoccupation majeure pour les organisations, qui rencontrent des difficultés pour les sécuriser. La manipulation des données dans le cloud requit un niveau élevé de sécurité pour assurer la confidentialité, notamment des échanges de clés, des accès des utilisateurs,etc. Les fournisseurs de services cloud offrent des mesures de sécurité pour protéger les données de leurs clients.

Les fournisseurs de services cloud ont également des politiques de sécurité strictes en place pour garantir que les données ne sont accessibles que par les personnes autorisées. Les mesures de sécurité telles que la gestion des identités et des accès, les authentifications multifactorielles, la journalisation des activités, les audits de sécurité et les tests de pénétration sont utilisées pour assurer la sécurité des données. Les technologies émergentes telles que l'apprentissage automatique, l'intelligence artificielle et la blockchain [59] sont également utilisées pour renforcer la sécurité des données dans le cloud.

2.2 La sécurité du cloud computing

La sécurité des systèmes, des données et de l'infrastructure basés sur le cloud nécessite un effort collaboratif de règles et de contrôles appelés sécurité du cloud computing. L'objectif de ces mesures est de protéger les données dans les nuages, d'assurer la conformité aux exigences légales, et de préserver la confidentialité des clients. La mise en place de la configuration parfaite pour la sécurité du cloud est essentielle afin qu'elle puisse répondre aux besoins spécifiques de l'entreprise. Les processus de sécurité cloud doivent être complémentaire entre le propriétaire d'une entreprise et le fournisseur de solutions cloud. La méthode pour assurer la sécurité du cloud est finalement déterminée par les solutions ou les fournisseurs de sécurité du cloud spécifiques utilisés [21].Ainsi que ces solutions doivent assurer un bon nombre de points notamment : Garantir la sécurité des données, détecter immédiatement tout événement inhabituel, suivre les événements inattendus et réagir,etc.

2.3 Critéres de la sécurité du cloud computing

Récemment, le cloud computing est devenu le milieu le plus riche en ressources stockées, données sensibles et services, c'est pour cette raison que la sécurité dans le cloud computing représente un grand sujet d'inquiétude [30], et pour atteindre un certain niveau de sécurité de ces ressources plusieurs caractéristiques de sécurité sont recommandées en particulier les quatre suivantes :

.bullet L'authentification dans le cloud computing

L'authentification permet de confirmer l'identité de l'utilisateur ou de l'entreprise, afin de garantir que l'accès aux données sensibles et aux ressources protégées et même aux services cloud se fait seulement par les entités autorisées [56]. Il existe plusieurs types d'authentification, notamment :
- L'authentification basée sur des informations d'identification
- L'authentification à facteurs multiples (MFA)
- L'authentification basée sur des certificats
- L'authentification basée sur des empreintes digitales
- L'authentification basée sur la reconnaissance vocale ou fiscale

• L'intégrité dans le cloud computing :

Signifie la capacité de cloud à assurer la sécurité des données en garantissant que les données stockées n'ont pas été modifiées, altérées, ou corrompues intentionnellement ou accidentellement.

• La disponibilité dans le cloud computing :

La disponibilité est un aspect très important dans le cloud computing, dont elle vise à garantir que les ressources, données et services soient accessibles et opérationnels pour les utilisateurs autorisés à tout moment sans interruption, en dépit des perturbations éventuelles.

• La confidentialité dans le cloud computing :

La confidentialité, est une proprieté pour assurer que les données et les informations stocker ne les divulguer qu'à des personnes autorisées, afin d'assurer la confidentialité dans le cloud computing et proteger les ressources Informatiques, des différentes mesures de sécurité peuvent être appliqué tel que le cryptage, les autorisations d'accès, les contrôles d'accès physique,etc.

2.4 Sécurité des infrastructures (sécurité physique)

La sécurité physique est un pilier de la sécurité du cloud. Il s'agit d'une combinaison de mesures visant à empêcher l'accès direct et l'interruption du matériel hébergé dans le centre de données de votre fournisseur de cloud. La sécurité physique doit inclure une protection complète :

• Contrôle de l'accès au périmètre des locaux et de l'installation :

Des clôtures en métal et en béton recouvrent tout le périmètre, des caméras sont installées et visionnées en permanence et à l'arrivée devant le bâtiment ; une demande d'accès est obligatoire avant d'être autorisée à entrer, les demandes sont contrôlées une par une pour minimiser le nombre de personnes qui en ont accès, et de même elles n'ont accès qu'à la zone pour laquelle leur justification est approuvée avec un nombre limité d'autorisation et un temps déterminé [7].

• Accès des visiteurs :

Pour qu'un visiteur est accès au centre de données faudrait avoir un badge temporaire et qui sera approuvé par l'équipe de contrôle pour donc avoir accès, mais en compagnie d'une escorte qui va leur faire la visite et qui va examiner les actions des visiteurs, si la visite est finie Le badge sera réquisitionné à la sortie de bâtiment [7].

• Contrôle à l'intérieur du bâtiment :

Chaque couche dans le bâtiment est sécurisée notamment l'entrée du bâtiment, l'intérieur du bâtiment et l'étage où se trouvent les données ; des caméras de surveillance sont installées partout et surveillent en permanence chaque côté des serveurs. L'entrée au centre des données est surveillée

par des professionnels de la sécurité ; après l'entrée dans le bâtiment une authentification à deux facteurs est essentielle, avec un passage sous un portique de détection de métaux à l'entrée de l'étage du centre de données ; aussi à la sortie de l'étage faudrait repasser sous le portique, et finalement pour quitter le centre de données faudrait passer par un autre scan de sécurité [7].

2.5 Sécurité logique du cloud computing

Comprends les techniques utilisées et les mesures de la sécurité prises par les fournisseurs et les responsables de sécurité de cloud, afin de garantir l'accès seulement aux entités autorisées, d'assurer la confidentialité et la sécurité des ressources Informatique, données sensibles, services,etc. Ainsi de réduire les risques relatifs à la sécurité Informatique. Pour attendre ces objectives des mécanismes de sécurité logique en cloud computing comprennent l'identification et l'authentification, la gestion des accès aux données, la gestion des mises à jour de sécurité,etc. Sont utilisé [40].

2.6 Contrôle et gestion des flux :

La gestion de flux dans le cloud signifie la façon comment les données et les ressources Informatiques circulent entre les différentes applications et services hébergés dans le cloud. Cela permet d'utiliser efficacement les ressources et d'assurer que les données sont en sécurité. Nous pouvons utiliser des techniques comme l'optimisation du réseau ou la compression pour améliorer ces performances. C'est important pour avoir un système qui fonctionne bien et qui est sécurisé. Une stratégie de gestion des flux dans le cloud peut inclure la mise en oeuvre de pare-feu, de systèmes de détection d'intrusion, afin de protéger les données en transit entre différentes applications et services.

2.7 Dispositifs de sécurité du cloud :

Voici quelques-uns de ces dispositifs qui assure la protection des données des clients dans le cloud computing :

• Pare-feu :

Firewall en anglais, est un système qui permet de protéger un ordinateur ou un réseau, autrement dit, c'est une passerelle de filtrage de paquets qui comprend au moins une interface pour le réseau interne et une autre interface pour le réseau externe, et un ensemble de règles prédéfinies permettront ensuite la mise en place de politiques de sécurité. Les pares-feux sont utilisés pour contrôler l'accès au réseau cloud. Ils bloquent les connexions non autorisées et filtrent le trafic pour protéger le réseau des attaques, et d'assure le relais et le masquage [25].

• Chiffrement :

Le chiffrement est une technique de sécurité qui consiste à rendre les données illisibles par des personnes non autorisées. Le but du chiffrement est de protéger les données confidentielles en les rendant incompréhensibles à toute personne qui tenterait de les intercepter ou de les espionner, sauf pour celui qui a créé le message et le destinataire [70]. Le chiffrement est utilisé dans les applications de sécurité Informatique, y compris la protection des communications en ligne, le stockage des données,etc. Il existe deux types de chiffrement :

- Chiffrement symétrique :

Le chiffrement symétrique, utilise une seule clé pour le chiffrement et le déchiffrement cependant un canal sécurisé doit être mis en place, parmi algorithmes de chiffrement symétrique les plus connus on trouve : DES et le AES.

- Chiffrement asymétrique :

Le chiffrement asymétrique ; utilise deux clés différentes une clé publique pour le chiffrement, et une clé privée pour le déchiffrement, la clé publique est connue de tous mais la clé privée doit rester confidentielle les algorithmes de chiffrement asymétrique les plus connus sont : RSA, El-Gamal.

• Contrôle d'accès :

Le contrôle d'accès est une technique par lequel un système autorise ou interdit le droit à des entités d'accéder à des données, un fichier, applications,etc. Le contrôle d'accès se base surtout sur l'authentification et l'autorisation pour gérer l'accès aux ressources, les fournisseurs de services cloud utilisent des mécanismes d'authentification et de contrôle d'accès pour s'assurer que seuls les utilisateurs autorisés ont accès aux données stockées avec un niveau approprié dans le cloud[70].

• Politique de sécurité :

Il s'agit d'un ensemble de lois, de règles et de pratiques régissant la manière dont les informations sensibles et autres ressources sont gérées, sécurisées et distribuées au sein d'un système particulier, et précise pour chaque sujet les droits d'accès qu'il autorise pour chaque finalité, dans le but de sécuriser les informations, en assurant la confidentialité, l'intégrité et la disponibilité. Les politiques de sécurité peuvent inclure : Les procédures de gestion des mots de passe, les règles de confidentialité et de protection des données personnelles, les règles d'utilisation du système et des ressources Informatiques,etc. Les politiques de sécurité peuvent être divisées en trois grandes catégories ; politiques discrétionnaires (DAC) Exemples : modèle Lampson, modèle HRU. La politique

obligatoire (MAC) qui est souvent utilisée dans l'armée et le gouvernement où la sécurité est un point critique, par exemple : le modèle Bell-La Padula, le modèle Biba. La politique basée sur les rôles (RBAC) ou on associe des permissions pour chaque rôle [38].

• Gestion des leurres :

Les leurres sont de fausses informations délibérément introduites dans une base de données dans le but de tromper ou de bloquer des individus ou des programmes tentant d'obtenir des informations à partir de ces données, généralement pour protéger l'existence d'informations sensibles et pour empêcher les attaquants déjà infiltrés de causer des dommages au réseau, son plus grand L'avantage : puisqu'il ne s'appuie pas sur les signatures d'attaque, il est très efficace pour obtenir une visibilité en temps réel [25].

• IDS :

Un système de détection d'intrusion est un outil de surveillance préventif surveillant toute action malveillante, c'est une partie importante d'une infrastructure de sécurité réseau et réside dans un système Informatique. Ils vérifient le trafic du système ou du réseau pour détecter les activités malveillantes et déclencher des alertes de sécurité en cas d'intrusion. L'IDS surveille tout ou une partie du réseau, ciblant des taux de détection d'attaque élevés et de faibles taux de faux positifs, c'est un outil important pour lutter contre les cyberattaques sur le réseau, aidant à détecter les activités suspectes et à prendre des mesures préventives avant que les attaques ne causent des dommages sur des données importantes [34].

• Honeypot :

Le honeypot est un dispositif de sécurité Informatique conçu pour être un piège avec des vulnérabilité intentionnelles, simulant un service, un système, une application ou une base de données afin d'attirer les attaquants vers une cible imaginaire et de les éloignés des réseaux et des systèmes réels, il est souvent utilisé comme outil de surveillance ou de recherche en sécurité Informatique pour identifier les tactiques des attaquants et mieux comprendre les menaces potentielles. Les informations collectées sont utilisées pour améliorer les stratégies de sécurité et de défense, détecter les nouvelles tendances et les vulnérabilités émergentes, et sensibiliser les utilisateurs et les administrateurs sur les risques encourus en sécurité Informatique [25].

2.8 Menaces liées au cloud computing

Les menaces dans le cloud font référence à tout événement susceptible qui sert potentiellement ou directement à endommager le système, et atteindre des ressources Informatiques telles que les services et les données sans autorisation, soit d'une façon intentionnelle ou accidentelle.

La figure 2.1 nous montre le contexte des menaces dans le Cloud Computing.

FIGURE 2.1 – Le contexte des menaces dans le Cloud Computing [22].

• Les menaces accidentelles :

Ce sont des accidents de sécurité causés par des agissements accidentels ou négligents d'employés, ou d'entrepreneurs,etc. [29]. Ces menaces peuvent provoquer des actions telles que la perte ou la destruction involontaire de données sensibles, l'installation de logiciels malveillants ou la divulgation involontaire d'informations confidentielles [16]. Ces incidents peuvent être dus à des :

- Erreurs humaines.
- Pannes de métastructure et d'applistructure.
- Catastrophes naturelles.
- Incendies involontaires.
- La mauvaise configuration et les erreurs de manipulation.

• Les menaces intentionnelles :

Ce sont l'ensemble des actions malveillantes faite en connaissance de causes, des risques et des conséquences [16][29] nous pouvons citer :

• Les APIs et UI mal sécurisées :

A travers des techniques d'injections les attaquants peuvent facilement exploiter ce genre de faille et exécuter du code malveillant afin d'accéder aux données sensibles.

• Un contrôle insuffisant des identifiants d'accès :

Cela peut survenir lorsque les processus d'authentification et d'autorisation comportent des vulnérabilités. Par exemple, des identifiants d'accès faibles ou faciles à deviner, des mots de passe non chiffrés,etc.

• Les attaques malveillantes :

Une tentative délibérée de porter atteinte à la sécurité ou à l'intégrité d'un système de l'IT, plus précisément l'environnement cloud en utilisant des déférentes techniques d'attaque dont les attaques de reconnaissance, les attaques de session, le phishing,etc.

• Les fuites de données :

Les fuites de données peuvent être causées par divers facteurs, tel que les failles de sécurité, la figure 2.2 montre le résultat d'une étude faite par CSA (Cloud Security Alliance).

Enquête question Rang	Enquête Résultats Score	Nom de la Moyenne
1	7.729927	Insuffisance de l'identification, des justificatifs, de la gestion des accès et des clés, des comptes à privilèges
2	7.592701	Interfaces et API non sécurisées
3	7.424818	Mauvaise configuration et contrôle inadéquat des modifications
4	7.408759	Absence d'architecture et de stratégie de sécurité pour l'informatique dématérialisée
5	7.275912	Développement de logiciels non sécurisés
6	7.214493	Ressources tierces non sécurisées
7	7.143066	Vulnérabilités du système
8	7.114659	Divulgation accidentelle de données dans le nuage/ divulgation
9	7.097810	Mauvaise configuration et exploitation des charges de travail sans serveur et des conteneurs.
10	7.088534	Criminalité organisée/ Hackers/ APT
11	7.085631	Stockage en nuage Exfiltration de données

FIGURE 2.2 – Les 11 top menaces en 2023 selon CSA [16].

2.9 Attaques du cloud computing :

Les attaques sont des techniques malveillantes de cybercriminalité utilisées par les attaquants afin de violer la confidentialité et la sécurité des données sensibles stocker, ou les ressources Informatique en générale ; dans le monde de la sécurité Informatique il existe deux classifications d'attaque chaqu'une a des fins différentes [69].

-Les attaques passives : Ces attaques sont appelées aussi les attaques de reconnaissance car elles visent à obtenir et collecter les informations seulement, sans perturber le fonctionnement de système ou apporter des modifications sur les données, c'est pourquoi ils sont difficiles à détecter. Les attaques passives peuvent prendre différentes formats tel que l'analyse des trafics, le renfilage de paquets, l'espionnage,etc.

-les attaques actives : Pas comme les attaques passives, les attaques actives sont plus dangereuses, dont elles consistent à violer le principe de la confidentialité, de l'intégrité et même de la disponibilité ; ce type attaque sert à détruire le bon fonctionnement de réseaux ou de systèmes cibles, ainsi d'endommager, modifier et supprimer les données sensibles et les entités précieuses.

La figure 2.3 montre la différence entre les attaques passives et actives

	Attaque active	Attaque passive
De base	Une attaque active tente de modifier les ressources du système ou d'affecter leur fonctionnement.	L'attaque passive tente de lire ou d'utiliser les informations du système mais n'influence pas les ressources du système.
Nuire au système	Cause toujours des dommages au système.	Ne provoque aucun mal.
Modification de l'information	Se produit	N'a pas lieu
Menace à	Intégrité et disponibilité	Confidentialité
Tâche effectuée par l'attaquant	La transmission est capturée en contrôlant physiquement la partie d'un lien.	Juste besoin d'observer la transmission.

FIGURE 2.3 – La différence entre les attaques passives et actives [14] .

• Les attaques DDOS (Distributed denial of service :

L'attaque DDOS est une attaque malveillante sert à interrompre et perturber la circulation normale de trafic d'un réseau, service, serveur,etc. Dans le but d'empêcher les utilisateurs légitimes d'accéder à un service, en envoyant une quantité écrasante de trafic au serveur de service. Il existe plusieurs types d'attaque DDOS, les plus communs sont les suivants [35][31] :

- Attaques de la couche application :

Ces attaques cible directement la couche application du modèle OSI, cette couche est chargée de gérer les requêtes et les interactions des utilisateurs avec les serveurs Web ou les applications. Il y en a de nombreux types et modèles, (Exemple : HTTP Flood) .

- Attaques protocolaires :

Les attaques protocolaires aussi appelées attaques d'épuisement d'état, dont elles causent des pannes en consommant trop de ressources serveur et/ou d'équipements réseau tels que des pares-feux et des répartiteurs de charge, (Par exemple SYN flood).

- Attaques volumétriques :

Les attaques volumétriques, sont les attaques les plus courantes des DDOS , ce type d'attaques sont basées sur la consommation de toute la bande passante disponible entre la cible et l'Internet en utilisant généralement une forme d'amplification et même les demandes d'un botnet. Il existe plusieurs types d'attaques volumétriques, notamment les attaques par inondation UDP/TCP/ICMP .

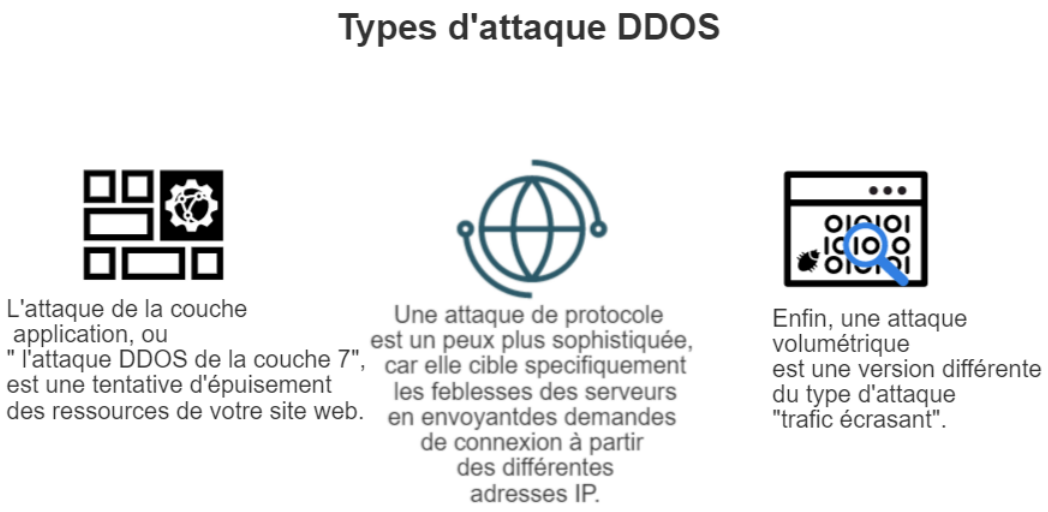

FIGURE 2.4 – Les types d'attaques DDOS [67].

• Les attaques d'injection de code :

Ces attaques consistent a exploiter les failles et les vulnérabilites situe au niveau des serveur et des applications web en insérant des codes malveillants. Ce genre d'attaque permet aux attaquants de perturber les requêtes d'une application dans sa base de données, afin de récupérer les données qu'ils ne sont normalement pas en mesure de récupérer, et aussi provoquer des changements persistants au contenu . Le succès de l'attaque dépend généralement fortement de la base de données sous-jacente et des systèmes interconnectés qui sont attaqués. Les attaques d'injection de code les plus courantes sont l'injection de SQL et l'injection de code JavaScript [12].

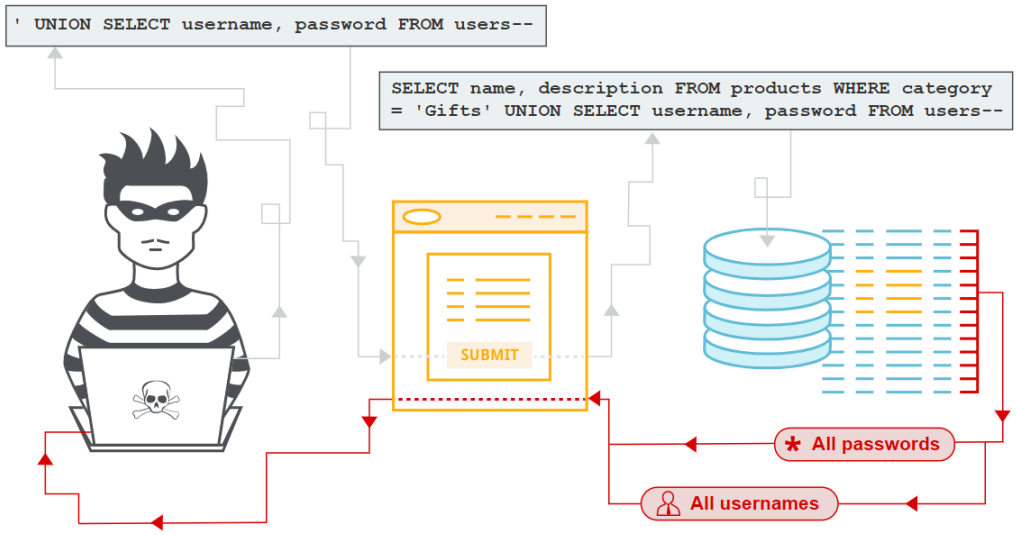

FIGURE 2.5 – Concept d'attaques d'injection de code [43].

• Les attaques de l'homme au milieu (Man-in-the-Middle :

L'attaque de l'homme au milieu est classée parmi les attaques les plus dangereuses dans le domaine de la sécurité Informatique généralement et le cloud computing spécifiquement, cette attaque est généralement utilise pour enregistrer et analyser les données, afin de collecter des informations sur la cible, son principe est d'intercepter les communications échangées entre deux entités sans qu'ils savent que leurs trafic a été lu, modifie, remplacer ou même supprimer [57].

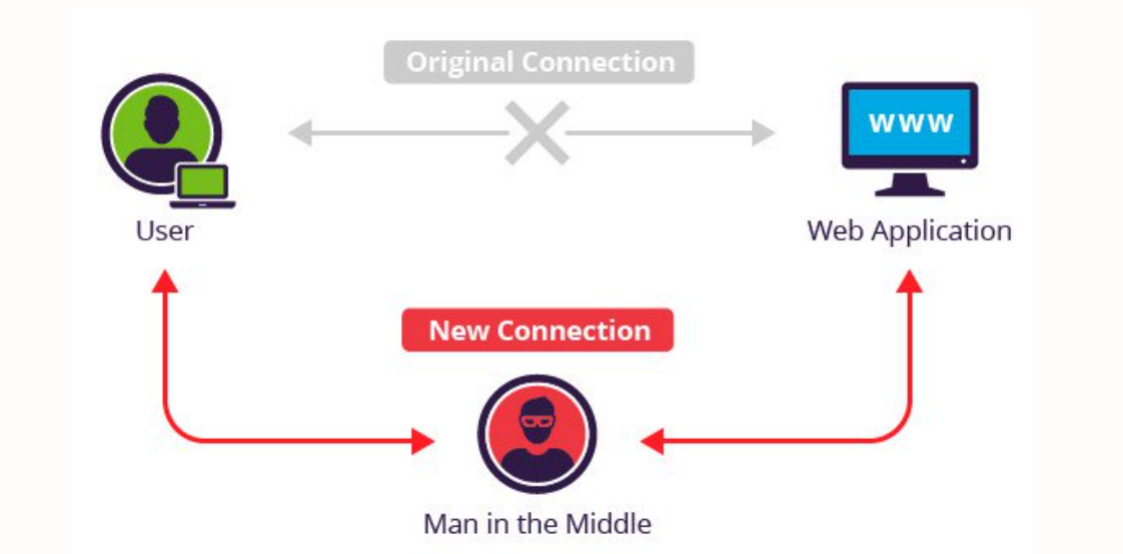

FIGURE 2.6 – L'attaque Main In The Middle [55].

• Les attaques hameçonnage ou phishing :

Ces attaques peuvent être définies par des tentatives d'obtention et de divulgation des informations et des données sensibles qui appartient aux entités victimes, en trompant ces dernières à travers des messages, courriels et fausses pages identique aux pages réels pour qu'ils fournissent leurs informations personnelles tel que les informations sur les comptes bancaires, les mots de passe,etc. Ensuite les utilisées pour des intérêts malveillants [52].

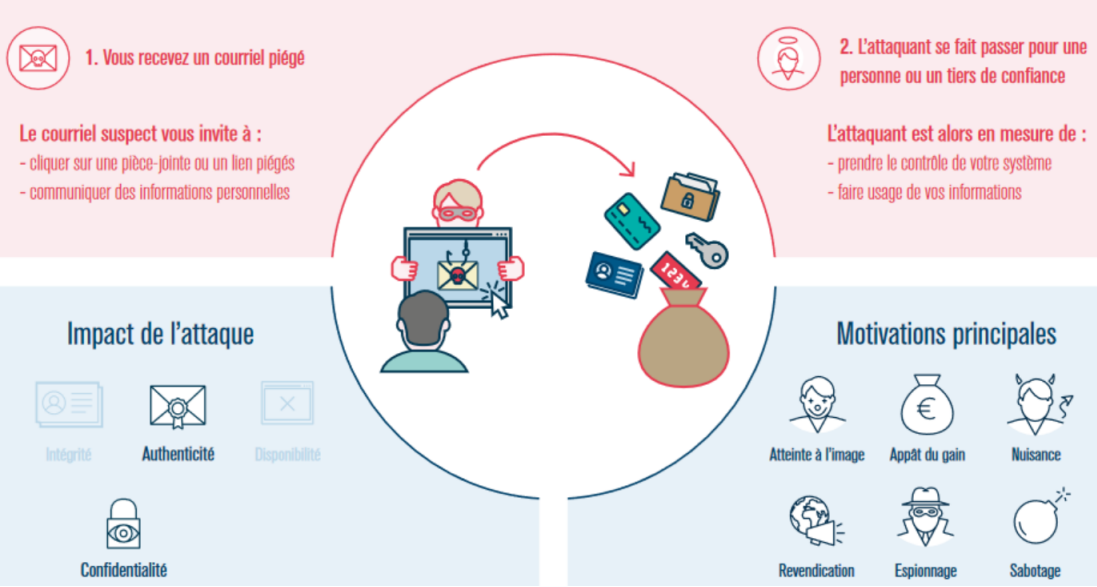

FIGURE 2.7 – Schéma d'une attaque d'hameçonnage [51] .

• L'attaque zéro day :

Cette attaque est aussi classer parmi les plus dangereuses cyberattaques en sécurité Informatique, car elle peut compromettre, des réseaux, des serveurs ou des applications sans être détecter, en exploitant des failles ou des vulnérabilités qui ne sont pas encore connues, jusqu'à ce qu'elle soient exploitées [66].

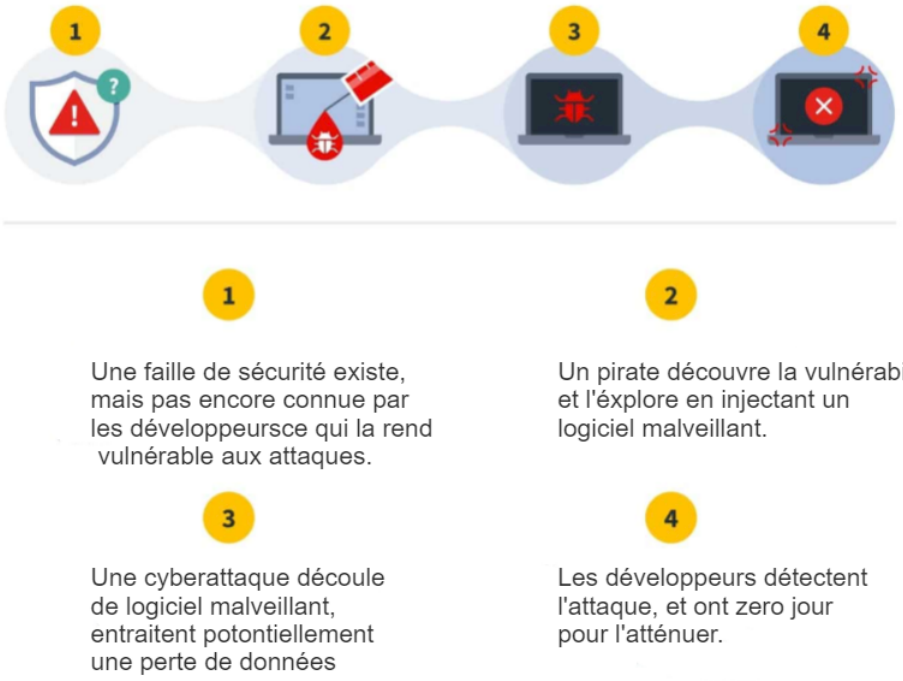

FIGURE 2.8 – La procédure d'une attaque Zéro Day [50].

2.10 Défis de la sécurité cloud

Parmis les principales inquiétudes des fournisseurs cloud [48], nous trouvons :

- la surveillance de nouvelles vulnérabilités.
- la mauvaise configuration de plateforme.
- la conformité légale .
- La violation de la confidentialité des données .
- la défense contre les malwares.
- la fuite de données.

2.11 Revue de la littérature

Ci-dessous une étude sur les travaux antérieures qui caractérise et vise à assurer la sécurité des données dans le cloud plus précisément le critère de la confidentialité dans le cloud computing :

2.11.1 Data Confidentiality Using Improved Security Approaches in cloud Environment

Les auteurs dans cet article [44] ont présenté une technique cryptographique, de sorte que le mécanisme de récupération et de stockage des données dans le cloud inclue le chiffrement par flux chacha20, avec l'utilisation du processus serveur cryptographique (cs) qui assure la gestion des clés, cryptage, décryptage et les droits d'accès des utilisateurs, l'échange de clés se fait avec le protocole Diffie- Hellman amélioré(EDHKEP). Le fonctionnement du mécanisme reste simple le propriétaire des données envoie une demande de téléchargement de fichiers à l'auditeur, ce dernier crypte le fichier avec Chacha20 puis le stocke sur le serveur cloud ; quand l'utilisateur demande ce fichier il sera vérifié par l'auditeur et il va générer la clé avec Diffie- Hellman, l'utilisateur reçoit alors le fichier crypté avec la clé, au final le fichier sera décrypté par l'utilisateur avec chiffrement par flux chacha20 . Cependant et étant donné que l'algorithme chacha20 est un algorithme symétrique faudrait que les canaux de transferts soient constamment sécurisés afin d'assurer la confidentialité des données en transit.

2.11.2 A Confidentiality Scheme for Storing Encrypted Data through Cloud :

Les auteurs dans cet article [64] ont présenté une approche afin de protéger les données contre les attaques malveillantes, en garantissant que même si un serveur est compromis il ne pourra pas accéder aux données stockées, ce schéma repose sur la distribution de clés de chiffrement entre différents serveurs du cloud, les données seront d'abord chiffrées d'une façon asymétrique en utilisant une clé publique, les données chiffrées seront ensuite devisées en parties, les chiffrées ensuite à travers des clé de chiffrement générer aléatoirement par l'utilisateur, l'ensemble de ces clés sera aussi chiffrés en utilisant la clé publique utilisée lors de premier chiffrement, enfin les données seront envoyées au stockage, et clés seront distribuées entre différents serveurs. Pour accéder aux données l'utilisateur télécharge d'abord l'ensemble des clés de chiffrement chiffrées, les déchiffrées ensuite en utilisant la clé privée, afin de déchiffre au final les données besoin à l'aide des clés déchiffrées. Néanmoins, le schéma peut également être vulnérable aux attaques par interception de clé de chiffrement. Si un attaquant parvient à intercepter l'ensemble des clés de chiffrement chiffrées, générer par l'utilisateur, les déchiffre, il peut être en suit en mesure de déchiffrer les chiffrées distribuées, et donc les données associées.

2.11.3 Modified Identity and Broadcast Based Encryption Scheme to Secure Cloud

Les auteurs dans cet article [49] ont présenté une méthode IBE basée sur l'externalisation de l'Informatique, ainsi un schéma qui écarte les clients malveillants est mis en oeuvre. Cette étude consiste en une méthode de cryptage intégrée basée sur l'identité et la diffusion IBEE visant à maintenir la confidentialité en attribuant une autorisation restreinte aux données utilisateurs stockées sur le cloud. L'approche proposer vise à modifier l'IBE standard (identity based encryption) en ajoutant deux phases pour assurer la sécurité. Premièrement l'IBE de base se compose de quatre modules comme suit module de configuration, extraction et génération, cryptage et enfin décryptage. L'approche proposée qui modifie cette dernière est la suivante, elle est composée de cinq modules : module de configuration au niveau de la racine, module de configuration au niveau feuille, module d'extraction, cryptage et enfin le décryptage. Cette approche cherche à améliorer le niveau de sécurité de la méthode IBE et maintenir la confidentialité avec autorisation restreinte aux données des utilisateurs. Si un tel scénario se produit par exemple un attaquant a usurpé l'identité d'un utilisateur et a commis un déni de service qui inonde dans un court laps de temps pourra causer des faux positifs, et puisque l'identifiant sera mis dans la liste révoquée quand le vrai utilisateur voudra avoir accès à ses propres données, l'accès sera refusé causant ainsi l'indisponibilité des ressources.

2.11.4 A Survey on Privacy Inference Attacks and Defenses in Cloud Based Deep Neural Network

Les auteurs dans cet article [71] s'orientent vers la planification d'une autre stratégie de sécurité de sorte qu'ils ont cité les quatre plus grandes attaques qui cible la confidentialité des données dans le cloud basé sur les DNN (Deep Neural Network) : attaque par inférence l'appartenance, attaque par inférence de propriété, attaque par inversion de modèle et attaque par extraction de modèle. Les approches de défense contre ces attaques se résument en quatre solutions : La confidentialité différentielle, visant à obscurcir les propriétés intermédiaires du modèle en ajoutant du bruit aléatoire pour garder les données privées ; l'apprentissage automatique contradictoire, visant à introduire le jeu min-max dans le processus de formation pour améliorer la robustesse du modèle ; le tatouage numérique, visant à intégrer des fligrants comme régulateurs pour la vérification de la propriété du modèle ; les techniques cryptographique, visant à chiffré le modèle cible et les données de requête. Par ailleurs des défis sont soulevés tel qu'un protocole de calcule efficace et mixte pour les applications d'apprentissage automatique, apprentissage collaboratif contre les participants malveillants, principe de conception de l'utilisation d'algorithmes ML (Machine Learning) pour atténuer les attaques d'inversion de modèle, concevoir une protection de la vie privée personnalisée pour l'utilisateur final. Par conséquent le but c'est d'empêcher la révélation des connaissances sensibles, des paramètres du modèle et des échantillons de requête de l'ensemble des

données de formation, tout de même le compromis majeur qui caractérise ces différentes approches c'est qu'il faut toujours chercher l'équilibre de Nash entre confidentialité, utilité et efficacité.

2.11.5 Privacy Preserving Data Sharing in Cloud Using EAE Technique

Les auteurs dans cet article [33] ont proposé une solution appelée "Enhanced elliptic curve Attribut based Encreption", pour assurer un transport confidentiel des données, et pour reduire le pourcentage des fausses recherches à retourner . L'EAE est basé sur l'utilisation de l'ECC (Elliptical Curve Cryptography), une méthode de cryptage par attributs dont le chiffrement est asymétrique et qui se compose de deux phase cryptage et décryptage. Lors de cryptage, l'ECC choisit deux nombres aléatoirement 'P' et 'Rk' de la gamme [1,R-1], pour créer une clé publique 'K' qui va être utilisée pour générer deux messages chiffrés afin d'avoir le texte chiffré final 'Ci'. l'EAE utilise une fonction de hachage pour chiffrer le texte et le stocker dans le cloud ; la valeur de hachage se régénère lors de chaque accès aux données, les deux valeurs seront comparer du coté de récepteur si les valeurs sont égales l'utilisateur reçoit les données demandées. Cependant, cette technique peut être vulnérable aux attaques par canaux auxiliaires, où un attaquant peut utiliser des informations sur la consommation de puissance, la durée d'exécution et d'autres caractéristiques pour déterminer les clés secrètes utilisées pour chiffrer les données.

2.11.6 A Confidentiality-based data Classification-as-a-Service (C2aaS) for cloud security

Les auteurs dans cet article [1] ont proposé une nouvelle technique de classification afin de séparer les données confidentielles et publiques, et alors assurer la confidentialité des données sensibles dans le cloud, le concept de cette classification (C2aaS) est de traiter les données dynamiquement selon leurs niveau de sécurité, en séparant les données non-confidentielles (publiques) et confidentielles (sensibles) en mettant une politique de sécurité pour les données classer en tant confidentielles. Cette classification fonctionne sous un ensemble des machines virtuelles (VMs) gérer par un composant essentiel appelé hyperviseur ou moniteur des machines virtuelles, en utilisant un simulateur CloudSim, la phase de classification fonctionne à l'aide de l'algorithme KNN (K plus proches voisins) sémantique qui se base sur la classification des attributs des données après une prédiction d'une liste des attributs confidentiels et non-confidentiels, dont chaque donnée sera associée à la classe de son attribut. Au final seules les données sensibles qui seront cryptées et sécurisées avant de les stocker, par contre les données publiques seront envoyées directement pour stocker dans le cloud, et donc la politique de sécurité sera choisie en fonction de la sensibilité des données.

Cependant, le KNN sémantique n'assure pas une classification idéale à 100% , ce qui signifier que parmi les données publiques non sécurisées l'attaquant peut accéder à des informations

sensibles mal classifiée sans aucun effort et même d'accéder à des informations simples mais qui peuvent être utilisé pour atteindre des informations plus importantes.

2.11.7 Privacy Preserving using Enhanced Shadow Honeypot technique for Data Retrieval in Cloud Computing

Les auteurs dans cet article [58] ont proposé une version améliorée de l'architecture shadow honeypot appelée l'ESH (Enhanced Shadow Honeypot), intégrée à un réseau en tant que système de leurre, avec l'utilisation d'une base de données KDD99. Au cas d'une attaque un processus d'organisation est utilisé pour définir le type d'attaque en comparant les informations enregistrées sur cette interaction lors de l'attaque à l'aide d'un classificateur CNN, d'autre part le système établit une connexion avec le serveur, sachant qu'un IDS est implémenté au pot de miel, et mutualisé avec le pare-feu, cette collaboration rend la détection de l'attaque plus accessible.

Avant de stocker les données dans le Cloud un chiffrement sera appliqué sur les données, générer ensuite une clé hachée, lorsque l'utilisateur demande ses données, ce dernier doit s'authentifié pour les exigences d'authentification, les données seront déchiffrées et mises à la disposition de l'utilisateur si le haché est authentifié. Tout ça dans le but de stationner les attaques en premier temps et donc assurer la confidentialité des données. Néanmoins, l'emplacement de shadow honypot représente un grand danger sur les données, les hackers inventent des nouvelles attaques à chaque fois l'une de ces dernières pourra contourner ce protocole d'une façon indétectable, vu que le jeux des données KDD99 ne pourra jamais contenir toutes des nouvelles attaques, donc au cas où un attaquant arrive à dépasser le honypot les hachés des mots clés des données ne seront pas suffisants pour assurer leurs confidentialités, sachant que y en a plusieurs outils pour casser le chiffrement par hachage dont 'la table Rainbow',etc.

2.11.8 Framwork for protecting the confidentiality of outsourced data on cloud

Les auteurs dans cet article [23] ont proposé une architecture bi-cloud pour but d'assurer la confidentialité des données externalisées, confidentialité des requêtes émises et confidentialité des modèles d'accès aux données. l'idée est d'intégrer le travail de différentes cloud (cloud traitement et cloud de stockage) de sorte que ni l'un ni l'autre ni l'intrus n'accède aux données stockées d'ailleurs ils ont utilisé le chiffrement homomorphe de paillers qui permet d'analyser les données sans qu'elles soient déchiffrées auparavant. L'architecture est divisée en quatre schémas différents chacun a ses propres points forts et point faible mais assurant tout de même la confidentialité des données externalisées, c'est le type d'application qui détermine le schéma à choisir ; néanmoins cette technique comporte un inconvénient malgré cet environnement complexe les données pourraient être mieux sécurisé vu que dans cette approche les données ne sont pas classifiées et supposer qu'elles aient toutes le même niveau de sécurité.

Le tableau 2.1 ci-dessous montre une comparaison entre les différentes approches déja cité dans l'état de l'art :

	Sécurité			Faisabilité		Réglementation
	integrité	confidentialité	disponibilité	applicabilité	facilité d'utilisation	conformité
Articles						
A confidentiality-based data classification-as-a-service(C2aaS) for cloud security	-	0	+	-	+	-
Privacy Preserving Data Sharing in Cloud Using EAE Technique	+	0	-	+	0	+
Privacy Preserving using Enhanced Shadow Honeypot technique for Data Retrieval in Cloud Computing	+	0	+	0	+	+
A Confidentiality Scheme for Storing Encrypted Data through Cloud	+	+	0	+	-	+
CCSC-DHKEP :Data Confidentiality Using Improved Security Approaches in cloud Environment	0	-	+	+	+	+
Modified Identity and Broadcast Based Encryption Scheme to Secure Cloud	+	+	-	+	+	+
A Survey on Privacy Inference Attacks and Defenses in Cloud Based Deep Neural Network	-	+	-	0	-	+
Framwork for protecting the confidentiality of outsourced data on cloud	+	+	0	+	-	-
NB : (+) high ; (-) low ; (0) medium .						

TABLEAU 2.1 – Tableau de comparaison entre les approches cloud.

2.12 Conclusion

Pour conclure, ce chapitre a mis en évidence les techniques proposées par les chercheurs en matière de la gestion des risques de sécurité, notamment la protection de la confidentialité des données sensibles dans les environnements de cloud computing, ainsi que les défis et les limites de ces recherches. En tenant compte de cela, notre recherche vise à explorer des solutions innovantes pour améliorer la confidentialité des données dans le cloud. Nous avons examiné et étudié les approches proposées pour assurer la confidentialité les données dans le cloud, afin de présenter une architecture améliorée qui permet de s'attendre à un degré élevé de confidentialité en protégeant les données contre les violations de sécurité, les failles de confidentialité et les accès non autorisés. En contribuant à combler ces lacunes, notre recherche est basé sur l'utilisation de différentes techniques et technologies récentes notamment en matière de contrôle d'accès ce qui permettra d'offrir une meilleure protection des données pour les entreprises et les utilisateurs comme elle prendra en compte leurs besoins spécifiques.

Chapitre 3

Notre Contribution

3.1 Introduction

Le cloud computing introduit de nombreux défis en matière de la sécurité des données dont la confidentialité. Dans le présent chapitre nous allons présenter une solution que nous avons développée pour répondre aux problèmes identifiés dans le chapitre précédent ; nous avons constaté que les données non classifiées et celles classifiées en seulement deux catégories représentent un impact négatif sur la confidentialité des données, et aussi l'utilisation d'un honeypot placé sur le même réseau que le cloud est concédé comme un grand risque sur la sécurité de ces données. Afin d'assurer la confidentialité et la sécurité de ces données nous allons présenter une architecture nommée ConfidCloud+ qui assure la confidentialité grâce à l'utilisation de multi-cloud, au même temps joue un rôle passive grâce à l'environnement de honeypot contribuant à prévenir des éventuelles futures fuites et violation de la confidentialité.

3.2 Motivation

Dans un monde de plus en plus numérique, les particuliers, les entreprises et les organisations génèrent de plus en plus de données. Cependant, la gestion de ces données présente de nombreux défis, notamment en ce qui concerne la confidentialité des données stockées dans le cloud. En fait, le stockage de données dans le cloud peut constituer une menace pour la vie privée car les données sont stockées sur des serveurs appartenant à des tiers et peuvent être consultées par des tiers non autorisés. Cela pourrait inclure des employés malveillants de cybercriminels, des sociétés de cloud computing et même des gouvernements. Ces menaces peuvent compromettre la vie privée, nuire à la réputation d'une entreprise et entraîner des pertes financières considérables. Dans l'ensemble, les entreprises manquent de confiance dans leur capacité à protéger les données dans le cloud, dont les études scientifiques ont prouvé que plus de la moitié des entreprises 57% déclarent que la confiance est faible à moyenne. Cette méfiance devient encore plus prononcée lorsqu'il s'agit de données sensibles, 40% des entreprises déclarent que 50% où moins des données sensibles sont adéquates dans le cloud suffisamment de sécurité, et seuls 4% ont déclaré que la sécurité était

suffisante pour sécuriser 100% de leurs données cloud. Cela montre que les entreprises éprouvent des difficultés lorsqu'il s'agit de la protection des données sensibles plus que les données à moins sensibilité [8].

Ci-dessous une figure qui montre le pourcentage des données sensibles sécurisées :

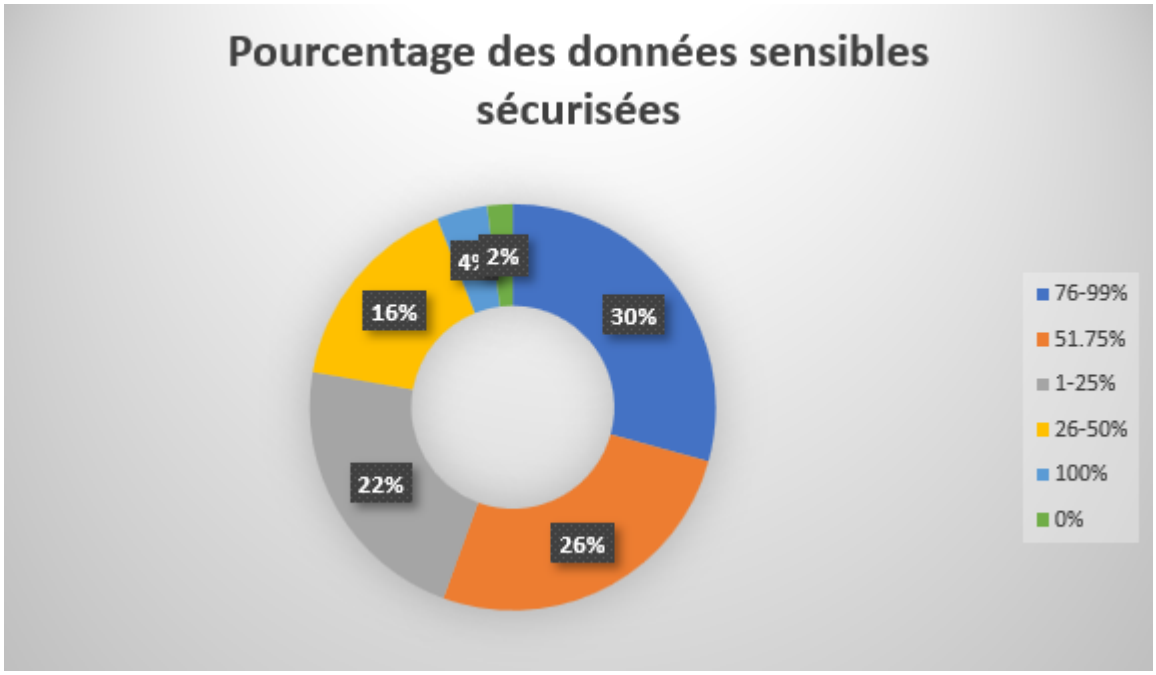

FIGURE 3.1 – Pourcentage des données sensibles securisées [18].

Dans ce contexte, la mise en oeuvre d'une approche efficace de confidentialité des données dans le cloud est cruciale pour la protection des données. Les approches de confidentialité existantes dans la littérature assurent un bon niveau de sécurité, néanmoins présentent quelques limites. Par conséquent, notre approche vise à assurer la confidentialité des données dans le cloud qui soit à la fois innovante et efficace tout en répondant aux exigences réglementaires et légales en matière de confidentialité des données.

Notre approche ConfidCloud + vise à tirer profils des points positifs des précédentes approches [23] [1] [58],tout en évoquant cet ensemble de motivation à savoir : Q1 : Comment prévenir les fuites de données avant qu'elles ne se produisent ? Q2 : Comment assurer un niveau élevé de confidentialité de données en transit et en stockage ? Q3 : Comment éloigner les suspectes des données sensibles ?

3.3 Problématique et objectifs

Tout en s'efforçant de protéger les données sensibles dans le cloud, les entreprises ont également du mal à suivre ces dernières. En effet la confidentialité de ces données est un sujet de préoccupation majeure pour les utilisateurs de cloud, particulièrement lorsqu'il s'agit de stocker des données sensibles notamment les renseignements personnels ou les données financières, dont les atteintes à ces données peuvent avoir des effets dévastateurs. Les approches actuelles employées pour garantir la confidentialité des données dans le cloud représentent de multiples défis, parmi lesquels l'utilisation de bi-cloud, l'inconvénient de cette technique est la non-classification des données, le faite qu'elles sont stockées dans un endroit unique, constitue un risque sur les données dans le cas où l'attaque à réussit à casser la structure du chiffrement utilisé ; voire même la classification des données en seulement deux niveaux confidentielles chiffrées et non confidentielles publiques, l'attaquant peut attendre facilement une donnée sensible si elle s'est mal classifiée ; nous citons aussi le cas d'utilisation d'un honeypot, au même niveau que le cloud ; l'emplacement de cette architecture présente un potentiel de dangers significatifs pour les données sensibles, où le hachage constitue l'unique moyen de protection à leur disposition.

Cette étude vise à évaluer quelques approches de protection de la confidentialité dans le cloud et à proposer une nouvelle approche dont l'objectif est :

- Explorer l'utilisation de multi-cloud afin de séparer le cloud de et le cloud de stockage dans le but d'optimiser la performance et de réduire les risques de défaillance du système.

- Classer les données en trois niveaux, chacun sécurise d'une façon déférente dans le but de renforcer la sécurité des données sensibles, et éviter au même temps les fuites de données au cas d'une erreur de classification.

- Utilisation d'un honeypot contenant des données leurres et l'implémenter avec un système de détection d'intrusion pour empêcher l'accès des attaquants aux systèmes réels, et détecter les tentatives d'attaque sur l'honeypot d'une manière instante, ce qui mène à détecter les données les plus visées, et donc renforcer leurres sécurité en appliquant de nouvelles mesures de sécurité selon leurs niveaux de sécurité précédent.

- Avoir un coup d'avance sur les potentielles attaques future grâce à l'utilisation de l'architecture honeypot.

3.4 Notre contribution "ConfidCloud+"

Dans cette section, nous allons présenter notre solution en mettant en évidence les points essentiels de la manière suivante.

3.4.1 Architecture proposée

Ci-dessous un schéma représentant notre approche ConfidCloud+ :

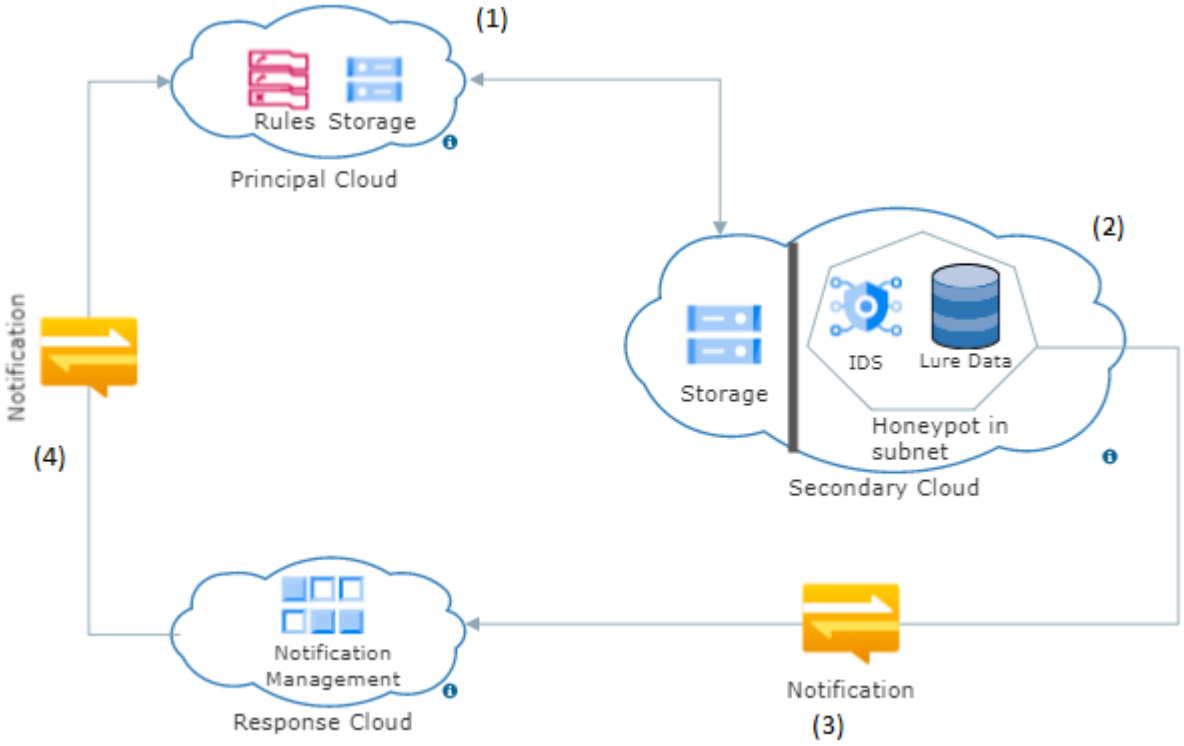

FIGURE 3.2 – Schéma d'architecture de ConfidCloud+.

3.4.2 Déroulement des actions

Nous présentons dans cette section un scénario d'application de notre architecture ConfidCloud+ : quand un attaquant décide d'altérer ou d'intercepter des données dont l'accès est interdit, il accède direct à l'environnement, désirant récupérer ou identifier des failles, il va délibérément découvrir la faille intentionnelle de l'honeypot, tout en asseyant de voler les données voir (1).

Dans le cas où l'attaquant arrive à retrouver la base de données leurre mises dans l'honeypot (qui est pour l'attaquant une base de données cible), dès qu'il essaie d'accéder à ces données, à ce moment une notification est envoyée immédiatement au cloud d'intervention voir (2) comprenant le nom de la table et le nom de la colonne (exemple : table employée, colonne numéro de carte de crédit), le cloud d'intervention retransmet cette notification vers le cloud principal voir (3) afin de mettre en oeuvre les mesures de sécurité adéquates par rapport au niveau de sécurité des données.

Dans le cas contraire, ou l'attaquant n'est pas arrivé à exploiter la faille, les données restent protégées et la confidentialité est assurée ; qu'ils réussissent ou ils ne réussissent pas ça n'affecte en aucun cas le cloud de stockage puisque l'honeypot est isolé dans le sous-réseau du cloud secondaire.

3.4.3 Différents modules de l'architecture

Cette partie vise à présenter les différents environnements utilisés lors de la création de notre approche ConfidCloud+.

3.4.3.1 Environnement multi cloud

L'intégration du multi-cloud est considérée comme une composante essentielle de notre approche. Ainsi, nous présenterons les différents cloud utilisés dans notre stratégie.

• Cloud principal :

Le cloud principal fournit le service de stockage dont les bases de données sont Chiffrées, Classifier et fragmenter comme suit :

- Le cryptosystéme de paillier :

Le chiffrement homomorphe de Paillier qui une extension du cryptosystème Paillier [46] permettant d'effectuer des opérations mathématiques sur des messages chiffrés sans les déchiffrer au préalable. Ceci est possible grâce à la propriété d'homomorphisme qui permet des opérations sur des nombres chiffrés tout en préservant la confidentialité des données.

Le processus de chiffrement dans le cryptosystème homomorphe de Paillier utilise une propriété mathématique appelée homomorphisme additif. Le processus de décryptage dans ce dernier utilise également la clé privée pour le calcul inverse afin de restaurer les informations d'origine.

Cependant, en raison de la propriété d'homomorphisme, il est possible d'effectuer des opérations mathématiques sur des messages chiffrés et d'analyser les données compromises.

- Classification par régression multinomiale logistique :

Avant de stocker ces données une classification en trois niveaux sera effectuée sur ces dernières, en utilisant la méthode de classification par régression logistique multinomiale [11], dont les niveaux de classification sont définis comme suit :

top confidential (Niveau 1) : Les données associées à cette classe sont les données les plus sensibles qui nécessitent un niveau de confidentialité extrêmement élevé par rapport aux deux autres niveaux. Les données de cette classe sont chiffrées et fragmentées en trois fragments en utilisant l'algorithme de Shamir secret sharing.

High confidential (Niveau 2) : Cette classe contient les données avec un niveau de confidentialité élevé, dans cette optique, un double chiffrement sera mis en place pour garantir leur sécurité.

Médium confidential (Niveau 3) : le niveau trois est réservé aux données non pas faibles, mais moins confidentiel par rapport aux autres, puisque la relation entre des données basiques peut s'avérer intéressante entre les mains des tiers non autorisés pour une ultérieure exploitation, dans

le but de protéger ces données ils seront chiffrées par un chiffrement homomorphe plus précisément le chiffrement de Paillier [46].

Voici une illustration présentant la classification des données :

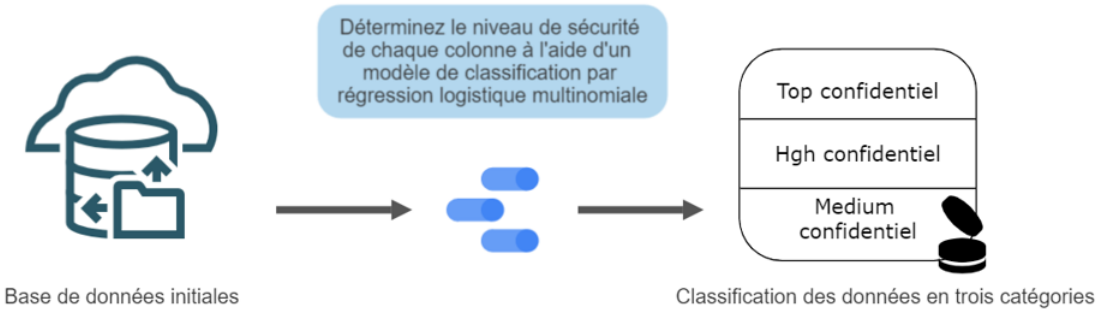

FIGURE 3.3 – Classification des données par régression logistique multinomiale.

Lorsqu'une notification est reçue des mesures adéquates sont appliquer aux données réelles en suivant une structure arborescente où le principe consiste à appliquer les mesures de sécurité du niveau supérieur aux données du niveau actuel. Les informations recueillies peuvent être utilisées pour améliorer la sécurité du système et prévenir des futures attaques. Les mesures à appliquer dépendent du niveau de confidentialité de la donnée.

Le tableau 3.1 illustre notre proposition pour une éventuelle politique de mesures de sécurité :

Niveau de sécurité de la donnée	Mesure avant infiltration	Mesure après infiltration
Top confidentiel	Fragmentation	Restriction d'accès temporel
High confidentiel	Double chiffrement	Fragmentation
Medium confidentiel	Chiffrement	Double chiffrement

TABLEAU 3.1 – Représentation des mesures appliquées aux données.

- Fragmentation des données sensible dans le cloud :

Une fragmentation des données sur celles classées tops confidentiels est appliqué afin d'apporter plus de confidentialités aux données, et que l'accès à un seul fragment ne sera pas nuisible. Nous avons opté pour l'algorithme de Shamir secret sharing [47], qui est une méthode de fragmentation de données permettant de diviser une donnée en plusieurs fragments et les distribuer entre plusieurs parties prenantes ; dans notre cas deux fragments seront stockés dans le cloud principal et uns dans le cloud secondaire. Cette méthode est particulièrement utile pour la sécurité des données sensibles dans le cloud, car elle permet de diviser une donnée en plusieurs fragments et de les stocker sur différents serveurs cloud, ce qui rend plus difficile la reconstitution des données complètes par des tiers non autorisés .

• Cloud secondaire :

Le Cloud secondaire est un cloud à deux fonctionnalités majeures, la première c'est qu'il va stocké une seule partie des fragments des données top confidential et les deux autres sont stockés dans le cloud principal, le fragment stocké dans ce cloud est comme une « garantie » ; disant qu'un attaquant réussisse à compromettre le cloud original et a réussi même ont retrouvé les deux fragments de la donnée cible, il pourra jamais retrouver la donnée entière puisque le troisième fragment lui serait nécessaire afin de retrouver la donnée complète. De plus le fait d'avoir mis qu'un seul fragment dans le cloud secondaire cela veut dire que même si l'honeypot vient à être exploité par l'attaquant à des fins malveillantes, il ne pourrait obtenir qu'un seul fragment au maximum. La deuxième fonctionnalité c'est qu'il va héberger l'honeypot, la base de données leurre et l'IDS dans son sous-réseau, le rôle de ce cloud est très important, c'est lui qui nous permettra d'anticiper des attaques futures et donc de prévenir des risques de fuite des données à des tiers non autorisé et d'assurer la confidentialité des données qui peuvent être des cibles d'attaques.

• Cloud d'intervention :

C'est un cloud supplémentaire ajouter à l'architecture ConfidCloud+, son rôle est de transmettre la notification reçue de la part de l'IDS vers le cloud principal afin d'assurer un transport sûr de cette dernière, de plus pour protéger l'emplacement du cloud principal et la confidentialité des données qui y sont stockés ; le cloud d'intervention sert aussi à éliminer tout contacte directe possible entre le cloud principal et l'honeypot, et garantir que même si l'attaquant a découvert qu'il s'agis d'un honeypot et a pu suivre la notification il finira par l'atteinte du cloud d'intervention qui comporte à la base aucune donnée et non pas le cloud principal.

3.4.3.2 Environnement honeypot

Dans cette section, nous expliquerons les configurations apportées aux dispositifs utilisés dans notre approche ConfidCloud+.

• Déploiement du Honeypot :

Le honeypot joue le rôle d'appât en imitant des vulnérabilités connues ou des ports ouverts susceptibles d'être ciblé par les attaquants. Il est installé dans le sous-réseau du cloud secondaire, en raison d'isolation du reste de l'infrastructure, et mis une base de données leurre afin d'attirer l'attaquant à l'exploiter et interagir avec elle, ce qui nous permettra de surveiller quelles sont les motivations de l'attaquant et prédire quel type de donnée seront ciblées de fuite avant que cela ne se produise.

La base de données leurre sera chiffrée et sécurisée pour donner impression que les données

sont réelles, de plus un IDS sera mis en place dans le même sous réseau, configurer avec la base de données leurre pour détecter en temps réel les tentatives d'accès, et renvoyer ensuite une notification au cloud d'intervention ; la notification comprend le nom de la table et le nom de l'attribut ex : utilisateurs, et le nom de l'attribut ex :mots de passe .

- **Déploiement de l'IDS :**

Le honeypot sera équipé d'un système de détection d'intrusion (IDS) pour assurer une surveillance des activités et détecter les comportements suspects ou malveillants, et ce en notifiant le cloud d'intervention de manière altérante, ce scénario aura lieu au moment où un attaquant tente d'accéder à une telle donnée ; la notification se présente sous forme du nom de la table et le nom de l'attribut qui décrit la donnée ciblée. Afin de garantir un bon fonctionnement de déploiement de l'IDS, la configuration suivante doit être suivie :

Pour commencer, il est nécessaire de choisir les journaux à surveiller en fonction de l'objectif de l'opération, qui dans ce cas est la détection des activités malveillantes et des violations de sécurité dont ces journaux peuvent contenir des informations sur les tentatives d'authentification, les modifications de configuration, l'accès aux fichiers sensibles,etc.

Par la suite, établir les règles de détection et identifier les éléments-clés de ces activités. Dans notre cas nous prenons en considération toute tentative d'accès à une donnée comme activité suspecte par conséquent des alertes vont être générée lorsqu'un événement correspondant à la règle est détecté. Une fois une activité suspecte est détectée, une notification comportant uniquement les noms des tables et des attributs auxquels les données visées sont associées sera envoyé directement au cloud d'intervention pour faire face à la situation.

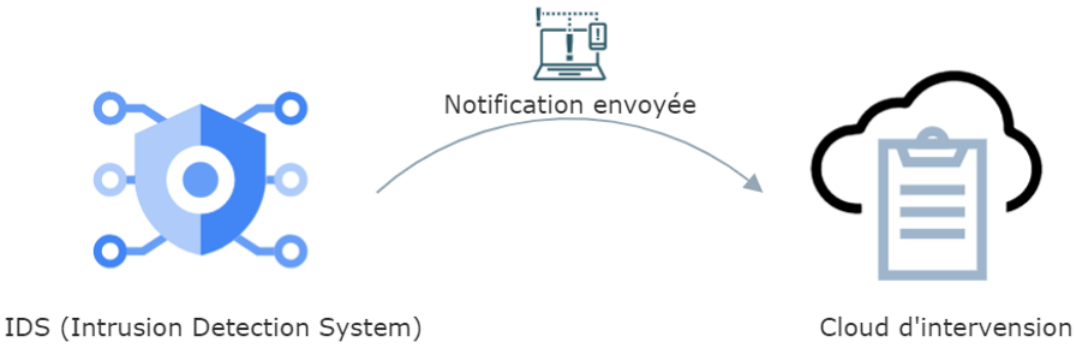

FIGURE 3.4 – Envoie de la notification.

3.5 Conclusion

La majorité des entreprises stockent leurs données sensibles dans le cloud, faisant confiance à sa sécurité. Cependant, cela implique de confier ces données à des prestataires tiers, ce qui rend la confidentialité cruciale pour garantir la fiabilité et la sécurité des données stockées ; notre approche ConfidCloud+ qui se base sur le multi-cloud est à la fois défensive et offensive grâce au honeypot implémenté.

Dans notre proposition, toutes données, y compris les informations basiques, sont considérées comme importantes dont ils sont chiffrés avec pailler, les données high confidentielles bénéficient d'un double chiffrement, tandis que les données top confidentiel sont fragmentées en trois fragments, deux stocker sur le cloud principal et un sur le cloud secondaire. Si une atteinte à la confidentialité est prédite grâce au honeypot, les mesures de sécurité seront renforcées de manière plus intensive, à un niveau supérieur, afin d'assurer une confidentialité maximale de ces données. Dans le chapitre prochain, nous évaluerons la performance de notre approche en utilisant divers outils et techniques tel que la complexité, l'analyse de sécurité,etc.

Chapitre 4

Validation de la solution

4.1 Introduction

Dans ce chapitre, nous nous concentrons sur l'évaluation de cette approche afin de mesurer ses performances et son efficacité. Nous avons utilisé une méthodologie rigoureuse pour évaluer la confidentialité, la sécurité, la performance, et la scalabilité de notre approche. Nous avons utilisé une série d'outils pour évaluer chaque aspect de notre approche, en utilisant une analyse de sécurité avec scenarios pour évaluer la confidentialité et la sécurité de l'approche , un calcul de complexité pour évaluer sa complexité ainsi que une maquette représentante l'approche ConfidCloud+ .

Les résultats de notre évaluation sont présentés en détail dans ce chapitre. Nous discutons des avantages et des inconvénients de notre approche, en mettant en évidence les domaines où elle excelle, et les domaines où elle peut être améliorée. Nous présentons également des recommandations d'utilisation de notre approche et ses limites.

4.2 Analyse de sécurité :

Cette section fournit une analyse de sécurité traduisant des sénarios mener sur ConfidCloud+ ; ainsi que sur les autres approches étudies :

Objectifs : Evaluer ConfidCloud+ qui assure la confidentialité dans le cloud computing.

Menaces potentielles : Interception de données, vol de données sensibles, prendre le contrôle du fournisseur, accès non autorisé au données, Exfiltration de données de stockage dans le cloud, Ressources tierces non sécurisées.

Mesures de sécurité : Architecture multicloud, chiffrement homomorphe des données, classification des données par niveaux de confidentialité, fragmentation des données, dispositifs de sécurité, technologie des leurres, détection et réponse aux activités suspectes, mise à jour de la politique de sécurité.

Scenario 1 : Attaque de chaine d'approvisionnement (supply chain attack).

L'attaque de chaîne d'approvisionnement vise des fournisseurs cloud afin de les compromettre, le danger de cette attaque réside dans le fait qu'elle peut être indétectable et passait inaperçu pendant un bon moment car elle se passe en amont de la cible réelle, elle consiste à exploiter la confiance accordée aux fournisseur cloud , en introduisons des logiciels malveillants, ainsi récolter des informations sensibles ou réussir un vol de données [15].

-Approche ConfidCloud+ : L'architecture multicloud de notre approche ConfidCloud+ limite l'attaque de la chaîne d'approvisionnement puisque le risque que le fournisseur sera affecté sera répartit sur trois Cloud différents (cloud principal, cloud secondaire et cloud d'intervention) au lieu que le risque sera centralisé sur un seul cloud et comme toute attaque passe par la phase de recherche de vulnérabilité sera attiré par les vulnérabilités du honeypot et donc le plus grand risque sera sur le cloud secondaire qui ne contient pas de vrai données sensibles ce qui minimise les conséquences de cette attaque sur les autres Cloud notamment le cloud principal.

- Approche[23] : L'information secrète et les données censées être confidentielles sont partagées sur deux nuages notamment le cloud de stockage qui stocke les bases de données chiffrés ainsi que les clés publiques, le cloud de traitement qui effectue les calculs et les traitements sur les données ainsi le stockage des clés privées, si une attaque de chaîne d'approvisionnement arrive à compromettre l'un des deux clouds cela va affecter les données sensibles qui y sont stockées, ce qui peut poser un vol de données, altération et modifications de données ou prendre le contrôle total sur le cloud.

-Approche [58] et [1] : Elles se sont basées sur une architecture mono cloud qui par conséquent tout le risque d'une attaque de chaîne d'approvisionnement sera sur un seul fournisseur cloud qui contient toute les données de ce dernier donc si l'attaquant arrive à compromettre et avoir la confiance de ce cloud ,le cloud se transformera en un cloud malveillant cela va causer la compromission de tout le cloud qui est une situation critique.

Scenario 2 : Attaque par injection SQL.

Dans ce type d'attaque, l'attaquant identifie d'abord une application web qui utilise une base de données SQL telle que Amazon Web Services, Microsoft Azure et Google Cloud Platform, afin d'exploiter une vulnérabilité d'injection SQL, via des sous-requêtes et des piles en insérant des codes SQL malveillants dans la requête du client qui consiste à injecter plusieurs requêtes SQL dans une seule requête SQL [12]. Cette technique permet à l'attaquant d'exécuter plusieurs actions en une seule injection.

Lors de l'exploitation de cette attaque, l'attaquant peut avoir un accès à des bases de données sensibles, et même à des configurations des outils de défense Informatiques tel que l'IDS qui représente un outil de surveillance de trafic réseau et des activités système pour détecter les tentatives d'intrusion ou les activités suspectes, cela signifie que la mauvaise ou la fausse configuration de

l'IDS peut avoir des conséquences graves sur la sécurité et la confidentialité des données grâce à son rôle important.

-Approche ConfidCloud+ : La configuration de l'IDS est très simple, dont l'alerte qui se déclenche lors d'une tentative de lecture des données leures, qui veut dire que dans le pire des cas l'alerte prévue ne va pas se déclencher pour un certain temps (augmentation de faux négatifs) en attendant la mise à jour et les corrections des administrateurs, ce qui garantit la confidentialité des données stockées.

-Approche [58] : Consiste à utiliser une architecture qui combine entre un honeypot, bases de données KDD99 et un IDS configuré d'une façon spéciale pour réagir correctement aux attaques, afin de garantir la confidentialité des données sensibles. L'inconvénient dans ce cas est la possibilité de ruiner le bon fonctionnent de l'IDS, si un attaquant décide d'utiliser l'injection par SQL cela vas mettre la confidentialité des données dans un grand risque, dont l'utilisation de cette attaque peut permettre aux attaquants d'accéder à la configuration de l'IDS et donc modifier et changer la façon dans il réagisse ; dans ce cas l'IDS ne vas pas réagir de la façon prévue afin d'arrêter l'attaque après la réception du journal de l'attaque capturer, et donc l'attaquant peux accéder aux données sensibles ; en utilisant des techniques comme la table arc-en-ciel pour obtenir les clés de chiffrement hachées, il aura l'accès au données stockées.

- Approche[23] : La solution ne mentionne pas de mécanismes de contrôle d'accès afin de restreindre l'accès aux bases de données pour seulement les personnes autorisées, ce qui permettra à l'attaquant d'accéder à la base de données en tant qu'utilisateur légitime et d'exécuter des commandes SQL malveillantes. À base de ça il peut obtenir des résultats chiffrés en utilisant les mêmes opérations homomorphes que celles utilisées pour chiffrer les données originales en injectant par exemple une requête qui retourne la clé secrète utilisée pour chiffrer les données, ce qui lui permettrait ensuite de déchiffrer les données stockées sur le cloud, autrement dit l'attaquant peut manipuler les résultats chiffrés pour obtenir les informations qu'il souhaite.

Scenario 3 :Éscalade de privilèges.

Pour qu'un attaquant réussisse l'éscalade de privilège faudrait d'abord avoir un accès en mode utilisateur et après tenter d'avoir l'accès priviligié avec l'utilisation d'exploits et de payloads [32].

-Approche ConfidCloud+ : L'attaquant va trouver les vulnérabilités intentionnelles du honeypot et il va tenter de les exploiter pour réussir à avoir un accès de bas niveau ensuite enchainer un escalade de privilèges ; l'attaquant ne sait pas qu'il est sur un système leurre contenant pas de vraies données sensibles car il y a que une base de données leurre , et un fragment stocké, dans le pire des cas si l'attaquant détecte qu'il est tombé dans le piège d'un honeypot ce qui peux arriver si l'attaquant est un black hat il va essayer de l'exploiter pour faire de l'escalade de privilèges et atteindre sa cible. Dans notre cas le honeypot est situé dans le sous réseau du cloud secondaire afin de limiter la propagation d'une attaque en cas de compromission et les activités malveillantes

peuvent pas sortir et atteindre le fragment stocké dans le cloud secondaire et même si l'attaquant réussisse son attaque et put avoir le fragment cela ne lui serait pas de grande utilité car il lui faudra deux autres fragments pour réussir a récolter l'information complète, et cela vas être difficile pour lui puisque les deux autres fragments sont stocké ailleurs dans le cloud principal. Dans le cas ou l'attaquant a eu accès à la base de données leurres une notification contenant le nom de la table et de l'attribut ciblé sera envoyée à l'instant notifiant l'accés obtenu aux données souhaité et une mesure de sécurité sera mis à jour en tenant compte du niveau de confidentialité de l'attribut.

- Approche[23] : Le cloud de traitement est supposé être semi-honnête ce qui veut dire que la capacité du cloud à protéger les données est limitée et que les clients du cloud ont un contrôle total sur la sécurité de leurs données, un attaquant peut essayer d'exploiter le fait que le cloud est semi-honnête et le compromettre pour accéder aux clés privées qui y sont stockées et réussir à faire une escalade de privilèges pour obtenir un accès non autorisé aux données des clients.

-Approche [58] : Le positionnement du honeypot n'est pas en faveur de la sécurité des données, puisque quand la vulnérabilité est exploitée l'attaquant essaiera d'obtenir un accès plus élevé et étendu au système, ce qui lui permettra d'obtenir une bonne position pour lancer d'autres attaques et compromettre la confidentialité des données qui y sont stockées puisque le honeypot n'est pas isolé du reste de l'environnement. Par conséquent dès que le honeypot est franchi et plus précisément franchit avec un accès privilégié les données seront directement exposées à l'intrus.

Scenario 4 : Attaque de l'homme au milieu (MITM) .

En utilisant l'attaque "Man In The middle", l'attaquant intercepte la communication entre deux parties afin d'espionner, modifier et voler les données transmisses sur n'importe quel réseau, et même violer l'identité d'une ou des deux parties impliquées [57], en d'autres termes cette attaque touche beaucoup plus la confidentialité des données sensibles. Pour se faire l'attaquant peut utiliser l'attaque par redirection de port pour intercepter la communication en changeant le chemin qu'elle doit suivre, et pour éviter toute détection, l'attaquant utilise des techniques de camouflage pour masquer sa présence, comme l'utilisation de faux certificats SSL ou de fausses adresses IP.

-Approche ConfidCloud+ : Quand l'attaquant arrive à exploiter une faille et retrouver la base de données leurres mise dans le cloud secondaire et découvre le fonctionnement de l'architecture adaptée, il vas essayer d'intercepter la notification envoyée de l'IDS vers le cloud d'intervention afin de collecter plus d'informations, et même modifier le contenu de la notification dans le but de réussir ses attaques suivantes. Cette opération ne lui servira à rien, car la notification générer et envoyée par l'IDS au cloud d'intervention ne contient aucune information sensible ou compréhensible, la notification va apporter seulement le nom de la table et le nom de la colonne de la donnée (attribut) que l'attaquant a essaié de lire, alors même si l'attaquant arrive à exploiter ces information cela ne lui donnera aucun résultat. Au même temps le retracement de la notification envoyée ne va lui apporter rien de plus, car la notification va être reçue par un cloud d'intervention qui ne contient aucune donnée, ainsi qu'il est placé à distance du cloud principal comme politique

de sécurité contre ce genre de menaces. Même si l'attaquant est arrivé à atteindre le vrai cloud et intercepter la communication entre le cloud principal et le cloud secondaire il ne pourra pas rassembler des données interceptées car seulement un fragment qui sera envoyé au cloud secondaire pour le stockage.

-Approche [58] : Si un attaquant exploite l'attaque MITM il aura la possibilité d'accéder aux données sensibles stockées dans le cloud, dans ce cas l'attaquant va intercepter le journal stocké et enregistré lors d'une tentative d'attaque et qui sera envoyé à l'IDS afin d'appliquer la méthode de défense adaptée pour chaque attaque selon des mesures identifiées d'avant, ce journal contient des informations sur l'attaque capturée tel que le type, les caractéristiques,etc. La modification ou la suppression de ce genre de données peuvent causer des dommages et exposer la sécurité des données stockées à des risques, car cette procédure va rendre l'attaque non détectable et donc aucune mesure de sécurité ne sera appliqué, et alors l'attaquant pourra atteindre des données très facilement, et utiliser ensuite des techniques (table arc-en-ciel, force brute...) afin de récupérer les données d'origine.

-Approche [1] : Elle comporte une faille très sensible dont l'attaquant peut l'exploiter afin d'obtenir les données stockées dans le cloud cela en utilisant l'attaque MITM qui vise à intercepter les données transmise, dans ce cas l'ensemble des données public ou moins confidentielles transmise aux clients suite à leurs demande et qui ne sont ni chiffrer ni protéger lors de cette opération, donc si un attaquant arrive à atteindre ces données il pourra surement trouvé certaines données mal classifiées et classées en tant que public alors qu'elles sont confidentielles et sensibles, en ajoutant que si l'identité du client qui demande ces données est connue par l'attaquant il pourra combiné ces données avec des données collectes à travers l'ingénierie sociale et donc atteindre des informations sensibles .

Ci-dessous le tableau 4.1 met en évidence les points essentielles de comparaison :

	Approches			
Attaques	ConfidCloud+	Shadow-Honeypot	Bi-Cloud	C2aaS
Supply chain	-Risque répartis sur trois clouds. -Le plus grand risque est redirigé (grace au honeypot) vers le cloud contenant pas de vraies données sensibles.	-Risque centralisé en raison de l'architecture mono-cloud.	-Risque repartis en deux clouds,mais les deux clouds contiennent des données sensibles.	-Risque centralisé en raison de l'architecture mono-cloud.
MITM	-Pas de données sensibles dans les transmissions. -Pas de chemin direct entre le cloud principal et l'environnement honeypot. -Le cloud intervention joue le rôle d'intermédiaire.	-Possibilité d'intercepter les journaux logs et divulguer des informations sur des mesures de sécurité.		-La partie des données non confidentielles est stockée et transitée en clair ;la relation entre ces données pourrait être sensible en cas d'interception.
Sql Injection	-Aucun risque d'altérer les données sensibles même en cas de contrôle sur l'ids.	-L'atteinte à la configuration de l'ids, pour qu'il détectera plus les attaques et passer inaperçue.	-Pas de dispositifs de prévention.	
Escalade de privilèges	-Honeypot isolé. -Le résultat d'une escalade de privilège sera une base de donnéesfictive et un fragment de données non significatif à lui tout seul.	-Mal positionnement du honeypot. -L'accès privilégié exposera directement les données sensibles à l'intrus.	-L'un des deux clouds est supposé être semi honnête. -Un escalade de privilèges pourrait faire gagner la confiance du cloud semi-honnête et prendre le contrôle sur lui et avoir des accès non autorisé aux clés privées des clients.	

TABLEAU 4.1 – Confidcloud+ VS shadow honeypot, bi-cloud et C2aaS.

4.3 Étude de complexité

La complexité est une mesure de la difficulté d'un système, un algorithme, une approche,etc. Qui dépend du nombre de ressources tel que le temps, l'espace, puissance de calcul et plusieurs autre facteurs nécessaires pour résoudre un tel problème [42]. Il existe plusieurs formes de complexité, la classification suivante représente la complexité de la plus petite à la plus grande :

Complexité constante : O(1)

Complexité logarithmique : O(log n)

Complexité linéaire : O(n)

Complexité quadratique : $O(n^2)$

Complexité exponentielle : $O(2^n)$

Le calcul de complexité exacte de notre approche ConfidCloud+ qui se compose de plusieurs environnements est extrêmement difficile, dû au manque d'informations tel que la taille de entreprise, les types de services cloud utilisés,etc. Pour cela nous avons essayer de définir la complexité d'une façon approximative ; pour ce faire nous avons définit la complexité de chaque environnement des trois cloud et combiner ensuite les résultats trouvés.

• Complexité du cloud principal :

La méthode la plus important dans ce cloud est la classification ce qui signifie que le résultat dépends de la complexité de la classification, dans notre cas si nous supposons que le nombre de caractéristiques d'entrée est (n) et le nombre de classes de sortie est trois, la complexité temporelle de l'algorithme de classification par régression logistique multinomiale serait de l'ordre de O(3nT), où T est le nombre d'itérations nécessaires pour converger vers une solution, sachant que la régression logistique multinomiale est considérée comme relativement rapide et efficace pour la classification de grands ensembles de données, ce qui signifie que la complexité de cette méthode de classification relativement faible quoi que ce soit (n). Donc la complexité du cloud principal peut s'écrire sous forme O(n).

• Complexité du cloud secondaire :

La complexité du cloud secondaire dépend des éléments suivants :

La complexité d'un honeypot standard pourrait être estimée à O (log n) en fonction de la nature spécifique du honeypot et de ses fonctionnalités.

La complexité de la base de données leurres dépend de certaines caractéristiques telles que la taille, la méthode de génération des données,etc. Dans notre cas nous supposons que la base de données est de taille moyenne, correctement indexée et que les données sont stockées de manière

cohérente dont l'accès est aléatoire, la complexité d'accès aux données peut être relativement faible et elle sera généralement proportionnelle à la taille de la base de données donc (O(n)).

La complexité d'un IDS se base sur trois principaux facteurs, la taille de la base de règles, la fréquence de mise à jour et la méthode de détection, ainsi que la fréquence des accès aux données leurres. Dans notre cas la taille de la base de règles est d'envergure limitée et qu'aucune règle ne s'ajoute lors de la mise à jour, la fréquence des accès, est supposée moyenne ce qui signifie que la complexité de cette IDS est (O(n)) dont n représente le nombre d'accès aux données par unité de temps.

En combinant ces résultats, la complexité du cloud secondaire peut se variée à O (log n).

- **Complexité du cloud d'intervention :**

La complexité des cloud dépend de plusieurs facteurs, notamment de l'architecture du cloud, de la configuration matérielle et logicielle, de la bande passante du réseau, du nombre d'utilisateurs simultanés,etc. Pour le cloud d'intervention qui ne contient aucune donnée, juste la notification envoyée de la part de l'IDS qui comporte seulement deux attributs qui vas être ensuite transmisse, donc sa complexité est très faible dont elle se change en fonction de nombre de notification reçu (n).

- **Complexité de la ConfidCloud+ :**

Au final en combinant les résultats obtenus nous pouvons conclure que la complexité de notre approche ConfidCloud+ peut se résume par O (log n), qui est dans les normes.

4.4 Maquette de la ConfidCloud+

Dans cette section nous présenterons une maquette de ConfidCloud+ visant nous permettre d'avoir une représentation visuelle des idées, de fonctionalités et de flux de travail et ce afin d'obtenir des premières retours de l'expérience utilisateur.

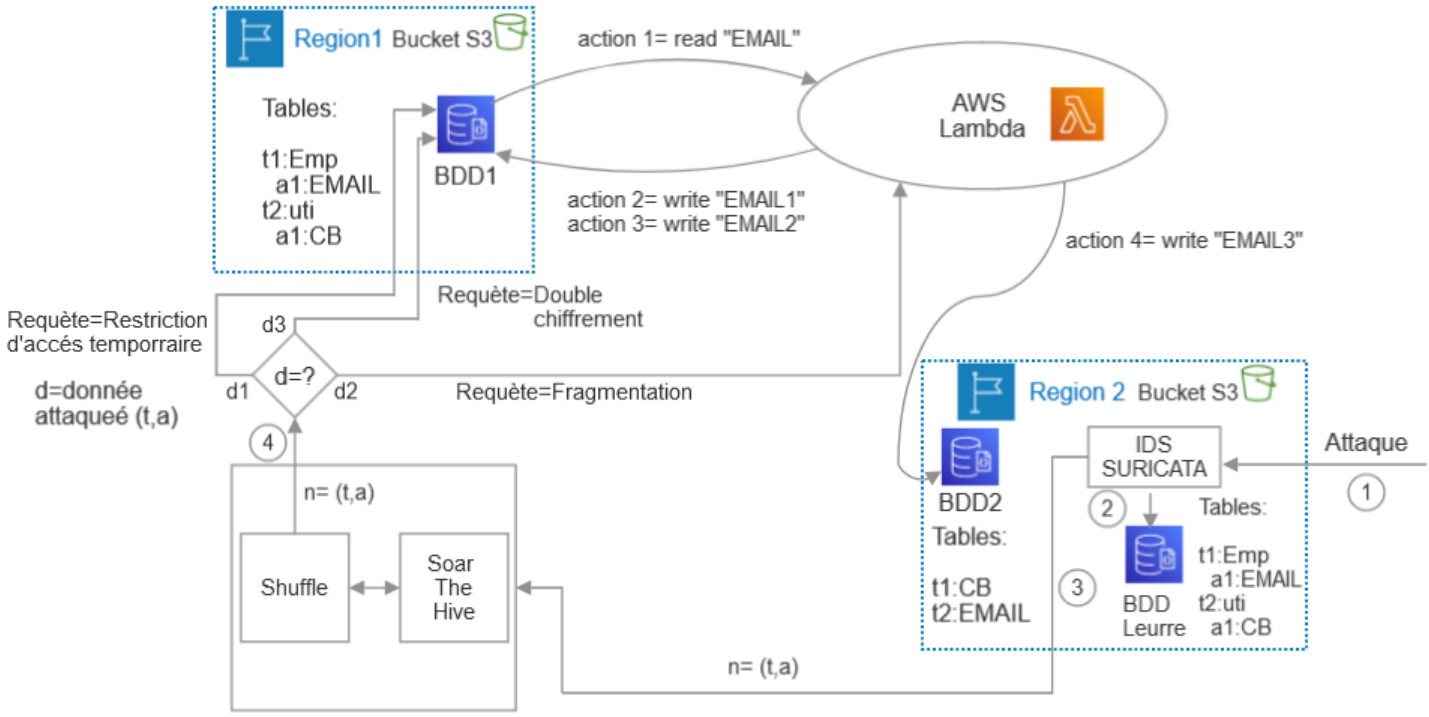

FIGURE 4.1 – Maquette pour la ConfidCloud+ sur AWS.

L'attaque arrive vers la base de données leurre et essaie de voler une donnée,par exemple la donnée dont : le nom de la table est EMP et le nom de l'attribut est EMAIL, l'ids suricata envoie instantanément une notification n=(t,a). Le couplet (t,a) représente Le nom de la table et le nom de l'attribut réspectivement, dédiés à la donnée ciblée. Ensuite la notification sera remontée et le niveau de confidentialité de la donnée est comparé d= ?

Cas N°1 : d = d1 :la donnée est de niveau top confidential ; une requête= restriction d'accès temporaire est envoyé.

Cas N°2 : d = d2 : la donnée est de niveau high confidential ;une requête=fragmentation est envoyé vers AWS lambda pour effectuer la fragmentation, dans notre exemple "EMAIL" est de niveau high confidential donc il sera fragementé en trois ; EMAIL1 et EMAIL2 seront envoyé vers BDD1 et EMAIL3 vers BDD2.

Cas N°3 : d = d3 : la donnée est de niveau medium confidential ; un double chiffrement sera appliqué comme mesure de sécurité.

Les tableaux 4.2 et 4.3 représentent respectivement les paramétres employés et les notions fondamentales de la maquette ConfidCloud+.

Paramètres	Signification
a	attribut
t	table
n	notification
d	donnée attaquée
d1	donnée top confidential
d2	donnée medium confidential
d3	donnée high confidential
Email1	fragment 1
Email2	fragment 2
Email3	fragment 3

TABLEAU 4.2 – Paramètres lies à la maquette

Notions	Déscription
Region	Un environnement, ou centre des données, dont les données sont stockées et sauvegardées au niveau de cloud [3].
Bucket S3	Un conteneur logique qui peut contenir un nombre illimité d'objets tels que des bases de données,des documents,etc.Hautement évolutif sécurisé et durable fourni par AWS [4].
Ids SURICATA	Un type d'IDS open source de haute performance, avec une capacité de détection d'intrusion en temps réel, de prévention des intrusions en ligne, de surveillance de la sécurité du réseau (NSM) et de traitement des pcap hors ligne,développé par l'OISF (Open Information Security Foundation)[5].
AWS lambda	Un service de calcul sans serveur fournit par AWS, il permet d'exécuter des codes et des applications ainsi que des services sans avoir à gérer l'infrastructure, d'une façon simple et facile [26].
Shuffle	Un shuffle est une opération qui consiste à déplacer des données depuis une machine (physique ou virtuelle) vers une autre machine, et ce qu'elle soit son emplacement dans le même réseau local ou un réseau diffèrent [65].
Soar	SOAR (Security Orchestration, Automation and Response) fait référence aux technologies qui permettent aux organisations de collecter des données sur les menaces de sécurité surveillées par les équipes d'exploitation à partir de plusieurs sources, telles que les alertes provenant d'autres technologies de sécurité et aussi l'analyse des incidents et le triage fort combinant la puissance humaine et celle des machines aidant ainsi à réduire le délai de réponse, les tickets des alertes simples non liées à des menaces réelles peuvent être fermées automatiquement [24].
The hive	Une plateforme évolutive et collaborative pour répondre aux incidents de sécurité qui nécessitent une enquête et une action rapides.Et se charge de la gestion des alertes si des alertes doivent être envoyées pour la réponse aux incidents ou non [45].

TABLEAU 4.3 – Tableau des notions utilisées dans la maquette

4.5 Conclusion

En résumé, après la génération de quatre scénarios d'attaque afin de déterminer les points forts de notre approche, ainsi réaliser une étude de complexité sur les techniques les plus importantes que nous avons utilisées lors de la génération de ConfidCloud+, et proposer la maquette pour l'approche ConfidCloud+, nous pouvons conclure que ConfidCloud+ se révèle très efficace selon diverses métriques et garantit la confidentialité des données dans le cloud en résistant à plusieurs attaques et techniques de vol d'informations néanmoins notre approche ne peut pas être déployer par les particuliers en raison du cout élevé de celle-ci, seuls dans les cas où la confidentialité est un point critique ; généralement dans l'armée et les grandes entreprises .

Conclusion et perspectives

Le cloud computing, est une technologie qui consiste à utiliser l'Internet pour fournir des services Informatiques dont le stockage et le traitement de données, l'accès aux ressources,etc. Au lieu de les gérer localement sur des serveurs physiques ; ainsi qu'il offre une multitude d'avantages, notamment la flexibilité, l'accessibilité mondiale,etc. Cependant, ces données stockées nécessitent un niveau élevé de confidentialité, pour faire face aux menaces croissantes, ce qui représente le grand challenge de cette technologie.

Ce travail met en évidence les concepts fondamentaux de cloud computing, tel que les modèles, les services et les tendances du cloud, ainsi que la sécurité dans le cloud dont le principe de la confidentialité et de la sécurité en générale. Plus les revus littérateur opté lors de la génération de l'approche ConfidCloud+, dans le but d'assurer un niveau élevé de sécurité des données stockées dans le cloud.

Pour atteindre cette objectif, nous avons généré une approche nommée ConfidCloud+, cette derniére est basée sur l'utilisation de plusieurs techniques pour garantir la confidentialité des données. L'approche ConfidCloud+ repose sur un environnement multi-cloud, qui se compose de trois cloud, un cloud principal dont les données sont classifiées en trois niveaux, ensuite sécurisé selon leurs niveau de confidentialité, un cloud secondaire qui apporte un honeypot afin de piéger et perturber les attaquants, et un cloud d'intervention, chargé de recevoir des notifications en cas d'attaque sur le honeypot, les transmettre après au cloud principal pour renforçer la sécurité des données ciblées . Ces mesures de sécurité permettent de diminuer le taux des accès non autorisés aux données ainsi le vol de données sensibles, et même d'assurer un niveau élevé de confidentialité des données sensibles et publiques.

En perspective, nous espérons une amélioration de la mise en oeuvre de ce travail, en menant des similations sous AWS, et augmenter le niveau de confidentialité des données sensibles stockée dans le cloud, ainsi réussir à déminuer au maximum les fuites de données au niveau de cloud.

Ce travail nous a apporté des bénéfices sur le plan théorique beaucoup plus en enrichissant nos connaissances existantes, au même temps en acquérant de nouvelles connaissances.

Bibliographie

[1] Munwar Ali, Low Tang Jung, Ali Hassan Sodhro, Asif Ali Laghari, Samir Birahim Belhaouari, and Zeeshan Gillani. A confidentiality-based data classification-as-a-service (c2aas) for cloud security. *Alexandria Engineering Journal*, 64 :749–760, 2023.

[2] amazon. Qu'est-ce que la 5g ? `https://aws.amazon.com/fr/what-is/5g/`, Consulté le 31 avril 2023.

[3] Aws Amazon. Régions et zones. `https://docs.aws.amazon.com/fr_fr/AWSEC2/latest/UserGuide/using-regions-availability-zones.html`, Consulté le 12 juin 2023.

[4] AWS Amazon. Amazon s3. `https://aws.amazon.com/fr/s3//`, Consulté le 31 Mai 2023.

[5] AWS Amazon. Ids suricata. `https://aws.amazon.com/marketplace/search/results?prevFilters=%7B%22ref%22%3A%22portal_asin_url%22%7D&searchTerms=IDS+suricata`, Consulté le 31 Mai 2023.

[6] Azure. Qu'est-ce que l'edge computing ? `https://azure.microsoft.com/fr-fr/resources/cloud-computing-dictionary/what-is-edge-computing/`, Consulté le 27 Mars 2023.

[7] Azure. Sécurité locale et physique des centres de données azure. `https://learn.microsoft.com/fr-fr/azure/security/fundamentals/physical-security`, Consulté le 28 Mars 2023.

[8] Hillary Baron. Understanding cloud data security and priorities in 2022. `https://cloudsecurityalliance.org/artifacts/understanding-cloud-data-security-and-priorities/`, Consulté le 01 Mars 2023.

[9] Hillary Baron. Technology and cloud security maturity. `https://cloudsecurityalliance.org/artifacts/cloud-security-and-technology-maturity-survey/`, Consulté le 07 Mai 2023.

[10] Solimene Bechoua. *La proposition d'une approche pour la découverte et la sélection des services Cloud à base d'ontologie*. PhD thesis in Computer Science, University of LAARBI BEN MHIDI OUM EL BOUAGHI, 2018-2019.

[11] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[12] J. Clarke. *SQL Injection Attacks and Defense*. Elsevier Science, 2012.

[13] CLOUDFLARE. Qu'est-ce que le multi-cloud ? | définition du multi-cloud. https://www.cloudflare.com/fr-fr/learning/cloud/what-is-multicloud/, Consulté le 09 Avril 2023.

[14] Chirag Bhalodia et al. difference between active and passive attack. https://www.chiragbhalodia.com/2021/09/difference-between-active-and-passive-attack, Consulté le 05 Avril 2023.

[15] Dansimp et al. Supply chain attacks. https://learn.microsoft.com/en-us/microsoft-365/security/intelligence/supply-chain-malware?view=o365-worldwide, Consulté le 14 Mai 2023.

[16] Jon-Michael C Brook et al. Top threats to cloud computing pandemic eleven. https://cloudsecurityalliance.org/research/working-groups/top-threats/, Consulté le 07 Avril 2023.

[17] Joseph Bradley et al. The security guidance for critical areas of focus in cloud computing v4.0 (guidance v4.0). https://www.cisco.com/c/dam/en_us/about/ac79/docs/re/Impact-of-Cloud-IT-Consumption-Models_Study-Report_fr.pdf, Consulté le 26 Mars 2023.

[18] Josh Buker et al. Understanding cloud data security and priorities in 2022. https://cloudsecurityalliance.org/artifacts/understanding-cloud-data-security-and-priorities/, Consulté le 07 mars 2023.

[19] Rich Mogull et al. The security guidance for critical areas of focus in cloud computing v4.0 (guidance v4.0). https://cloudsecurityalliance.org/download/security-guidance-v4/, Consulté le 24 Mars 2023.

[20] Wesley Chai et al. What is cloud computing ? https://www.techtarget.com/searchcloudcomputing/definition/cloud-computing, Consulté le 06 Avril 2023.

[21] forcepoint. Sécurité locale et physique des centres de données azure. https://www.forcepoint.com/fr/cyber-edu/cloud-security, Consulté le 07 Avril 2023.

[22] LA REDACTION DE FUTURA. Contexte des cyber-menaces avira gmbh. https://www.futura-sciences.com/tech/questions-reponses/cybersecurite-sont-nouvelles-cyber-menaces-proteger-15252/, Consulté le 01 Avril 2023.

[23] Tasneem Ali Ghunaim, Ibrahim Kamel, and Zaher AL Aghbari. Framework for protecting the confidentiality of outsourced data on cloud. In *2020 14th International Conference on Innovations in Information Technology (IIT)*, pages 29–34, 2020.

[24] Gartner Glossary. Security orchestration, automation and response (soar). https://www.gartner.com/en/information-technology/glossary/security-orchestration-automation-response-soar/, Consulté le 12 juin 2023.

[25] R.A. Grimes, P. Van Goethem, and A.S. Vilret. *Hacking/contre hacking*. Hors collection Sciences. De Boeck supérieur, 2019.

[26] Encora Inc. Aws lambda and serverless architecture workshop. https://aws.amazon.com/marketplace/pp/prodview-67l7o7wbcbvmm?sr=0-1&ref_=beagle&applicationId=AWSMPContessa, Consulté le 31 Mai 2023.

[27] Isaca, Information Systems Audit, and Control Association. *IT Control Objectives for Cloud Computing : Controls and Assurance in the Cloud*. Information Systems Audit and Control Association, 2011.

[28] james carroll. The workbench for machine learning. https://snapcraft.io/weka, Consulté le 25 Avril 2023.

[29] Adem K. cloud-computing et cybersecurite toutes les informations a connaitre. https://www.cyberuniversity.com/post/cloud-computing-et-cybersecurite-toutes-les-informations-a-connaitre, Consulté le 03 Avril 2023.

[30] Mohamed KABA. disponibilite-integrite-et-confidentialite. https://ciberobs.com/2021/03/12/disponibilite-integrite-et-confidentialite/, Consulté le 31 avril 2023.

[31] H. Kim. *Information Security Applications : 22nd International Conference, WISA 2021, Jeju Island, South Korea, August 11–13, 2021, Revised Selected Papers*. Lecture Notes in Computer Science. Springer International Publishing, 2021.

[32] Amos kingatua. Attaques d'escalade de privilèges, techniques et outils de prévention. https://geekflare.com/fr/privilege-escalation-attacks/, Consulté le 14 Mai 2023.

[33] N Jagadish Kumar, K V Vinisha, C Balasubramanian, Dasam Sowmya, and K S Prathibapriya. Privacy preserving data sharing in cloud using eae technique. In *2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, pages 384–388, 2022.

[34] Université Paris-Est Marne la Vallée. Introduction aux ids. http://www-igm.univ-mlv.fr/~dr/XPOSE2004/IDS/IDSPres.html, Consulté le 31 avril 2023.

[35] R. Lee. *Computer and Information Science 2021Summer*. Studies in Computational Intelligence. Springer International Publishing, 2021.

[36] losif peterfi et al. Ethernity cloud decentralized confidential computing on ethereum compatible blockchain. https://ethernity.cloud/, Consulté le 27 Mars 2023.

[37] T. Lynn, J.G. Mooney, L. van der Werff, and G. Fox. *Data Privacy and Trust in Cloud Computing : Building trust in the cloud through assurance and accountability*. Palgrave Studies in Digital Business & Enabling Technologies. Springer International Publishing, 2020.

[38] Université montreal. Politiques de sécurité. https://www.iro.umontreal.ca/~salvail/securite/notes2014/securite11-2014court.pdf, Consulté le 31 avril 2023.

[39] K. Munir. *Cloud Computing Technologies for Green Enterprises*. Advances in Business Information Systems and Analytics (2327-3275). IGI Global, 2017.

[40] HAMDANI nadir et al. *Etude et comparaison des failles de sécurité d'OpenStack et OpenNebula*. PhD thesis in Computer Science, University of Mouloud Mammeri, 2019.

[41] NIST. what+is+cloud+computing. https://www.nist.gov/searchs=what+is+cloud+computing&index=all-meta-engine, Consulté le 03 Avril 2023.

[42] S. Perifel. *Complexité algorithmique*. Références sciences. Ellipses, 2014.

[43] port swigger. sql injection. https://portswigger.net/web-security/sql-injection, Consulté le 05 Avril 2023.

[44] SureshBabu R Prabahar L, Sukumar R. CCSC-DHKEP : Data confidentiality using improved security approaches in cloud environment. *Wireless Personal & Communications*, 122 :3633–3647, 2022.

[45] THEHIVE PROJECT. Thehive a 4-in-1 security incident response platform. https://thehive-project.org/, Consulté le 12 juin 2023.

[46] W. Puech. *Sécurité multimédia 2 : Biome ?trie, protection et chiffrement multime ?dia*. Number vol. 2 in Encyclopédie sciences, Image, Compression, codage et protection des images et vidéos. ISTE editions, 2021.

[47] Sumedh N. Pundkar and Narendra Shekokar. Cloud computing security in multi-clouds using shamir's secret sharing scheme. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pages 392–395, 2016.

[48] David Puzas. Menaces et defis lies a la securite cloud. https://www.crowdstrike.fr/cybersecurity-101/cloud-security/cloud-security-risks-threats-challenges/, Consulté le 07 Avril 2023.

[49] Urvashi Rahul Saxena and Taj Alam. Modified identify and broadcast-based encryption scheme to secure cloud. In *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, pages 289–294, 2022.

[50] Salman Ravoof. Qu'est-ce qu'un exploit zero-day ? et pourquoi soint-ils dangereux ? https://us.norton.com/blog/emerging-threats/how-do-zero-day-vulnerabilities-work#, Consulté le 10 Avril 2023.

[51] Mickael RIGONNAUX. sensibilisation-phishing-hameconnage. https://www.sitec.corsica/sensibilisation-phishing-hameconnage/, Consulté le 04 Avril 2023.

[52] T. Roux. *L'art de la guerre cyber : Vers une intelligence des crises*. Nunkee Éditions, 2020.

[53] N.B. Ruparelia. *Cloud Computing*. The MIT Press Essential Knowledge series. MIT Press, 2016.

[54] Samy SADI. *Techniques de Checkpointing pour la Tolérance aux Fautes dans le Cloud Computing*. PhD thesis in Computer Science, University of Mouloud Mammeri, 2017.

[55] Muhamed Sahrudin. pengertian dan teknik man in middle. https://www.inputekno.com/2017/01/pengertian-dan-teknik-man-in-middle.html, Consulté le 06 Avril 2023.

[56] sailpoint. authentication-methods-used-for-network-security. https://www.sailpoint.com/fr/identity-library/authentication-methods-used-for-network-security/, Consulté le 31 avril 2023.

[57] Betsy Samuel and Vivek Somasundaran. Prevention of man-in-the-middle attacks using blockchain vpn. 12 2022.

[58] T. Saravanan and S. Saravanakumar. Privacy preserving using enhanced shadow honeypot technique for data retrieval in cloud computing. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pages 1151–1154, 2021.

[59] Mélissa SY SAVANE. Les technologies émergentes à surveiller en 2023. `https://fr.linkedin.com/pulse/les-technologies-%C3%A9mergentes-%C3%A0-surveiller-en-2023-m%C3%A9lissa-sy-savane`, Consulté le 31 avril 2023.

[60] SK Singh. *Cloud Computing : Cloud Computing Fundamentals | IaaS | PaaS | SaaS | FaaS | Serverless Computing | Virtualization | Virtual Machine | Hypervisor | Docker*. KnoDAX, 2022.

[61] Aina Strauss. definition-cloud-computing-selon-nist. `https://www.hebergeurcloud.com/definition-cloud-computing-selon-nist/`, Consulté le 09 Avril 2023.

[62] C. Surianarayanan and P.R. Chelliah. *Essentials of Cloud Computing : A Holistic Perspective*. Texts in Computer Science. Springer International Publishing, 2019.

[63] Lucid Content Team. Understanding the basics of cloud computing. `https://www.lucidchart.com/blog/cloud-computing-basics`, Consulté le 24 Mars 2023.

[64] Karim Timraz, Tawfiq Barhoom, and Tamer Fatayer. A confidentiality scheme for storing encrypted data through cloud. In *2019 IEEE 7th Palestinian International Conference on Electrical and Computer Engineering (PICECE)*, pages 1–5, 2019.

[65] Abdelwahab Touil. Comprendre comment spark traite les shuffles. `https://meritis.fr/spark-shuffle/`, Consulté le 12 juin 2023.

[66] W. Tounsi. *Cybervigilance et confiance numérique : La cybersécurité à l ?ère du Cloud et des objets connectés*. Collection réseaux et télécommunications. Iste editions, 2019.

[67] unknown. how does a ddos attack work. `https://sectigo.com/resource-library/how-does-a-ddos-attack-work`, Consulté le 05 Avril 2023.

[68] waikato. Weka 3 : Machine learning software in java. `https://www.cs.waikato.ac.nz/ml/weka/`, Consulté le 12 Mai 2023.

[69] A.S. Zaidoun. *Sécurité informatique : Concepts et outils*. G - Reference,Information and Interdisciplinary Subjects Series. ISTE Editions Limited, 2023.

[70] A.S. Zaidoun. *Sécurité informatique : Concepts et outils*. G - Reference,Information and Interdisciplinary Subjects Series. ISTE Editions Limited, 2023.

[71] Xiaoyu Zhang, Chao Chen, Yi Xie, Xiaofeng Chen, Jun Zhang, and Yang Xiang. A survey on privacy inference attacks and defenses in cloud-based deep neural network. *Computer Standards Interfaces*, 83 :103672, 2023.

Chapitre A

Annexe

A.1 Introduction

Ce chapitre, concentre sur des aspects complémentaires des techniques utilisées lors de la contribution de ConfidCloud+. Il fournit des informations détaillées sur la méthode de classification par régression logistique multinomiale, ainsi que le chiffrement de Shamir Secret Sharing, pour approfondir la compréhension de l'approche ConfidCloud+.

A.2 La régression logistique multinomiale

La classification par régression logistique multinomiale utilise une fonction d'activation exponentielle douce nommée "Softmax" pour produire une distribution de probabilité sur plusieurs classes, elle prend en entrée un vecteur de scores (également appelé logits) et renvoie un vecteur de probabilités normalisées de nombres réels compris entre 0 et 1, où chaque élément représente la probabilité d'appartenance à une classe spécifique, la fonction Softmax est définie comme suit : Sof tmax(zi) = e (zi)/(sumje (zj)) [11] Où e est la fonction exponentielle, z est un vecteur de nombres réels représentant la somme pondérée des variables explicatives pour chaque classe, et sumje (zj) représente la somme exponentielle de toutes les valeurs dans le vecteur z. Prédire ensuite la probabilité qu'un attribut (colonne) appartienne à une des trois classes générés et donc l'appartenance de chaque donnée, cela en analysant la relation entre une variable de réponse sous la forme d'un indice de décès et plusieurs variables indépendantes.

A.3 Le partage secret de Shamir :

Le partage secret de Shamir représente un moyen de fragmentation de données, Le secret est réparti entre un groupe de n participants, et tout le monde profite du secret. L'avantage de cette approche est que le secret ne peut être reconstruit que si k actions sont combinées ; les partages individuels sont inutiles en eux-mêmes, de sorte que les personnes ayant moins de k partages sur n n'ont aucune information supplémentaire sur le secret par rapport aux personnes ayant 0 partages.

C'est le seuil(n,k) n :nombre de fragments , k :nombre de fragments nécessaires pour reconstruire le secret ,les étapes en principe c'est comme suit :

* le secret est S.
*]Il est divisé en N parties : S1, S2, S3, ..., Sn.
* Après l'avoir divisé, choisir un nombre de seuil (k) qui déterminera le nombre minimum de sous-parties nécessaires pour reconstituer la partie originale. Ensuite, générez k-1 nombres aléatoires et utilisez-les pour calculer les autres sous-parties, k est choisi par l'utilisateur afin de décrypter les parties et de trouver le secret d'origine.
* Il est choisi de telle sorte que si nous connaissons moins de K parties, alors nous ne pourrons pas trouver le secret S (c'est-à-dire que le secret S ne peut pas être reconstruit avec (K-1) parties ou moins.
* Si nous connaissons K ou plusieurs parties de S1, S2, S3, ..., Sn, alors nous pouvons calculer/reconstruire notre code secret S facilement. C'est ce qu'on appelle conventionnellement (K, N) le schéma de seuil.

A.4 Conclusion

En conclusion, ce chapitre apporte une valeur ajoutée à la discussion principale en fournissant des informations supplémentaires et des perspectives approfondies sur l'approche ConfidCloud+, pour mieux comprendre les étapes suivies, et les algorithmes appliquer afin d'atteindre le résultat final.

RÉSUMÉ

Dans le domaine de l'information, la préservation de la confidentialité des données stockées ou partagées est essentielle, notamment en ce qui concerne le cloud computing, la protection des données sensibles contre les accès non autorisés est un défi majeur à relever. Malgré l'adoption de différentes approches l'avancement dans cet aspect, se fait de manière progressive. Afin de résoudre cette problématique et dans le but d'assurer un niveau élevé de confidentialité des données stockées dans le cloud, nous avons proposé une contribution nommée ConfidCloud+ basée sur l'utilisation d'une architecture multi-cloud, composant de trois cloud dont le cloud principal, cloud secondaire et le cloud d'intervention ; une classification des données par régression logistique multinomiale en trois niveaux comme suit : top confidentiel, high confidentiel et medium confidentiel, selon la sensibilité de la donnée ; ainsi un environnement honeypot placé dans le sous réseau du cloud. A travers l'approche ConfidCloud+ qui combine entre plusieurs technologies, il pourrait être possible d'accroître le niveau de confidentialité des données à un niveau supérieur, dans un avenir proche.

Mots clés : Confidentialité des données ; multicloud ; classification ; fragmentation ; chiffrement ; honeypot.

ABSTRACT

In the information sector, preserving the confidentiality of stored or shared data is essential, particularly where cloud computing is concerned. Protecting sensitive data from unauthorized access is a major challenge. Despite the adoption of various approaches, progress in this area is gradual.

To address this issue, and with the aim of ensuring a high level of confidentiality for data stored in the cloud, we have proposed a contribution called ConfidCloud+ based on a multi-cloud architecture, consisting of three cloud including the primary cloud, the secondary cloud and the intervention cloud ; a data classification by multinomial logistic regression into three levels as follows : Top confidential, high confidential and medium confidential, depending on sensitivity of the data ; a honeypot environment placed in the cloud's subnetwork. Through the ConfidCloud+ approach, which combines several technologies, it may be possible to raise the level of data confidentiality to a higher level in the near future.

Key words : Data confidentiality ; multicloud ;classification ; fragmentation ; encryption ; honeypot.