Mémoire de recherche pour l'obtention du master

MIAGE parcours Système d'Information et Innovation (S2I)

Université Paris 1 Panthéon Sorbonne

Les méthodes de sécurité pour la protection des données stockées dans le cloud

Réalisé par Richard Nakarmi

Tuteur enseignant : Corinne Plourde

Maître d'apprentissage : Farouk Berrouba

Année : 2022/2023

Table des matières

Remerciement

Tout d'abord, je souhaite exprimer ma profonde gratitude envers Mme PLOURDE et l'ensemble de l'équipe pédagogique du master MIAGE de l'Université Panthéon-Sorbonne Paris 1 pour leur précieux soutien tout au long de l'élaboration de mon mémoire de recherche.

Je tiens également à adresser mes sincères remerciements à mon tuteur d'alternance, M. BERROUBA, pour sa disponibilité, ses conseils et sa précieuse expérience qui ont grandement contribué à la réussite de mon immersion dans le monde professionnel.

Enfin, je remercie mes camarades de promotions avec qui j'ai pu échanger et qui ont apporté une contribution précieuse à la conception de ce mémoire de recherche.

Glossaire

ABAC / RBAC : Attribute Based Access Control / Role Based Access Control

ABE / RBE : Attribute Based Encryption / Role Based Encryption

AES : Advanced Encryption Standard

ANSSI : Agence Nationale de la Sécurité des Systèmes d'Information

CNIL : Commission Nationale de l'Informatique et des Libertés

ECC : Elliptic Curve Cryptography

EHR : Electronic Health Record

IBE : Id-Based Encryption

k-NN : K-Nearest Neighbors

MRFC : Modified Random Fibonacci Cryptographic

NIST : National Institute of Standards and Technology

NTRU : N-th degree Truncated polynomial Ring Unit

RGPD : Règlement Général sur la Protection des Données

SE : Searchable Encryption

SecNumCloud : Sécurité Numérique du Cloud

Sk-NN : Semantic-kNN

SSO : Single Sign-On

1. Introduction

1.1. Contexte

L'informatique en nuage, plus connue sous son nom anglicisme Cloud Computing est une des technologies les plus émergentes de cette dernière décennie. En effet, ce nouveau paradigme informatique qui permet d'accéder à des services hébergés sur Internet occupe une place centrale au sein des entreprises, notamment avec l'émergence de fournisseurs cloud comme Google Cloud Platform (GCP) ou encore Amazon Web Service (AWS). Ce dernier présente 4 avantages à l'utilisation du cloud par les entreprises [40] comprenant :

- l'agilité : le cloud offre un accès simple à une vaste gamme de technologies favorisant l'innovation rapide [40]
- l'élasticité : les ressources sont ajustables en fonction des besoins réels [40]
- la réduction des coûts : le cloud repose sur un modèle de tarification basé sur le pay-as-you-go [6], c'est-à-dire que les utilisateurs ne paient que ce qu'ils utilisent.
- le déploiement rapide : le cloud permet un déploiement des activités à grandes échelles très rapidement [40]

Ainsi avec ces nombreux avantages, de plus en plus d'entreprises se tournent vers le cloud pour se fournir différents services notamment de stockage afin de gérer efficacement leurs grandes quantités de données.

1.2. Définition de la problématique

Malgré les avantages considérables offerts par le cloud computing, l'adoption de cette technologie soulève également des questions et des défis importants. L'une des problématiques majeures réside dans la sécurité des données. En effet, en confiant leurs informations sensibles à des fournisseurs de services cloud, les entreprises doivent s'assurer que les mécanismes de sécurité en place sont suffisamment robustes pour prévenir toute forme d'attaques et toute atteinte à la confidentialité des données. De plus, elles peuvent renforcer cette sécurité en mettant en place leur propre méthode de sécurisation des données.

C'est ainsi que ce mémoire de recherche se focalise sur la problématique suivante :

Quelles sont les méthodes que les organisations peuvent mettre en œuvre pour sécuriser leurs données stockées dans un environnement cloud ?

1.3. Objectif et plan du mémoire

L'objectif de ce mémoire de recherche consiste à présenter des méthodes de sécurité des données dans le cloud proposées dans la littérature. Les résultats de cette étude seront utiles aux entreprises qui pensent à utiliser le cloud pour leurs activités, ainsi qu'aux experts en sécurité qui cherchent à rester informés des méthodes de sécurité les plus efficaces pour protéger les données stockées dans le cloud.

La suite de ce mémoire est organisé comme suit :

La section 2 présente les définitions et concepts clés nécessaires à la compréhension de cette recherche. La section 3 présente des travaux antérieurs qui traitent de la sécurité des données dans le cloud. La section 4 résume la méthodologie suivie pour réaliser cette revue systématique de la littérature. La section 5 est la partie centrale de notre recherche où nous réalisons l'extraction et l'analyse des différentes méthodes de sécurité. Dans la section 6, nous tentons de répondre à nos différentes sous-questions de recherche avant de conclure.

2. Background

2.1. Définition du Cloud Computing

Le National Institute of Standards and Technology (NIST) est une agence américaine renommée pour son travail dans l'établissement de standards et de bonnes pratiques dans divers domaines, tels que la cybersécurité ou encore les technologies de l'information. En 2011, alors que le concept de “Cloud Computing” commence à émerger, le NIST a formulé une définition et un modèle de référence du Cloud, qui ont été largement adoptés par les entreprises pour décrire les caractéristiques, les modèles de service et les modèles de déploiement du Cloud. Ainsi, le Cloud Computing est défini comme “ un modèle permettant un accès réseau omniprésent, pratique et à la demande à un pool partagé de ressources informatiques”. Ces ressources informatiques sont de plus configurables avec un minimum d'effort de gestion ou d'interaction avec le fournisseur de services [32].

Les principales caractéristiques du cloud sont entre autres l’utilisation des services à la demande sans nécessiter d'intervention humaine, avec une accessibilité complète sur différentes plateformes (PC, tablette/mobile…) à condition d’avoir accès à un Internet. De plus, ces ressources cloud s’adaptent à tous les besoins de l’utilisateur, on dit que ces ressources sont “élastiques”, c’est-à-dire qui s'adaptent de manière dynamique en fonction des besoins attendus.

2.1.1. Types de services cloud

De nos jours, il existe de nombreux types de services cloud qui répondent aux besoins spécifiques des organisations en offrant des solutions variées :

Software as a Service (SaaS) : fournit un service à la demande pour les utilisateurs qui leur permet d'accéder aux données n'importe où sans avoir besoin d'installer préalablement un logiciel. Des exemples de SaaS célèbres sont Outlook, Dropbox, Saleforce…

Platform as a Service (PaaS) : fournit une plateforme de développement et de déploiement d'applications sur le cloud, sans avoir à gérer l'infrastructure sous-jacente de l’application

Infrastructure as a Service (IaaS) : fournit une infrastructure informatique complète (serveurs, réseaux et systèmes de stockage).

Database as a Service (DBaaS) : fournit aux utilisateurs l'accès à une base de données gérée et hébergée sur une plateforme cloud, éliminant ainsi la nécessité de gérer l'infrastructure de cette base. [7]

Integration as a service (iPaaS) : permet aux entreprises de faciliter l'intégration entre différentes applications, systèmes et services hétérogènes

Cette liste de services cloud mentionnés ci-dessus n'est évidemment pas exhaustive.

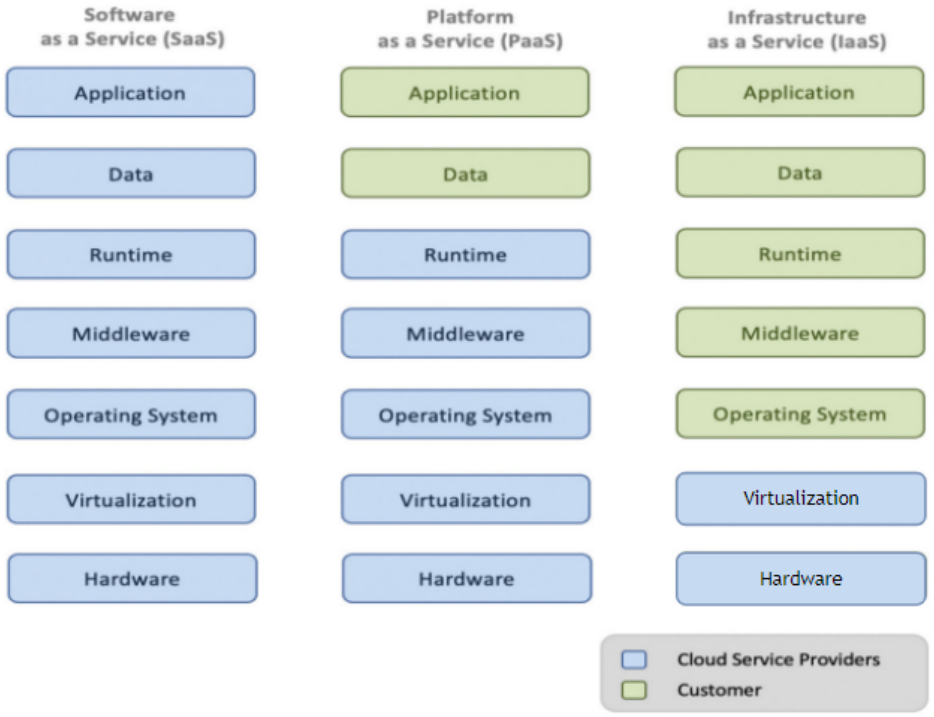

Figure 1 : Les 3 principaux types de services cloud et les responsabilités entre fournisseurs et clients [28]

2.1.2. Types de déploiement cloud

Lorsqu'une entreprise souhaite utiliser un service cloud, une des premières étapes importantes consiste à choisir la façon dont elle va configurer et mettre en œuvre ces services pour répondre au mieux aux besoins de l'entreprise. Pour cela, elle a le choix entre 4 types de déploiement cloud [32]:

Un **Cloud Public** qui permet aux organisations d'utiliser des services ou ressources informatiques par le biais d'un fournisseur cloud. Ces ressources sont partagées entre

plusieurs clients et sont accessibles à partir de n'importe quel endroit disposant d'une connexion Internet [19]. Ce type de cloud offre des avantages tels que le déploiement rapide, des économies de coûts [20] ou encore une grande évolutivité car les ressources informatiques peuvent être facilement augmentées ou réduites en fonction des besoins de l'organisation client [19].

A l'inverse, un **Cloud Privé** fournit les ressources informatiques, qui sont gérées et déployées au sein de l'organisation [19]. Il offre un plus grand niveau de contrôle, et sont utilisés exclusivement par l'organisation cliente. Il offre également un plus grand niveau de sécurité, car l'organisation cliente a un contrôle total sur la sécurité et la gestion des données.

Un **Cloud Hybride** combine les avantages du cloud public et du cloud privé. Les données et les applications sont réparties entre les deux types de cloud en fonction de leur niveau de sensibilité et de leur importance pour l'organisation [19].

Enfin un **Cloud Communautaire** est un environnement de cloud partagé entre des organisations liées par des intérêts et des besoins similaires en matière de sécurité ou de réglementation et de politique. Il peut être possédé, géré et exploité par une ou plusieurs des organisations de la communauté, par un tiers, ou par une combinaison de ceux-ci, et il peut exister sur site ou hors site. [32]

Ainsi, les solutions cloud restent le premier choix en entreprise pour leur projet où ils ne veulent pas gérer la maintenance des systèmes et d'une équipe de développement [28]. Néanmoins, l'une de leurs principales préoccupations, susceptible de freiner leur adoption du cloud, concerne la sécurité et la confidentialité des données stockées [28]. C'est donc dans cet environnement de plus en plus orienté vers le cloud que la gestion des données en entreprise relève d'une importance cruciale.

2.2. La réglementation dans le cloud

Le Cloud Computing soulève des préoccupations significatives en matière de sécurité des données. Pour relever ces défis, diverses réglementations et normes ont été mises en place pour encadrer l'utilisation du Cloud Computing et assurer la protection des données sensibles. Nous retrouvons la Sécurité Numérique du Cloud (SecNumCloud) [41], initiée par l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI) en France, qui illustre un exemple marquant de ces efforts réglementaires. Il y a également le Règlement Général sur la

Protection des Données (RGPD) [42] qui impose des exigences strictes en matière de protection de données personnelles. Ainsi, les entreprises traitant des données personnelles dans le Cloud doivent s'assurer que leurs fournisseurs de services respectent les exigences du RGPD et garantissent la confidentialité, l'intégrité et la disponibilité des données. Dans le cas contraire, elles peuvent mettre en place leurs propres méthodes de sécurité.

La Commission Nationale de l'Informatique et des Libertés (CNIL), qui est également une autorité en France chargée de protéger les droits des individus concernant la collecte, le traitement et la protection des données personnelles, recommande aux entreprises françaises de réaliser une analyse de risque et d'être très rigoureux dans le choix de son prestataire avant de recourir à un service cloud [33]. En effet, le document [33] présente des recommandations en matière de sécurité à destination des entreprises qui souhaitent souscrire à des services cloud. Dans un premier temps, elle suggère d'identifier les données qui seront hébergés via une solution cloud avec notamment :

- les données à caractères personnel
- les données sensibles (qui sont des informations personnelles qui révèlent des aspects particulièrement intimes ou privés de la vie d'une personne [33]
- les données stratégiques pour l'entreprise
- les données utilisées dans les applications métiers

De plus, la CNIL propose plusieurs recommandations concernant la définition des exigences de sécurité techniques et juridiques, la conduite d'une analyse de risques afin de mettre en place des mesures de sécurité adéquates pour les prévenir ou atténuer [33]. En ce qui concerne la responsabilité liée au traitement des données, la CNIL recommande une définition claire des responsabilités de chaque partie impliquée et fournit un partage des responsabilités comme suit :

Hypothèse	Formalités déclaratives	Information des personnes	Obligation de confidentialité et sécurité	Exercice des droits des personnes concernées auprès du …
Le prestataire est conjointement responsable du traitement	**Client**[5]	**Client**[6]	**Client + Prestataire**	**Client (avec le concours du prestataire)**[7]

Figure 2 : Partage des responsabilités proposée par la CNIL entre le client et le prestataire Cloud [33]

Les droits dans la dernière colonne de la Figure 3 font référence aux droits des individus dont les données personnelles sont traitées, incluant le droit d'accès ou encore au droit à l'oubli de leurs données personnelles.

3. Travaux antérieurs

Dans cette section, nous passons en revue différents travaux de recherche que nous avons repérés lors de la sélection de nos articles.

L'article [19] a pour objectif de fournir une vue d'ensemble des risques de sécurité liés au stockage cloud et de proposer des pratiques et des mesures pour les éviter. Il examine les différents types de risques de sécurité, notamment les risques liés à la confidentialité, à l'intégrité et à la disponibilité des données stockées dans le cloud. Les recommandations de l'article en matière de sécurité comprennent le choix d'un fournisseur cloud de stockage sûr et sécurisé, la mise en place de pratiques de sécurité au sein de l'organisation, la surveillance des appareils connectés et l'ajout de mesures de sécurité avancées pour les données sensibles.

L'article [20] fournit une revue détaillée des problèmes de sécurité liés au cloud computing. Il discute des problèmes de confidentialité, de disponibilité et de sécurité des données stockées dans le cloud. L'étude réalise également une revue des menaces dû à l'exploitation de vulnérabilités des systèmes ou de différents types d'attaques.

L'objectif de l'étude [22] est de mettre en évidence les vulnérabilités et les problèmes de sécurité liés à la confidentialité et à la protection des données stockées dans le cloud, ainsi que les risques associés à l'utilisation de réseaux sans fil publics pour accéder à des informations personnelles sensibles. [22] considère que les services cloud sont nécessaires aux entreprises et aux organisations pour le stockage de leurs données volumineuses mais qu'ils doivent s'assurer de la sécurité mis en place par les fournisseurs cloud.

L'article [25] fournit des informations sur les risques et les défis liés à l'utilisation de la technologie de stockage dans le cloud. Il présente différentes vulnérabilités de sécurité concernant les plateformes cloud d'éducation, de données mobiles ou encore de santé. L'étude présente un classement des vulnérabilités en fonction du pourcentage de problèmes engendrés dans un système de stockage. Ainsi, selon [25], la perte ou fuite de données représente le plus gros risque, suivi des problèmes de contrôle d'accès.

Les articles [21] et [28] sont ceux qui se rapprochent le plus de notre sujet. En effet, l'étude [21] présente les problèmes de sécurité des données stockées dans le cloud, en mettant

l'accent sur les problèmes d'intégrité, de confidentialité des données et d'authentification. Les auteurs fournissent ensuite une analyse comparative entre 6 méthodes de sécurisation en analysant notamment leurs avantages, limites et performances. Enfin, l'étude termine en énumérant les défis de sécurité importants pour protéger les données, qui comprennent les fuites de données, les accès aux données, les performances de chiffrement qui ne doivent pas affecter les performances cloud et la sécurité des données pendant les transmissions.

L'article [28] fournit une revue systématique de la littérature qui a pour objectif de présenter des solutions de sécurité cloud pour répondre à différentes problématiques de sécurité, et cela à quatres niveau de l'infrastructure de cloud computing (données, application, réseau et hôte). Il discute des diverses solutions proposées dans la littérature pour aborder les problèmes de sécurité à ces différents niveaux d'infrastructure. Toutefois, [28] souligne des lacunes sur certains sujets de sécurité concernant l'audit des tiers de service, des techniques pour assurer la disponibilité de données ou encore de contrôle d'accès et de gestion d'identités.

4. Protocole de recherche

4.1. Revue systématique de la littérature

Il est important de créer une méthode de recherche pour trouver des recherches scientifiques appropriées qui traitent des questions de recherche identifiées précédemment. Cette méthode doit garantir que les études trouvées soient pertinentes. Dans ce contexte, le processus appliqué pour cette revue systématique de la littérature (SLR) suit les lignes directrices proposées par Kitchenham et Charters [43], qui ont défini trois phases principales pour la réalisation de l'étude :

La première phase est la **planification** qui consiste à décrire les objectifs de l'étude, à définir les questions de recherche et à élaborer le protocole de recherche.

La deuxième phase est la **conduite de la recherche** qui comprend l'identification des articles pertinents, sélection des études primaires en évaluant leur qualité, ainsi que l'extraction et la synthèse des données pour pouvoir répondre à la problématique de recherche. Cela implique notamment la lecture et l'analyse approfondie des articles sélectionnés

La troisième phase de **synthèse** rapporte les résultats de manière systématique. Elle permet de répondre aux questions de recherche de l'étude et de formuler des conclusions ou des recommandations.

Cette section détaille la première phase du SLR.

4.2. Définition des questions de recherches

Une revue de la littérature de qualité doit se focaliser sur les concepts liés à la problématique de recherche. Nous avons ainsi décomposé notre problématique en plusieurs sous-questions selon les différents concepts que celle-ci soulève. Cela permet de mieux structurer la recherche et de répondre à chaque sous-question de manière plus spécifique afin d'élaborer une réponse complète à la question de recherche initiale.

Nous avons formulé nos différentes sous-questions de cette manière :

Q1 : Quelles sont les vulnérabilités de sécurité qui peuvent affecter les données stockées dans le cloud ?

Cette première sous-question (Q1) vise à examiner les principales vulnérabilités de sécurité spécifiquement liées aux données stockées dans le cloud ce qui permet d'évaluer les menaces potentielles et de mieux cibler les mesures de sécurité nécessaires.

Q2 : Quelles sont les méthodes existantes pour sécuriser les données stockées dans le cloud ?

La deuxième sous-question (Q2) se concentre sur l'exploration des méthodes existantes dans la littérature pour sécuriser les données stockées dans le cloud. Il est essentiel de connaître les différentes approches et solutions de sécurité disponibles afin de déterminer quelles sont les méthodes envisageables pour une utilisation en entreprise.

Q3 : Comment les organisations peuvent-elles choisir la méthode de sécurisation des données ?

Enfin, la troisième sous-question (Q3) aborde la question de la sélection de la meilleure méthode de sécurisation des données dans le cloud. Pour cela, nous allons tenter d'évaluer les méthodes à travers différents critères spécifiques puis les comparer afin de les classer.

4.3. La conduite de la recherche

Afin de sélectionner des études scientifiques pertinentes, nous avons suivi la méthode Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [36] pour cadrer cette revue systématique. Cette démarche est composée de 3 étapes principales que nous décrivons ci-dessous : l'identification, la sélection et l'inclusion des études primaires.

4.3.1. Phase d'identification

Cette première étape consiste en la recherche d'études scientifiques à partir de mots clés. Pour cette étude, nous avons utilisé le moteur de recherche Miage Scholar qui permet de collecter des articles parmi plusieurs moteurs de recherche scientifique. Les mots clés ont été choisis à partir de nos sous-questions de recherche, ce qui nous ont donné la chaîne de recherche suivante :

A partir de Q1:

TITLE-ABS-KEY(cloud)
AND (TITLE-ABS-KEY(data storage) OR TITLE-ABS-KEY(data stored))
AND (TITLE-ABS-KEY(security) OR TITLE-ABS-KEY(data protection))
AND (TITLE-ABS-KEY(data issues) OR TITLE-ABS-KEY(data risks))

A partir de Q2 et Q3 :

AND TITLE-ABS-KEY(organization)
AND (TITLE-ABS-KEY(technologies) OR TITLE-ABS-KEY(methods))
AND PUBYEAR > 2018

Cette chaîne de recherche a engendré un nombre important d'articles donc nous avons choisi de garder les articles dont la date de publication est ultérieure à 2018 pour garantir d'obtenir les études les plus récentes sur ce sujet de sécurité cloud.

4.3.2. Phase de sélection puis inclusion

Dans cette phase de sélection, nous soumettons chaque article à nos critères d'inclusion et d'exclusion pour sélectionner les études primaires de notre revue de la littérature.

Critères d'inclusion	Critères d'exclusion
- Les articles qui mentionnent des vulnérabilités de sécurité cloud liées aux données. - Les articles qui présentent une ou plusieurs méthodes de sécurité. - Les articles qui évaluent les méthodes proposées en fonction de critères de performances	- Les articles qui mentionnent la technologie "blockchain" - Les articles non accessibles ou payants - Les articles qui ne sont pas rédigés ni en français ni en anglais

Tableau 1 : Critères d'inclusion et d'exclusion

Concernant les articles qui présentent une méthode, il n'est pas nécessaire de fournir une évaluation explicite de celle-ci, mais les résultats des tests de performance doivent être inclus. Nous avons également volontairement omis les articles qui traitent de la technologie blockchain pour se concentrer sur des solutions qui peuvent être mises en place rapidement.

Ainsi, une lecture plus approfondie a été nécessaire. En effet, la lecture du titre, des mots-clés et de l'abstract nous permettent d'une part de cerner le sujet central de celui-ci. Mais cela n'est pas suffisant pour évaluer la qualité d'un article pour notre recherche.

5. Extraction et analyse des données

5.1 Analyse des vulnérabilités et menaces dans le cloud

Lors de la recherche des articles, nous avons trouvé de nombreuses revues de la littérature qui ont analysé les problèmes liés à la sécurité des données dans le cloud computing. Les données stockées dans des environnements de cloud computing sont exposées à diverses vulnérabilités et menaces potentielles, nécessitant une attention particulière lorsque les entreprises souhaitent utiliser un service cloud. Dans cette section, notre objectif est d'identifier ces vulnérabilités en nous appuyant sur nos articles de recherche.

L'étude [28] considère 4 niveaux de l'infrastructure cloud où les vulnérabilités peuvent être présentes et dont il est nécessaire de planifier des mesures de sécurité. Nous retrouvons le niveau des données, de l'application, du réseau et de l'hôte. Dans ce mémoire, c'est le niveau des données (*data level*) qui nous intéresse.

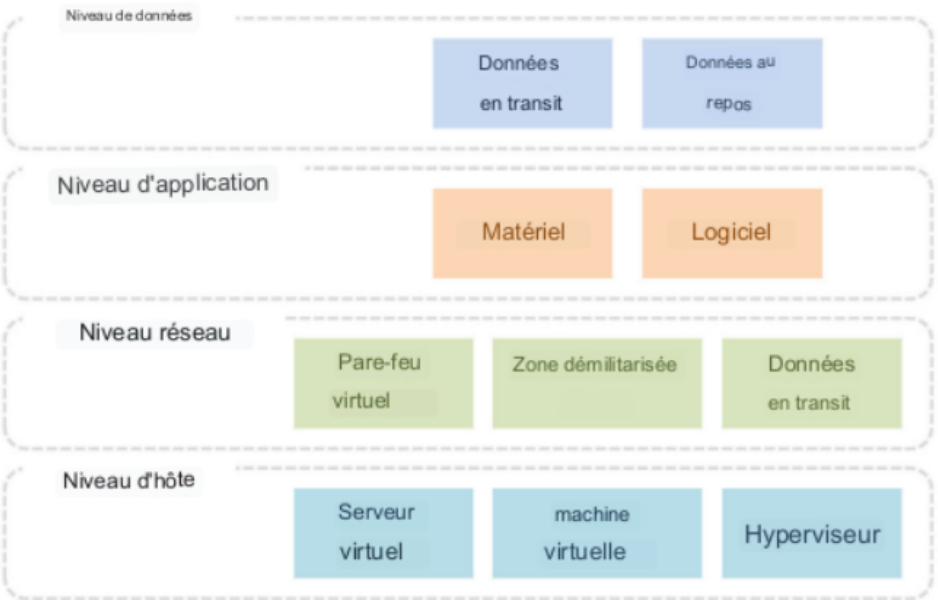

Figure 3 : Les niveaux d'infrastructure cloud [28]

Selon [28], il existe 7 types de vulnérabilité au niveau des données :

La **confidentialité des données** qui concerne la protection des données contre l'accès non autorisé. Les données sensibles doivent être préservées et ne doivent pas être divulguées à des tiers non autorisés. Des problèmes de coordination entre les fournisseurs de services cloud ou des failles de sécurité peuvent compromettre la confidentialité des données. Les données stockées peuvent être compromises en raison de problèmes de coordination entre les fournisseurs de services cloud.

L'intégrité des données qui se rapporte à la garantie que les données n'ont pas été altérées ou manipulées de manière non autorisée. Des erreurs de stockage ou des attaques malveillantes peuvent entraîner des modifications non autorisées des données, mettant en péril leur fiabilité.

La **disponibilité des données** qui concerne la capacité à accéder aux données stockées dans le cloud. Des attaques de type déni de service (DoS), des pannes matérielles ou des erreurs peuvent empêcher l'accès aux données, rendant ainsi les services indisponibles pour les utilisateurs.

La violation de données qui se produit lorsque des attaquants malveillants pénètrent dans le stockage de données, mettant en danger tout l'environnement cloud.

Les pertes ou fuites de données peuvent survenir accidentellement lors du transfert ou du stockage. Cela peut résulter d'erreurs de manipulation, de problèmes de sécurité ou d'attaques malveillantes, mettant en danger la sécurité des données.

La **ségrégation des données** dû aux environnements multi-locataire, où plusieurs clients partagent les mêmes ressources cloud, il existe le risque d'accès non autorisé aux données d'un autre client. Cette vulnérabilité nécessite des mesures de sécurité pour empêcher la fuite ou l'accès non autorisé à ces données partagées.

La **virtualisation des données** implique la création de machines virtuelles et de ressources partagées dans le cloud. Cependant, le déplacement de ces ressources peut entraîner des pertes de métadonnées, provoquant des erreurs et des interruptions de service.

En se basant sur ces vulnérabilités, nous avons examiné la fréquence de mention de chacune de ces vulnérabilités à partir de la littérature. Ainsi, en interprétant la fréquence de mention, nous pouvons identifier les vulnérabilités qui sont largement reconnues comme des points d'attention dans le domaine de la sécurité des données dans le cloud.

Type de vulnérabilité	Nombre de sources	Sources
Pertes ou fuites de données	5	[19], [20], [21], [25], [28]
Confidentialité des données	4	[20], [21], [22], [28]
Disponibilité des données	4	[20], [21], [22], [28]
Intégrité des données	3	[20], [21], [28]
Violation de données	3	[19], [20], [28]
Ségrégation des données	2	[19], [28]
Virtualisation des données	1	[28]

Tableau 2 : Fréquence d'apparition dans la littérature des vulnérabilités au niveau des données

5.2. Analyse des méthodes de sécurité existantes

Dans cette section, nous procédons à l'extraction des données pertinentes à notre recherche sur les méthodes de sécurisation des données stockées dans le cloud. Cette étape est cruciale pour comparer et évaluer les différentes approches existantes. Pour ce faire, nous avons sélectionné des articles en suivant le protocole de recherche décrit dans la section 4. Notre objectif principal est d'identifier les méthodes existantes pour répondre aux différents problèmes de sécurité des données. Ainsi, nous allons analyser ces méthodes à travers différents critères de sécurité des données (confidentialité, disponibilité, intégrité…) et évaluer le niveau d'adoption industrielle de chaque approche, ce qui nous permettra de formuler des recommandations pour les entreprises cherchant à renforcer leur sécurité des données dans le cloud.

Afin de réaliser cette exploration approfondie, nous avons élaboré un tableau pour recueillir et synthétiser les informations clés de chaque méthode de sécurité analysée. Ce tableau nous permettra de comparer les résultats et les conclusions des différents articles, en mettant en évidence les différences et les similitudes entre les méthodes de sécurité

proposées. Dans ce tableau, nous avons défini plusieurs colonnes que nous décrivons ci-dessous, afin de visualiser et analyser rapidement les principales caractéristiques et performances de chaque approche, facilitant ainsi la comparaison entre celles-ci.

Le tableau est composé des colonnes ci-dessous :

- **La problématique** décrit le sujet principal de recherche abordé dans l'article.
- **L'objectif** qui spécifie l'intention de la méthode de sécurité proposée dans l'article.
- **La méthode proposée** que l'article présente.
- **L'évaluation de la méthode** indique comment la méthode proposée a été évaluée ou testée dans l'article.
- **Le résultat** présente les principales conclusions de l'évaluation de la méthode. Elle indique les performances, les avantages ou les inconvénients de la méthode.

Tableau 3 : Référencement des articles sélectionnés

Source	Objectif	Méthode proposée	Evaluation de la méthode	Résultat
[1]	Proposer un modèle de contrôle d'accès basé sur l'ontologie pour les environnements cloud	Utilisation d'une ontologie pour modéliser les entités associées au contrôle d'accès et leurs relations; Exploitation de la propriété de subsumption pour réduire l'ontologie	Comparaison des performances du modèle proposé en termes de nombre de règles stockées, temps de raisonnement et efficacité par rapport aux modèles existants	• Réduction du nombre de règles à stocker • Réduction du temps de raisonnement • Meilleure efficacité par rapport aux modèles existants
[2]	Utiliser des techniques de "searchable encryption" pour le stockage sécurisé des données et proposer une méthode de partage de données	Proposition de la méthode "User Prediction in Role" pour réduire les risques d'attaques internes et partager les données en fonction des priorités des utilisateurs.	Non spécifié pour le chiffrement Les prédictions de rôle sont testées sur un petit dataset	• Réduction du risque d'attaques internes et de fuites de données non autorisées • Amélioration de la sécurité du partage de données dans un environnement basé sur les rôles
[3]	Assurer la confidentialité et l'intégrité des données stockées dans le cloud	Propose un framework de sécurité incluant la génération de clés, la confidentialité, l'authentification et l'échange de clés	Comparaison avec d'autres frameworks de sécurité cloud sur plusieurs critères : confidentialité, authentification,génération et échange de clés	• Amélioration de la sécurité des données stockées dans le cloud • Réduction des coûts de calcul par rapport aux frameworks existants
[4]	Proposer un schéma de gestion de réplication de données sécurisé (SDRMS)	Utilisation d'un Réplica Management Agent (RMA) pour créer des répliques en fonction de la fréquence d'accès aux données et de la charge du serveur cloud. Pour chiffrer les données, utilisation d'un Twin Layered Security Scheme (TSS) permettant de s'assurer de l'intégrité des données	Évaluation des performances du schéma proposé en termes d'accessibilité des données, d'exploitation du stockage, d'allocation des répliques et de temps de récupération	• Amélioration de l'accessibilité des données, de l'utilisation du stockage, de l'équilibrage de charge des serveurs cloud et de la facilité de récupération des données
[5]	Proposer un cadre de sécurité pour les systèmes EHR (Electronic Health Records), en tenant compte de l'intégrité, de la disponibilité et de la confidentialité des dossiers de santé	Solution pour le système EHR en 3 niveaux avec chacune des méthodes de sécurité (SSO,HTTPS/SSL, OAuth, attribute-based encryption (ABE) et attribute based access control (ABCA)	Non spécifié	• La solution proposée combine plusieurs méthodes de sécurités pour réduire les menaces de sécurités concernant les système EHR

[8]	Développer une technique de chiffrement basée sur les attributs et les clés de groupe pour assurer la sécurité des données sensibles dans le stockage cloud	Utilisation de la Cryptographie Fibonacci Aléatoire Modifiée (MRFC) pour diviser les données en groupes d'attributs sensibles et non sensibles, et crypter les sous-groupes sensibles avec des clés de groupe distinctes	Evaluation de la méthode sur le temps d'exécution, le temps de chiffrement/déchiffrement, mémoire utilisée	La méthode proposée assure la sécurité et la confidentialité des données sensibles dans le stockage cloud avec un coût de traitement minimal
[9]	Développer un système de détection et de réduction des cyberattaques basé sur l'apprentissage automatique dans un environnement de cloud computing	Combinaison d'un algorithme de machine learning (SVM) pour la détection précise des attaques avec un système de défense multicouche (SDN)	Test réalisé sur une simulation 3 critères d'évaluation : sensitivity, specificity, and accuracy (precision + recall + F1 score)	Toutes les mesures de performance ont été comprises entre 97,37 % et 98,79 %
[10]	Proposer un algorithme de mobilité et de sécurité des données (DMoS) pour assurer la sécurité de la mobilité des données dans les environnements de cloud	Combinaison d'une méthode de classification des données et application d'un algo de sécurité pour la mobilité (DMos)	Comparaison avec trois algorithmes existants, évaluation des résultats Évaluation sur le calcul du temps de classification des données, de la réponse aux données, du temps de retard des données et du débit de données.	• Réduction du temps de réponse de la transmission • Réduction du temps de traitement total • Meilleure débit de données dans le traitement
[11]	Proposer un modèle de service Confidentiality-based data Classification-as-a-Service (C2aaS) qui traite les données de manière dynamique en fonction de leur niveau de sécurité pour un stockage efficace dans le cloud	Classification des données en deux catégories : confidentielles et non confidentielles. Seules les données confidentielles sont chiffrées (avec AES-128), tandis que les données non confidentielles sont stockées en texte clair.	Évaluation de l'efficacité de la méthode en termes de sécurité des données, de surcharge du système et de coût.	• Réduction de la taille de stockage des données • Réduction du temps de traitement
[12]	Proposer un système intégré de prévention et de détection des intrusions (IIPDS) pour prévenir et détecter différents types d'attaques au niveau de l'infrastructure du système cloud.	IIPDS est composé de : 2 techniques de préventions : Trusted Third Party (TTP) et de protocoles SSL/TLS pour prévenir les intrusions, 2 techniques de détections : Détection d'anomalies selon des paramètres (IP source, temps d'envoie et de réception, nb de paquets) et Snort (logicielle qui analyse le réseau)	Méthode testée avec une simulation et 2 scénarios	• Réduction de la surcharge de traitement en raison du système distribué et de l'utilisation de deux techniques, IPS et IDS, avec deux sous-techniques • Capacité à détecter et bloquer un large éventail d'attaques

[13]	Proposer une méthode de chiffrement basée sur les attributs pour contrôler l'accès aux données stockées dans le cloud computing	La méthode permet au data owner de créer une politique d'accès aux données basée sur une hiérarchie avec différents attributs. Les données sont chiffrées en fonction de ces attributs	Comparaison de la cette méthode qui utilise une structure hiérarchique avec des méthodes standards (sans hiérarchie) sur plusieurs critères : temps de chiffrement et déchiffrement, coût de stockage	• Réduction du temps de chiffrement et de déchiffrement ainsi que l'espace de stockage requis pour les textes chiffrés • Externalisation des opérations de déchiffrement qui améliore l'efficacité du processus de déchiffrement
[14]	Proposer une méthode utilisant la Triple Data Encryption Standard (TDES) pour assurer la sécurité de données médicales dans l'environnement Cloud	Les données sont divisées en trois catégories en fonction de leur importance, et un chiffrement triple, double ou simple est appliqué en fonction de cette importance.	Critères de performances pour la comparaison avec d'autres méthodes existantes : temps d'exécution, % d'utilisation du CPU, % d'utilisation du réseau Comparaison de TDES avec d'autres méthodes de chiffrement existantes telles que Hybrid Encryption, MA-ABE et Tabu sur plusieurs critères : temps de chiffrement, de latence, le débit, taux de livraison des paquets	• Meilleure temps d'exécution et % utilisation de CPU mais plus gourmand en % d'utilisation de réseau • Meilleure temps d'exécution • Meilleure temps de chiffrement/déchiffrement • Moins de latence • Meilleure débit
[17]	Développer une architecture intégrant des services de base de données cloud avec une intégrité des données et la possibilité d'exécuter des opérations simultanées sur des données chiffrées	Utilisation de l'algorithme de chiffrement asymétrique NTRU	Non spécifié	Les principales caractéristiques de NTRU sont qu'il nécessite une faible capacité de mémoire et de calcul tout en fournissant un niveau de sécurité élevé.
[7]	Identifier et surmonter les menaces de sécurité accrues liées au cloud computing, afin de permettre l'utilisation de cette technologie pour les organisations (méthode Cryptographic Data Integrity Checking and Encryption (CryptDICE)	CryptDICE est un système qui se compose de trois couches : la couche Application, la couche Secure Data Access et la couche NoSQL Abstraction API. Cette méthode utilise une combinaison de techniques de chiffrement en fonction de la nature de la donnée	Mesure de l'efficacité et l'efficience de la méthode sur plusieurs critères : coût en ligne de code, capacité à traiter des requêtes sur des données cryptées (SUM, CRUD…)	CryptDICE réduit de manière significative le temps et les efforts nécessaires au développement pour activer le chiffrement des données et prend en charge une variété de requêtes de recherche. De plus, il propose des améliorations de performances visant à réaliser des requêtes d'agrégation à faible latence.

[26]	Proposer un cadre de sécurité des réseaux cloud pour améliorer les performances de sécurité des données dans les environnements cloud Big Data	La méthode proposée dans l'article est une approche cryptographique appelée SB-DS. L'objectif principal de l'approche proposée est de chiffrer les données et de les stocker de manière distribuée sur les serveurs cloud	Comparaison de SB-DS avec AES par rapport aux critères suivant : la taille des données d'entrée, le temps d'exécution du chiffrement et déchiffrement, le temps d'exécution total ou encore le taux de transfert de données,	Les résultats montrent que l'approche proposée est plus efficace que l'algorithme AES en termes de temps d'exécution pour le chiffrement et le déchiffrement des données.
[29]	Proposer un modèle de sécurité pour l'amélioration de la confidentialité des données dans le cloud computing	Propose une encryption multi-niveau. Les données sont d'abord chiffrées avec RSA puis AES	Comparaison de la méthode AES avec son ancien modèle appelé DES en termes de longueur de clé, taille de données en entrée et niveau de sécurité de données.	Le modèle offre une double sécurité de chiffrement (plus grande taille de clé, plus grande taille de données en entrée par rapport à DES)
[30]	Concevoir un système qui permet aux utilisateurs de réaliser un stockage sécurisé des données et un partage sécurisé des données externalisées dans des environnements cloud non fiables	Le système nommé OutFS, chiffre les données avant de les stocker, permet un partage sécurisé entre les utilisateurs et garantit une efficacité élevée pour la gestion des clés et des opérations de partage. Il est conçu pour fonctionner au-dessus d'outils de cloud tiers (Dropbox, Drive..)	Comparaison avec plusieurs autres systèmes existants en mesurant les performances d'écriture et de lecture avec des opérations cryptographiques sur diverses tailles de fichiers externalisés.	Meilleure en termes de confidentialité, d'authentification, d'intégrité des données
[31]	Prévenir les attaques telles que les attaques par force brute, les attaques par injection SQL	La méthode proposée est SUCDDES (Secure Uncrackable Cipher Dynamic Double Encryption Standard), qui utilise un double chiffrement et des clés secrètes partielles pour améliorer la sécurité et le contrôle d'accès aux données dans le cloud.	Comparaison avec d'autres méthodes (UCDDES, RSA) : temps de chiffrement/déchiffrement, téléchargement des données, temps d'uploading/downloading	Meilleure que UCDDES et RSA dans tous les critères

5.3. Classification par types de méthode de sécurité

D'après notre échantillon de 18 articles présentant une méthode de sécurité, nous avons identifié 5 types de méthodes de sécurité :

Type de méthode	Nombre de sources	Sources
Chiffrement	12	[2], [3], [5], [7], [8], [13], [14], [17], [26], [29], [30], [31]
Classification des données	5	[8], [10], [11], [14], [26]
Contrôle d'accès	4	[1], [2], [5], [13]
Prévention / détection de menaces	2	[9], [12]
Réplication de données	1	[4]

Tableau 4 : Types de méthode de sécurité de notre SLR

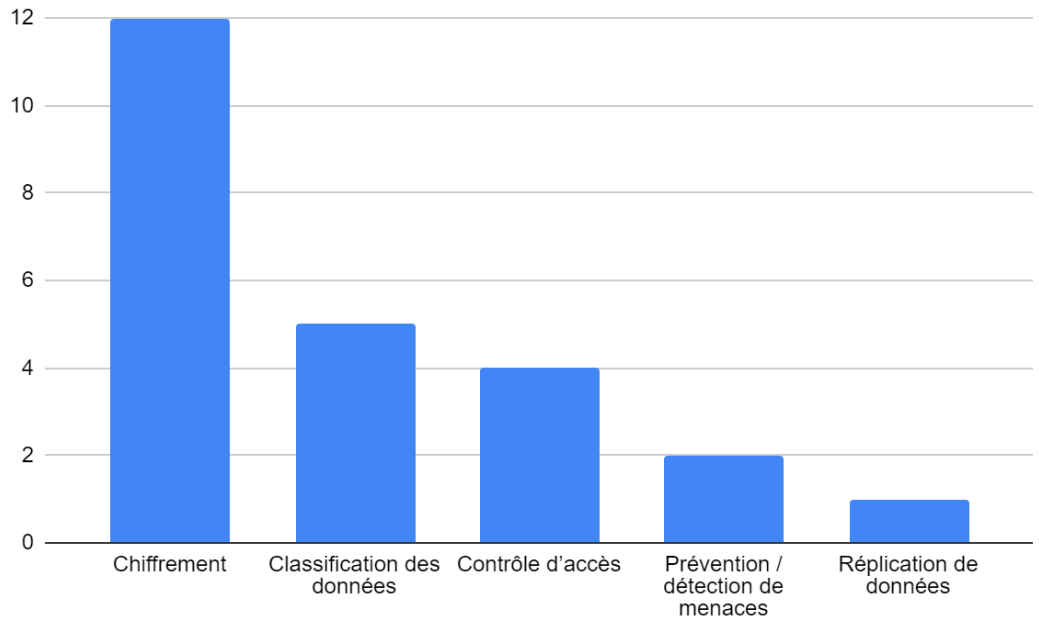

Figure 4 : Nombre d'articles par type de méthode de sécurité

5.4. Analyse des méthodes

Cette section fournira une vue d'ensemble des méthodes que nous avons analysées dans notre échantillon.

5.4.1. Le chiffrement

Le chiffrement est un processus de conversion de texte en clair en une forme codée qui ne peut être lue que par des parties autorisées qui possèdent la clé de déchiffrement. Il existe principalement deux types de chiffrement :

Le **chiffrement symétrique** utilise la même clé pour le chiffrement et le déchiffrement. C'est une méthode rapide et efficace, mais elle nécessite un moyen sécurisé de partager la clé entre l'expéditeur et le destinataire. Ce chiffrement est plus rapide et efficace pour chiffrer de grandes quantités de données [4].

Quant au **chiffrement asymétrique**, cette méthode utilise une paire de clé. une clé publique pour chiffrer tandis que la clé privée sert à déchiffrer. C'est une méthode plus sécurisée que le chiffrement symétrique, mais elle est plus lente et plus intensive en termes de calcul [2].

Mais nous allons voir qu'il y a d'autres types de chiffrement dans les méthodes que nous avons analysées. En effet, les auteurs de [2] suggèrent d'utiliser le chiffrement consultable ou Searchable Encryption (SE) permettant aux utilisateurs de récupérer uniquement les données nécessaires tout en empêchant le fournisseur cloud, qui stocke les données, de les déchiffrer. Pour cela, le propriétaire des données crée un type de "clé spéciale" appelée "trapdoor" pour chaque utilisateur autorisé. Cette clé spéciale permet à l'utilisateur de demander au fournisseur cloud d'accéder à des morceaux spécifiques de données. Néanmoins, lorsque le nombre d'utilisateurs augmente, le propriétaire des données doit générer davantage de ces "clés spéciales", ce qui peut devenir difficile à gérer. Par conséquent, [2] propose d'utiliser deux techniques qui sont le chiffrement basé sur les attributs (ABE) et le chiffrement basé sur les rôles (RBE). Comme son nom l'indique, ABE se base sur les attributs des utilisateurs tandis que RBE repose sur la hiérarchie de rôles au sein de l'organisation. Ainsi, on utilise RBE pour partager les données à un groupe d'utilisateurs plutôt qu'avec des utilisateurs individuels [2].

L'étude [5] propose un framework de sécurité pour les systèmes de dossiers électroniques de santé (EHR). Celui-ci utilise également ABE pour assurer la confidentialité des données des patients.

L'étude [13] précise qu'il existe 2 catégories d'ABE : le Ciphertext Policy-based Attribute-Based Encryption (CPABE) qui consiste à définir une politique d'accès basée sur les attributs dans le message. Tandis que dans le Key Policy-based Attribute-Based Encryption (KPABE), la politique d'accès est liée à la clé de déchiffrement, et permet de déchiffrer. Ainsi, [13] propose un système de sécurité avec CPABE prenant en charge une structure d'accès hiérarchique et un déchiffrement externalisé. En effet, d'une part la structure hiérarchique permet de chiffrer plusieurs messages en utilisant une seule structure d'accès, ce qui simplifie le processus de chiffrement en évitant de créer des structures d'accès distinctes pour chaque message et d'autre part l'externalisation des opérations de déchiffrement permet de résoudre les problème de surcharge associé au déchiffrement.

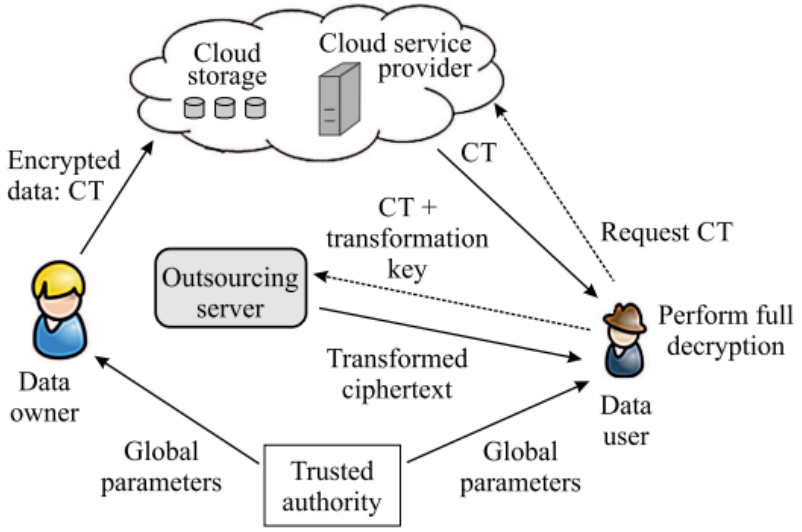

Figure 5 : Système proposé par [13]

(CT signifie "Ciphertext" pour texte chiffré)

Les auteurs de [3] proposent un framework de sécurité de stockage de données dans le cloud. Ce framework assure la confidentialité des données en utilisant une méthode de génération de clé nommée McClaurin series-Laplace combiné à l'algorithme de chiffrement Advanced Encryption Standard (AES) tout en proposant la gestion des clés de déchiffrement ainsi qu'un service d'authentification. AES est un algorithme de chiffrement symétrique qui

prend en entrée, des blocs de données fixés à 128 bits. Le framework proposé est composé de 7 phases :

1. L'enregistrement du propriétaire auprès du système ainsi que les utilisateurs auprès du fournisseur cloud.

2. La génération des clés qui utilise la transformé de Laplace qui sera utilisé pour le chiffrement

3. La phase de chiffrement des données à l'aide de la méthode McClaurin series-Laplace. Le propriétaire peut ensuite envoyer la clé de chiffrement ainsi que les données chiffrées vers un tiers de confiance cloud.

4. Les requêtes des utilisateurs qui souhaitent accéder aux données réalisent la demande au propriétaire qui lui-même demande au tiers de confiance cloud de fournir les données demandées.

5. La phase d'authentification mutuelle où le fournisseur cloud s'assure que les données des différentes parties sont correctes à l'aide d'un protocole d'authentification nommé Knowledge-Based Authentication (KBA)

6. La phase d'échange des clés entre les deux utilisateurs finaux est réalisé avec l'algorithme LU-Factorized

7. La phase de déchiffrement réalisé par l'utilisateur à l'aide de la clé qu'il reçoit de la part du propriétaire

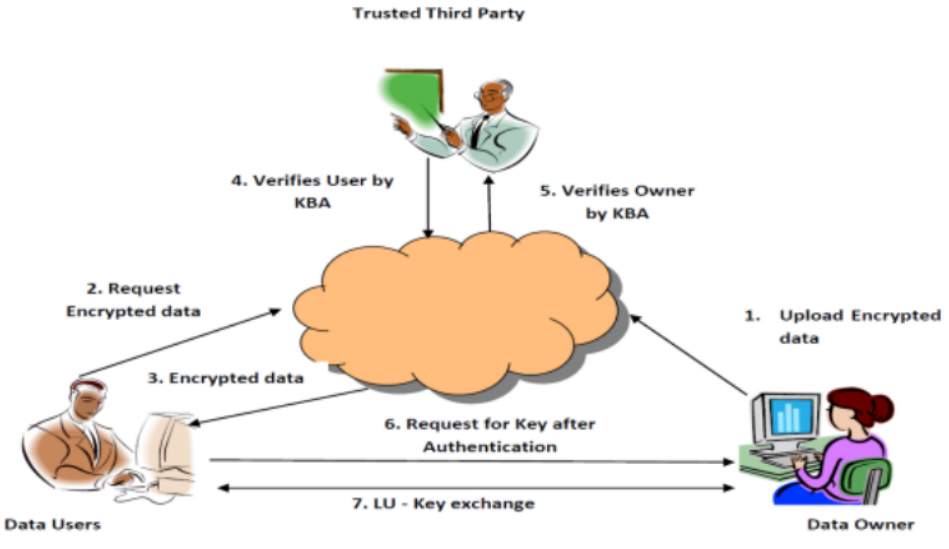

Figure 6 : Schéma du framework proposé dans l'article [3]

[14] propose une méthode de sécurité cloud pour les applications de santé en utilisant un chiffrement symétrique nommé Triple Data Encryption Standard (TDES). En effet, la méthodologie propose tout d'abord de diviser les données en 3 catégories de sensibilité (respectivement faible, moyenne et forte) avant d'appliquer un chiffrement avec TDES (respectivement simple, double et triple).

Les auteurs de [8] proposent un système de stockage sécurisé qui chiffre des données sensibles basé sur une technique de chiffrement nommé Modified Random Fibonacci Cryptography (MRFC). En effet, le propriétaire des données définit des groupes d'attributs sensibles. Chaque groupe sera chiffré à l'aide de l'algorithme MRFC qui combine la cryptographie à courbe elliptique, avec une technique d'échange de clés de Diffie-Hellman. Cela permet de générer et de partager des clés de chiffrement tout en assurant un haut niveau de sécurité dans le processus. Ainsi, chaque groupe possède sa clé de groupe spécifique. L'accès aux données chiffrées est contrôlé par les propriétaires de données et les administrateurs de groupe.

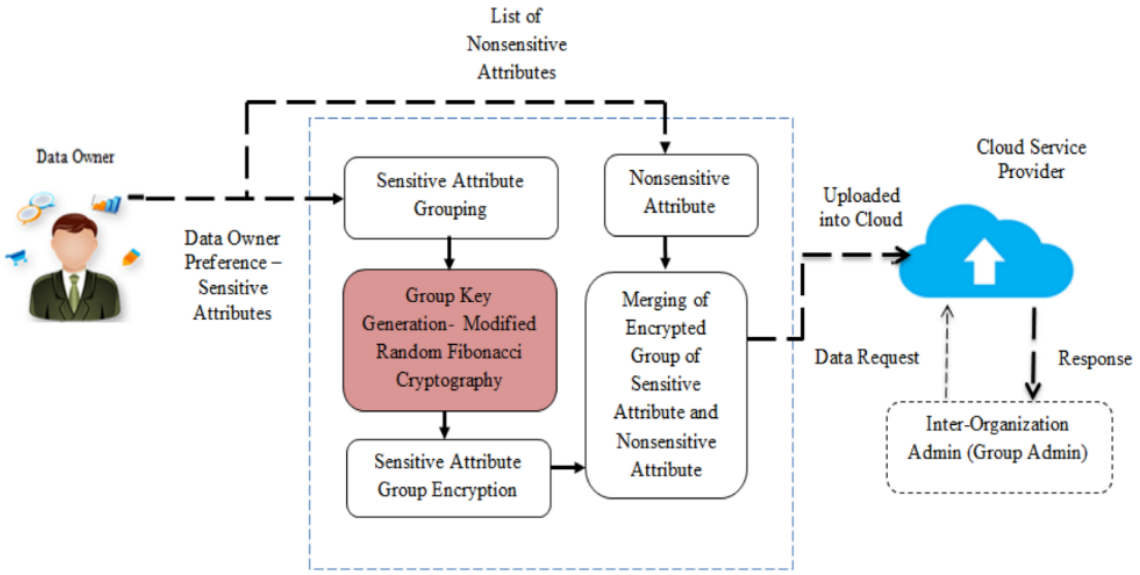

Figure 7 : Système de stockage sécurisé avec MRFC [8]

Les auteurs de [7] proposent un système de protection des données, en particulier pour les applications SaaS qui utilisent souvent des bases de données NoSQL où il existe un manque de sécurité. En effet, les auteurs considèrent qu'il existe une limite dans la confiance lorsqu'on utilise des applications SaaS comme le montre la Figure 7. En effet, les bases de

données NoSQL n'ont pas été conçues avec des aspects de sécurité des données [7]. Cependant, plusieurs éditeurs NoSQL utilisent désormais un chiffrement totale de la base de données à l'aide d'une technologie de chiffrement nommée Transparent Data Encryption (TDE). Cela soulève un problème au niveau des performances notamment lorsque l'on souhaite exécuter des requêtes de recherche et effectuer des calculs sur ces données chiffrées, qui devient très inefficace [7]. En effet, il faut récupérer toutes les données chiffrées, les déchiffrer puis l'exécution de la requête est rendue possible. Pour pallier ce soucis, l'étude [7] compare différents schémas chiffrements de données qui permettent d'exécuter des requêtes de recherches ou des calculs sur des données chiffrées.

Parmi ces schémas de chiffrement que nous classons du plus au moins sécurisé, on retrouve :

Le Random Encryption (RND) qui pour un même texte en clair en entrée aboutit à différents résultats chiffrés.

Le Deterministic Encryption (DET) qui pour un même texte en clair en entrée et la même clé de chiffrement aboutit toujours au même résultat chiffré.

L'Order-Preserving Encryption (OP) préserve les relations d'ordre entre le texte en clair et celui chiffré.

L'Homomorphic Encryption (HOM) permet de réaliser des requêtes qui nécessitent un calcul directement sur les données chiffrées.

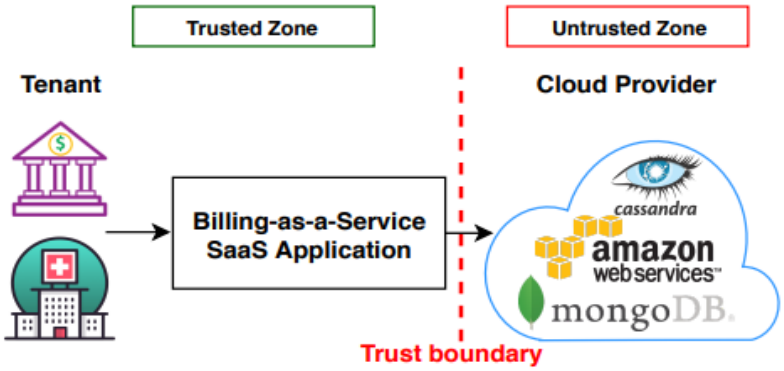

Figure 8 : Limite de confiance lors de l'utilisation d'une application SaaS [7]

L'étude [29] propose d'améliorer la confidentialité des données stockées dans le cloud en réalisant un chiffrement multi-niveau. En effet, tout d'abord le texte en clair est chiffré à un premier niveau avec un algorithme de chiffrement asymétrique nommé RSA qui est très utilisé dans les échanges de données confidentielles. Puis dans un deuxième temps, on soumet le résultat chiffré de ce premier niveau à l'aide de l'algorithme AES.

Les auteurs de [17] proposent de sécuriser les données à l'aide de l'algorithme asymétrique N-th degree Truncated polynomial Ring Unit (NTRU). Le système est composé d'un propriétaire de données qui génère des paires de clés via NTRU pour chaque utilisateur et s'assure que la clé publique est utilisée pour le chiffrement des données avant de les stocker dans le cloud. Les utilisateurs qui souhaitent accéder aux données doivent s'authentifier et avoir la clé de déchiffrement appropriée.

L'étude [30] propose un système de partage de dossier nommé OutFS qui fonctionne au-dessus d'un tiers de services cloud qui permet l'externalisation du stockage de fichiers. Dans ce système, il existe deux dossiers : un dossier personnel où toutes les données sont chiffrées de manière hybride, c'est-à-dire d'abord un chiffrement des données avec AES puis avec RSA pour chiffrer la clé obtenue. D'autre part, il y a un dossier partagé dont les données sont chiffrées à l'aide de la méthode Identity-Based Encryption (IBE). Cette approche se base sur un identifiant unique d'un utilisateur (comme un mail ou numéro de téléphone) pour chiffrer un message. Pour cela,un serveur de confiance nommé Private Key Generator (PKG) va tout d'abord générer une paire de clés principales et partager la clé publique principale à tous les utilisateurs. Quant au receveur, il partage son identifiant (le mail dans le cas de la Figure 9) au PKG pour générer une clé secrète. Enfin, l'envoyeur chiffre le message en utilisant la clé publique principale et l'identifiant du receveur avant que ce dernier le déchiffre avec la clé secrète qu'il a obtenue. L'utilisation d'IBE élimine la nécessité d'une gestion complexe des clés [30]. La Figure 9 résume les principales étapes d'IBE. Ainsi, en utilisant ce mécanisme IBE, OutFS permet le partage de dossier de manière sécurisée sans passer par des mécanismes d'authentification fournis par le service cloud.

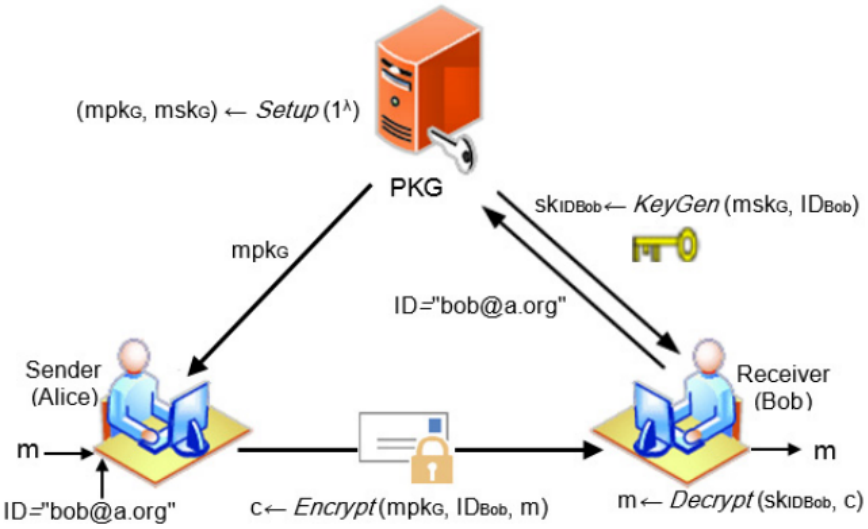

Figure 9 : Fonctionnement du mécanisme d'IBE [30]

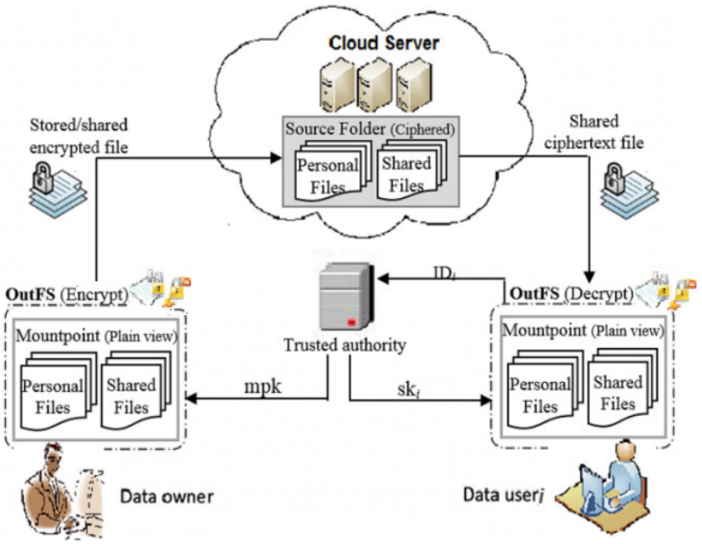

Figure 10 : Workflow du système OutFS [30]

Enfin, les auteurs de [31] proposent un chiffrement nommé Secure Uncrackable Cipher Dynamic Double Encryption Standard (SUCDDES) conçu pour chiffrer les informations de manière hautement sécurisée avec un double chiffrement à l'aide d'AES. SUCDDES prend un texte en clair en entrée puis applique plusieurs opérations de chiffrement pour produire un premier texte chiffré. Ce dernier est pris en entrée et soumis à des opérations similaires pour produire un deuxième texte chiffré.

Type de chiffrement	Technique de chiffrement	Nombres de sources	Sources
Chiffrement symétrique	AES	3	[29], [31], [30]
	TDES	1	[14]
	McClaurin series-Laplace	1	[3]
	SE	1	[2]
Chiffrement asymétrique	ABE/RBE	1	[5]
	CPABE	1	[13]
	RSA	1	[29]
	NTRU	1	[17]
	MRFC	1	[8]
	IBE	1	[30]

Tableau 5 : Nombre de sources par techniques de chiffrement

5.4.2. Le contrôle d’accès

Nous venons de voir les méthodes de chiffrement qui permettent de garantir une certaine confidentialité des données, mais elles n’empêchent pas des intrus d' y accéder [28]. Ainsi, le contrôle d’accès consiste à établir des règles et des droits pour contrôler l’accès aux ressources partagées [1]. Il existe différentes approches pour définir ces règles notamment en fonction des informations d'identification de l'utilisateur, du type de ressource et des choix de confidentialité du propriétaire de la ressource [1]. Cependant, la grande majorité des solutions de contrôle d'accès reposent principalement sur des politiques/règles préalables et administrées manuellement. De plus, dans de nombreux mécanismes de contrôle d'accès, il y a une diminution relative des performances à mesure que le nombre d'entités impliquées dans le traitement augmente [1].

C’est pourquoi l’étude [1] conçoit un mécanisme de contrôle d'accès nommé Ontology-Based Access Control Model (OBACM) modélisant les entités impliquées dans dans le contrôle d'accès et leurs interrelations et qui effectue de manière automatique la désignation, le refus et la vérification des droits d’accès. Pour ce faire, le mécanisme se base sur les ontologies permettant la modélisation d’un domaine d’information. Les auteurs de [1]

pensent que la modélisation des entités dans l'ontologie peut considérablement améliorer les performances du système notamment grâce à la propriété de subsumption des ontologies qui permet de la réduire. En effet, [1] fournit 4 théorèmes pour effectuer cette réduction :

1. Si A et B sont des concepts et que A subsume B, alors les règles appliquées à A le sont aussi pour B.
2. Si A est un individu et B un concept, alors les règles appliquées à B le sont aussi pour A
3. S'il existe une relation sémantique entre différentes propriétés, alors de nouvelles propriétés qui ne sont pas explicitement mentionnées dans une ontologie peuvent être déduites.
4. Si A et B sont des individus, les règles définies pour A sont aussi applicables pour B.

L'article [2] mentionne le modèle de contrôle d'accès nommé Role Based Access Control (RBAC) qui se base sur la hiérarchie des rôles dans une organisation pour définir la politique d'accès. Cependant les auteurs soulèvent des vulnérabilités à son encontre notamment concernant les utilisateurs qui peuvent avoir des excès de privilèges dû au fait qu'un utilisateur peut avoir plusieurs rôles. Ainsi, [2] propose une méthode de prédiction du rôle d'un utilisateur lui conférant un niveau de priorité pour l'accès aux données. En effet, en fonction de paramètres tels que le nombre d'heures passées sur un système, le niveau d'expertise sur les ressources informatiques ou le nombre de rôles, la méthode calcule un niveau de priorité de l'utilisateur entre 0 (pas prioritaire) et 1 (très prioritaire).

Dans leur framework de sécurité cloud pour EHR, les auteurs de l'étude [5] proposent deux méthodes de gestion des accès. D'une part, il y a le Single Sign-On (SSO) qui permet aux utilisateurs de se connecter à plusieurs services après une unique connexion. Pour cela, les utilisateurs s'enregistrent auprès d'un Identity Provider (IdP) pour recevoir un token. Lorsqu'un utilisateur s'enregistre auprès d'un service du système EHR, celui-ci vérifie auprès de l'IdP si le token est bien vérifié avant d'authentifier l'utilisateur. D'autre part, les auteurs proposent la méthode Attribute-Based Access Control (ABAC) qui permet de paramétrer un ensemble de conditions complexes en fonction d'attributs pour définir les accès. ABAC est donc plus flexible que RBAC qui est basé sur les rôles de la hiérarchie prédéfinie. [13] et [31] proposent également un contrôle d'accès sur les attributs (ABAC) dans leur système.

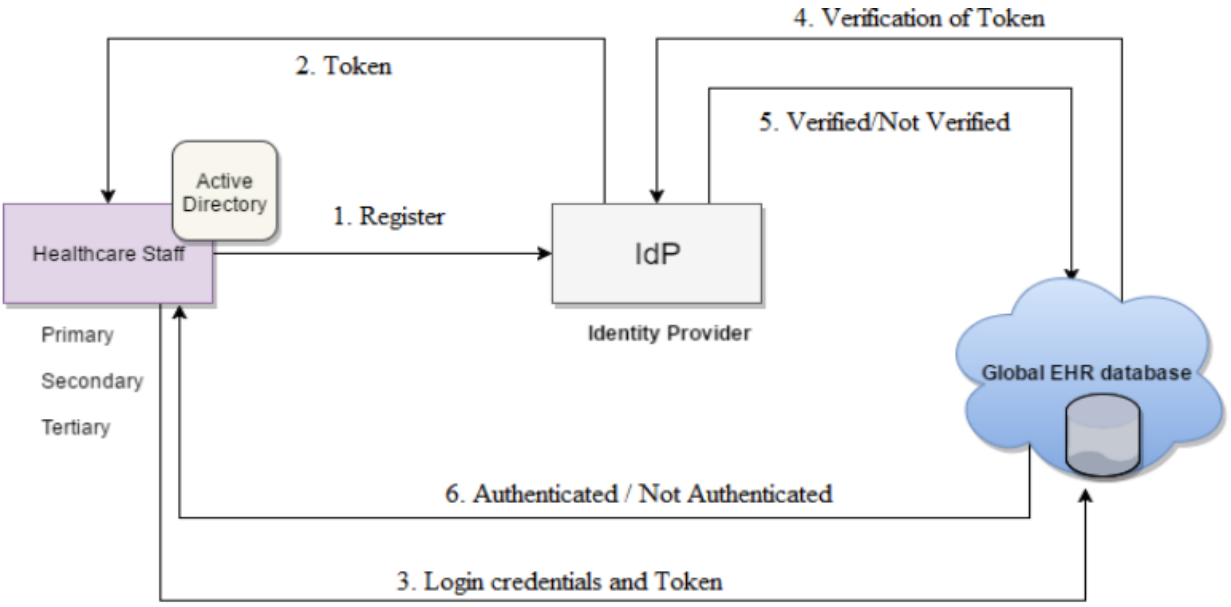

Figure 11 : Fonctionnement du SSO pour le système EHR [5]

A noter que les systèmes qui utilisent les algorithmes de chiffrement ABE ou RBE réalisent également un contrôle d'accès de manière implicite.

5.4.3. La prévention et détection de menaces

L'étude [12] présente le système Integrated Intrusion Prevention and Detection System (IIPDS) qui vise à améliorer les performances en matière de sécurité des systèmes cloud, notamment en ce qui concerne la confidentialité, l'intégrité et la disponibilité. Ce système repose sur deux approches préventives : l'utilisation d'un service tiers pour l'authentification et l'application de la couche de sécurité de transport appelée Transport Layer Security (TLS) avec le chiffrement des données en transit via le protocole SSL. De plus, il intègre deux méthodes de détection, dont l'utilisation d'un outil open source nommé Snort pour la détection d'attaques réseau basée sur des règles. Snort est capable de capturer les paquets de données circulant sur le réseau et de les analyser à la recherche d'empreintes

d’intrusion, telles qu'une utilisation excessive de caractères spécifiques. L'autre méthode de détection repose sur la détection d'anomalies pour identifier tout comportement anormal. Dans le cadre de la simulation de l'IIPDS, trois paramètres sont surveillés : l'adresse IP source, la différence entre les intervalles de temps d'envoi et de réception des paquets de données, ainsi que le nombre de paquets transmis durant une certaine période choisie. Lorsque IIPDS identifie une anomalie, il déclenche des alertes avec une réactivité variable en fonction de la gravité (low, moderate et high) des attaques présenté sur la Figure 9. Une attaque grave implique une alerte immédiate. IIPDS utilise un seuil de 3 pour changer la priorité d’un type d’attaque, c’est-à-dire qu’au bout de 3 attaques similaires, la priorité de celle-ci augmente.

Attack type	Priority
DoS or DDoS	H (high)
Data availability and integrity attacks	H
Infrastructure attacks	H
Data confidentiality attack	M (Moderate)
Confidentiality breach	L (Low)

Figure 12 : Priorité des types d’attaques [12]

L’étude [9] propose un système de défense multicouche basé sur les réseaux définis par logiciel (SDN) pour détecter et atténuer les attaques DDoS dans les environnements cloud .Pour ce faire, les auteurs utilisent un Software-Defined Networking (SDN) qui permet de séparer le contrôle du plan logique dans les réseaux. Il permet notamment de programmer des applications réseau essentielles comme les pare-feu, les systèmes de détection d'intrusion, et les algorithmes de routage. Le contrôle centralisé offre une gestion flexible des ressources réseau, améliorant la vitesse et l'efficacité de l'utilisation des ressources. Pour détecter les attaques, ils utilisent un algorithme de machine learning supervisé nommé Support Vector Machine (SVM). SVM est utilisé pour créer un hyperplan dans un espace de grande dimension afin de catégoriser le trafic réseau en trafic habituel et malveillant. L’algorithme utilise plusieurs paramètres lors de son entraînements qui sont le nombre d'adresses IP sources entrantes par unité de temps par seconde (SSIP), l'écart type des paquets de flux entrants (SDFP), la vitesse des entités de flux (SFE), et le nombre total de flux (RFIP). Ainsi, les données du trafic réseau sont collectées par le SDN et les fournis à la SVM pour qu'elle puisse effectuer la classification du trafic. En cas de détection d’une attaque, le contrôleur du

SDN met en œuvre des mesures de sécurité, telles que la redirection du trafic ou l'activation impact de listes de contrôle d'accès pour limiter l'impact de l'attaque.

5.4.4. La classification des données

L'étude [10] propose un algorithme de sécurité nommé Data Mobility Security Algorithm (DMoS) qui utilise une technique de classification de données nommée Hadoop Distributed File System (HDFS). C'est un outil utilisé dans les systèmes cloud pour le traitement de grandes quantités de données et qui consiste à diviser les données en petits fichiers. Tout d'abord, les données qui entrent sur le serveur cloud sont classées en 3 catégories en fonction de leur sensibilité : publiques, privées et spéciales. Pour ce faire, une valeur AMD (Attribute Metadata) est calculée pour chaque donnée puis ajoutée dans le fichier de métadonnées respectives à chaque catégorie. Les fichiers peuvent prendre différents types de format (txt, pdf, sql, doc, image, audio ou xml). Le système HDFS va diviser les données en bloc de 128 Mo avant de les traiter afin de leur attribuer la valeur AMD. Une fois le niveau de sensibilité identifié, les fichiers privés et spéciaux sont soumis à un chiffrement avec l'algorithme DMoS. Dans ce processus, les données sont chiffrées avant d'être envoyées. Une fois que les données arrivent à destination, elles sont vérifiées pour s'assurer qu'elles n'ont pas été altérées en cours de route. Cette approche garantit la confidentialité et l'intégrité des données pendant leur transit dans le cloud.

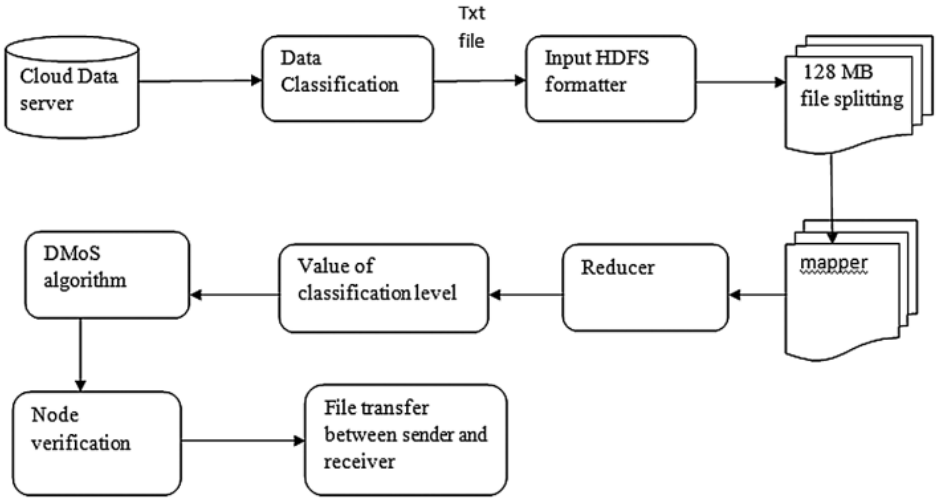

Figure 13 : Système proposé par [10]

Les auteurs de [11] proposent un nouveau type de service cloud qu'ils nomment Confidentiality-based data Classification-as-a-Service (C2aaS). Ce service aide les systèmes cloud à choisir les contrôles de sécurité et l'emplacement de stockage en fonction du niveau de sensibilité des données. Pour ce faire, C2aaS divise les données en 2 catégories : les données confidentielles et non confidentielles. Cette classification de données se base sur les noms d'attributs à l'aide de d'un algorithme de classification nommé Semantic-kNN. C'est un concept combinant les idées de la recherche des k-Nearest Neighbor (k-NN) qui est un algorithme de machine learning supervisé mais avec en plus des aspects sémantiques pour améliorer la précision de la classification étant donné que dans ce cas, la signification sémantique est importante pour connaître les données confidentielles. Une fois ces données confidentielles identifiées, elles sont chiffrées contrairement aux données non confidentielles.

L'étude [26] propose un algorithme nommé Security Based-Distributed Storage (SB-DS). Ce système permet de stocker des données de manière sécurisée et efficace grâce à un travail de découpage où les données sont labellisées pour être classées comme sensibles ou non. Les données sensibles sont ensuite divisées en 2 parties : une partie des données est sélectionnée aléatoirement et l'autre correspond aux données restantes. Ces deux parties sont chiffrées à l'aide d'une clé générée également de manière aléatoirement avant d'être envoyé séparément vers un serveur cloud respectif. Pour récupérer les données originales, une opération de fusion est réalisée entre les deux parties puis un déchiffrement à l'aide de la clé.

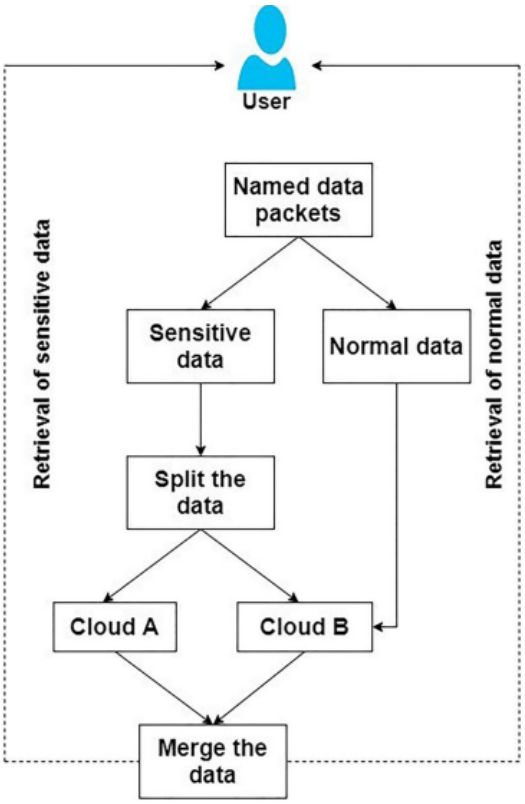

Figure 14 : Système SB-DS proposé par [26]

4.4.5. La réplication de données

L'étude [4] propose un système de réplication de données que les auteurs nomment Secured Data Replication Management Scheme (SDRMS). Ces derniers tentent de répondre à deux problématiques majeures concernant la gestion des réplicas. En effet, d'une part lorsque les données originales sont modifiées, les réplicas doivent également être mis à jour au risque de créer des incohérences de données. D'autre part, la création de réplicas pour des données sans importances peut consommer des ressources mémoires inutilement. Ainsi, [4] utilise un Replica Management Agent (RMA) pour créer les réplicas de données. Ce RMA contrôle plusieurs serveurs cloud et se charge de créer les répliques en fonction de la charge des serveurs et du taux d'accès à une donnée dont la formule est donnée dans la Figure 9. Pour ce faire, il construit une structure en arbre où les serveurs les moins chargés se retrouvent en première couche avec les données dont le taux d'accès est le plus élevé. Si un serveur cloud de première couche devient plus lourd qu'un serveur de deuxième couche, ils interchangent.

Concernant le taux de fréquence des données, le RMA calcule ce taux de fréquence toutes les 5 minutes et réorganise les répliques. De plus, il enregistre la location de celles-ci, ce qui permet de les mettre à jour rapidement en cas de modification des données originales.

$$Freq_{adata} = \frac{Request\ for\ data_i}{Total\ number\ of\ requests}$$

Figure 15 : Formule pour calculer le taux d'accès d'une donnée [4]

5.5. Analyse comparative

Dans la section précédente, nous avons passé en revue les 5 types de méthodes de sécurité utilisées pour protéger les données stockées dans le cloud. Dans cette section, nous entrons dans la phase d'analyse comparative de ces méthodes. L'objectif de cette section est d'évaluer et de comparer les différentes méthodes entre elles. Cependant, compte tenu du manque de méthode dans la littérature, les méthodes de prévention d'attaques ainsi que de réplication de données ne feront pas partie de cette analyse. Ainsi, l'analyse comparative se concentre davantage sur le chiffrement, le contrôle d'accès et la classification de données.

5.5.1. Comparaison des méthodes de chiffrement

Lors de l'évaluation des différentes méthodes de chiffrement de notre échantillon, plusieurs critères sont pris en compte que nous avons classé selon 3 catégories :

Les critères de sécurité se concentrent sur la capacité des méthodes de chiffrement à protéger les données contre les attaques. La NIST propose pour cela de comparer la longueur de clé de chiffrement qui est une valeur universelle exprimée en octet [34]. En effet, plus la clé est longue, plus l'algorithme est résistant à une attaque par force brute [34]. L'utilisation de nombres aléatoires dans le processus est également un critère qui renforce la sécurité de la méthode [8].

Les critères de performance évaluent l'efficacité des méthodes de chiffrement, notamment en termes de vitesse lors du chiffrement et du déchiffrement des données ou encore le débit de données qui est la capacité de la méthode à traiter un volume élevé de données en un temps donné.

Enfin, nous avons les critères d'utilisation des ressources qui examinent l'impact des méthodes de chiffrement sur les ressources informatiques, comprenant le coût de stockage nécessaire pour stocker les données chiffrées ou encore la mémoire nécessaire pour réaliser toutes les opérations.

Étant donné que chaque étude a utilisé des calculs spécifiques pour évaluer ces critères, il est difficile de comparer les résultats bruts entre celles-ci. C'est pourquoi, nous avons adopté une approche méthodique consistant à attribuer un score sur une échelle de 1 à 3 pour chaque critère évalué. Cette démarche nous permet de normaliser les résultats. Par défaut, si un critère n'est pas renseigné dans l'étude, le score attribué est 2 pour ne pas impacter le score global. Le score final est calculé en prenant une moyenne pondérée des scores attribués à chaque critère évalué, avec un coefficient de 2 pour les critères de sécurité et un coefficient de 1 pour les autres critères. Cette pondération permet de mettre davantage l'accent sur la sécurité tout en tenant compte des autres aspects importants de l'évaluation. Les méthodes ABE/RBE et SE [2] ne sont pas évaluées par manque de données.

Nous obtenons ainsi le Tableau 4 suivant :

	Sécurité		Performance		Utilisation des ressources			
Méthode de chiffrement	Longueur de la clé de chiffrement (en octets) (a)	Utilisation d'un nombre aléatoire (b)	Temps de chiffrement/ déchiffrement (c)	Débit de données (d)	Coût de stockage des données chiffrées (e)	Utilisation de la mémoire (f)	**SCORE FINAL**	Sources
TDES	2 (168)		2	3		1	**1.78**	[14], [8]
CPABE		3	3		3		**2.22**	[13]
NTRU	3 (256)	3				3	**2.33**	[17]
AES	2 (128-256)	3	1			2	**1.89**	[26], [29], [31], [3], [8], [30]
RSA	2 (128-256)	3	1				**1.89**	[29], [30]
IBE	2 (160)	3	3				**2.11**	[30]
SUCDDES	1 (32-64)	3	3				**1.89**	[31]
MRFC	3 (256)	3	2			3	**2.33**	[8]

Tableau 6 : Evaluation des méthodes selon les critères de sécurité, performance et utilisation des ressources sélectionnées

Légende du Tableau 6 :

Longueur de clé, Utilisation d'un nombre aléatoire : 1 - Faible, 2- Moyenne, 3 - Forte
Temps de chiffrement/déchiffrement : 1- Lent, 2- Moyen, 3 - Rapide
Débit de données : 1 - Faible débit, 2- Moyen débit, 3 - Fort débit
Coût en stockage : 1 - Fort coût, 2- Moyen coût, 3 - Faible coût
Utilisation de la mémoire : 1 - Forte utilisation, 2- Moyenne utilisation, 3 - Faible utilisation
Les cases vides sont les scores par défaut égale à 2 car ces critères ne sont pas renseignés dans les articles.

Score Final = (2 * (a + b) + c + d + e + f) / 9

Pour notre analyse, nous avons attribué un poids plus élevé aux critères de sécurité afin de rendre compte de l'importance de la sécurité dans le choix d'une méthode de chiffrement. Les autres critères sont significatifs mais moins critiques que la sécurité dans notre évaluation globale des méthodes de chiffrement. Ainsi, nous obtenons des scores finaux compris entre 1 et 3. Un score de 3 signifie que la méthode est significative par rapport à nos critères.

Les méthodes NTRU et MRFC obtiennent le meilleur score avec 2.33 grâce à leur longueur de clé de 256 octets et l'utilisation d'un nombre aléatoire dans leur algorithme. De plus, elles nécessitent une utilisation de la mémoire minimale.

La méthode CPABE obtient un bon score de 2.22 qui s'explique par l'utilisation d'un nombre aléatoire pour le chiffrement, un temps de chiffrement/déchiffrement rapide ainsi qu'un coût de stockage du texte chiffré faible.

Ensuite, deux méthodes ont obtenu un score de 2.11. Nous retrouvons la méthode IBE qui possède une taille de clé moyenne par rapport à d'autres méthodes avec une longueur de 160 octets mais qui utilise un nombre aléatoire tout en assurant une vitesse de chiffrement rapide.

Trois autres méthodes obtiennent un score similaire égale à 1.89. Ce résultat peut être expliqué par l'absence d'informations détaillées dans l'article, ce qui rend difficile l'attribution de scores spécifiques pour certains critères. Néanmoins, ce score nous offre une indication générale de leur performance. AES et RSA obtiennent exactement les mêmes scores dans les critères de sécurité, et ils affichent une vitesse de chiffrement plus lente que les autres méthodes. SUCDDES obtient le même score mais avec une longueur de clé très faible par rapport aux autres. En effet, la longueur de la clé varie de 32 à 64 octets contre 256 octets pour les meilleures méthodes.. Cependant, elle génère un nombre aléatoire dans son processus de chiffrement et possède une très bonne vitesse de chiffrement.

Enfin, nous retrouvons TDES en dernière position de notre étude avec un score de 1.78. Cela s'explique par une longueur de clé moyenne à 168 octets ainsi qu'une vitesse de chiffrement moyenne. S'ajoute à cela une forte utilisation de la mémoire pour son processus de chiffrement ce qui affecte ses performances.

5.5.2. Comparaison des méthodes de contrôle d'accès

Dans [37], NIST a catégorisé différentes métriques d'évaluation pour les systèmes de contrôle d'accès. Le document classe les critères d'évaluation des systèmes de contrôle d'accès en quatre catégories :

Administration : Les caractéristiques qui généralement influencent les coûts, l'efficacité et les performances de l'administration d'un système de contrôle d'accès.

Mise en œuvre : Les caractéristiques des mécanismes ou des algorithmes utilisés par le système de contrôle d'accès pour mettre en œuvre les modèles et règles.

Performance : Les caractéristiques qui affectent les performances en plus de la mise en œuvre des processus du système de contrôle d'accès.

Support : Les caractéristiques qui ne sont pas essentielles mais peuvent améliorer la convivialité et la portabilité d'un système de contrôle d'accès.

Pour notre analyse, nous décidons de réaliser notre comparaison basée uniquement sur 3 critères de performance. présenté dans le Tableau 5. La méthode de l'étude [2] n'est pas prise en compte par manque de données. Ainsi, nous attribuons les scores de 1 à 3 pour chaque méthode dans le tableau suivant :

Méthode de contrôle d'accès	Temps de réponse des demandes d'accès (a)	Stockage et récupération des politiques/règles (b)	Intégration de fonction d'authentification (c)	**SCORE FINAL** **(a+b+c) / 3**	Sources
OBACM	3	3		**2.67**	[1]
Framework EHR	2		3	**2.33**	[5]
CPABE (hiérarchique)	3	3		**2.67**	[13]

Tableau 7 : Evaluation des méthodes selon les critères de performances sélectionnées

Légende du Tableau 7 :

Temps de réponse des demandes d'accès : 1- Lent, 2- Moyen, 3- Rapide
Stockage et récupération des politiques/règles : 1- Stockage lourd, 2 - Moyen; 3- Faible
Intégration de fonction d'authentification : 1- Faible intégration, 2- Moyenne, 3- Forte

Les méthodes OBACM et CPABE obtiennent le meilleur score avec 2,67. Cela s'explique d'une part grâce à leur faible coût de stockage des règles d'accès qui sont réduites par rapport à des méthodes standards, grâce à la propriété de subsumption dans OBACM ou à une structure en hiérarchie dans CPABE. D'autre part, ces mêmes propriétés justifient leur temps de réponse rapide ainsi que l'utilisation d'un serveur tiers pour CPABE

Quant au framework de sécurité EHR, il obtient un score de 2.33 grâce notamment à l'intégration de solutions d'authentification avec le SSO permettant un temps de réponse rapide pour les demandes d'accès.

5.5.3. Comparaison des méthodes de classification de données

Contrairement aux deux méthodes précédentes, les articles que nous avons étudiés ne fournissent pas de métriques de performance ni de critères d'évaluation explicites qui nous permettraient de comparer les méthodes de classification de manière précise. Toutefois, nous pouvons toujours procéder à une comparaison en nous basant sur certaines caractéristiques spécifiques des méthodes de classification.

L'article [8] se distingue par son choix de classification basée sur les préférences du propriétaire des données plutôt que sur des algorithmes de machine learning, comme le kNN, reconnu pour son efficacité dans la résolution de problèmes de classification conventionnels [35]. Cependant, cette méthode est étroitement liée aux données utilisées pour l'entraînement et se révèle moins efficace pour évaluer le niveau de confidentialité des données, notamment dans des domaines sensibles comme la finance ou la santé [14].

En revanche, [11] adopte la technique Semantic-kNN (Sk-NN), une amélioration du kNN. Cette approche intègre un processus de filtrage sémantique ("semantic filtration process"), conçu pour être intelligent et capable d'identifier les données confidentielles et non confidentielles sans intervention humaine [35].

Dans [14], avant de procéder au chiffrement TDES, un administrateur crée un masque de données pour classer les attributs en trois catégories : faiblement sensibles, moyennement sensibles et hautement sensibles.

La méthode présentée dans l'article [10] utilise le système de fichiers distribué HDFS pour classifier les données en trois catégories : privées, publiques et spéciales, réservées à des destinataires spécifiques, caractérisées par leur haut degré de confidentialité [10].

En résumé, il est important de noter que ces méthodes de classification varient de la classification manuelle à la classification automatique grâce à des algorithmes. La comparaison directe entre ces méthodes est difficile en raison du manque de métriques d'évaluation.

6. Discussion

Dans cette section de discussion, nous allons tenter de répondre à nos trois sous-questions de recherche à partir de notre revue systématique de la littérature. Ces questions ont été formulées en amont de notre étude pour explorer en détail les vulnérabilités de sécurité qui menacent les données stockées dans le cloud, les différentes méthodes existantes pour sécuriser efficacement ces données, et enfin, pour comprendre comment les organisations peuvent choisir la méthode de sécurisation des données.

Nous allons donc répondre aux différentes sous-questions que nous nous sommes posées avant notre revue de la littérature.

Tableau 8 : Réponses à nos sous-questions de recherche

Sous-question de recherche	Réponse
Q1 : Quelles sont les vulnérabilités de sécurité qui peuvent affecter les données stockées dans le cloud ?	Nous avons identifié 7 vulnérabilités qui touchent les données stockées dans des environnements cloud. Ces dernières sont fréquemment mentionnées dans la littérature ce qui donne une indication de l'importance de chaque vulnérabilité en matière de sécurité des données stockées dans le cloud. La perte ou fuite de données semble être perçue comme la plus préoccupante par la communauté de la recherche, suivie de la confidentialité, la disponibilité, l'intégrité et la violation des données. Les vulnérabilités de ségrégation ainsi que de virtualisation des données ne sont pas fréquemment mentionnées.
Q2 : Quelles sont les méthodes existantes pour sécuriser les données stockées dans le cloud ?	Notre recherche nous a permis de classifier les méthodes de sécurité en 5 catégories qui sont le chiffrement, le contrôle d'accès, la classification de données, la prévention ou détection d'attaques et enfin la réplication de données. De la même manière que pour les vulnérabilités, nous avons examiné la fréquence d'apparition de chaque méthode dans la littérature indiquant leur pertinence dans le contexte de sécurité des données dans le cloud.

	Ainsi, les méthodes de chiffrement sont les plus proposées suivi des méthodes de classification de données, de contrôle d'accès. Nous n'avons trouvé qu'un article pour les méthodes de prévention d'attaques ainsi que de réplications de données.
Q3 : Comment les organisations peuvent-elles choisir la méthode de sécurisation des données ?	Pour répondre à Q3, nous avons réalisé une analyse comparative pour les méthodes de chiffrement, de contrôle d'accès et de classification des données. Concernant les méthodes de chiffrement, nous nous sommes basés sur une métrique d'évaluation proposée par la NIST et d'autres renseignées dans nos articles. Nous retrouvons la longueur de la clé de chiffrement, la génération de nombre aléatoire pour le chiffrement, la vitesse de chiffrement et déchiffrement, le débit de données, le coût de stockage des données chiffrées et l'utilisation de la mémoire pour les calculs. Pour les méthodes de contrôle d'accès, nous avons utilisé des métriques d'évaluation des systèmes de contrôle d'accès proposées par la NIST qui comprend le temps de réponse des demandes d'accès, le stockage des politiques ou règles d'accès ainsi que l'intégration de fonction d'authentification dans les systèmes. Cela nous a permis d'attribuer des scores de 1 à 3 et calculer un score final pour les méthodes de chaque type afin de les comparer entre elles. Nous n'avons pas trouvé de critères d'évaluation standard pour les méthodes de classification. Néanmoins, nous avons pu réaliser une comparaison basée sur leur mode de classification qui peut être manuel à l'aide de label ou automatique grâce à des algorithmes. Ainsi, nous recommandons aux entreprises d'utiliser des métriques d'évaluations standardisées pour comparer et choisir les méthodes de sécurité adaptées à leurs exigences comme nous l'avons réalisé

	dans cette recherche. Dans le cas contraire, il est tout de même possible de comparer les méthodes en fonction d'autres critères de comparaison, comme nous l'avons réalisé pour les méthodes de classification de données.

7. Conclusion

A travers notre recherche, nous comprenons que le cloud computing présente de nombreux avantages favorisant l'adoption de cette technologie au sein des entreprises. En effet, cette technologie continue de prendre de l'ampleur avec l'émergence de nouveaux services cloud à destination de celles-ci. Les facteurs d'adoptions principaux du cloud sont la facilitation de mise en place, l'évolutivité mais surtout la réduction des coûts. A l'inverse, l'un des principaux freins à son adoption concerne les risques de sécurité notamment liés aux données. Bien que les fournisseurs de services cloud sécurisent les données qu'ils stockent, les entreprises peuvent également mettre en œuvre des méthodes pour renforcer la sécurité de leurs données.

Ainsi, nous avons mené cette revue afin de recenser l'ensemble des méthodes de sécurité proposées dans la littérature. Nous avons identifié cinq catégories de méthodes de sécurité, englobant le chiffrement, le contrôle d'accès, la classification des données, la prévention et la détection d'attaques, ainsi que la réplication des données. Une analyse comparative a été réalisée pour identifier les meilleures méthodes, en prenant en compte des critères spécifiques à chaque type de méthode, notamment en termes de sécurité et de performance. Nous pouvons souligner un manque de littérature concernant les méthodes de réplications de données ou encore de préventions et détections d'attaques. C'est pourquoi une future recherche sur ces 2 types de méthodes pourrait être intéressante.

Nous avons identifié des approches techniques visant à renforcer la sécurité dans le cloud, bien que celles-ci puissent exiger une expertise spécifique. Toutefois, il est important de mener des analyses d'impact dès les premières étapes des projets afin de mieux les anticiper. Cette approche est d'ailleurs préconisée par le NIST à travers son concept nommé "Privacy by Design". En effet, ce cadre de sécurité suggère de considérer les problématiques de sécurité, notamment de confidentialité des données, dès la conception et tout au long du processus de développement des systèmes. Les entreprises doivent notamment se montrer entièrement transparentes sur la manière dont elles traitent les données et prendre toutes les mesures nécessaires pour protéger les données personnelles dont elles ont la responsabilité. S'ajoute à cela la formation et sensibilisation des équipes de développement qui doivent être formées et sensibilisées aux principes de Privacy by Design pour qu'elles puissent les mettre en pratique. Finalement, cette approche vise à changer la perception des entreprises vis-à-vis de la sécurité, en la faisant passer d'une simple contrainte dans les projets à une composante intégrée et essentielle.

8. Bibliographie

[1] Michael, Auxilia, Raja Kothandaraman, and Kannan Kaliyan. “Providing Ontology-Based Access Control for Cloud Data by Exploiting Subsumption Property among Domains of Access Control.” *International Journal of Intelligent Engineering and Systems* 12, no. 3 (2019): 280–91. https://doi.org/10.22266/ijies2019.0630.27.

[2] Jyosthna, Mrs. P., and Dr. Konala Reddy. “User Prediction in a Role for Secure Data Sharing Through Cloud.” *International Journal of Innovative Technology and Exploring Engineering* 8, no. 10 (2019): 162–66. https://doi.org/10.35940/ijitee.g5425.0881019.

[3] Namratha*, P, and C Shoba Bindu. “A Systematic Framework for Securing Cloud Data.” *International Journal of Recent Technology and Engineering (IJRTE)* 8, no. 3 (2019): 4191–96. https://doi.org/10.35940/ijrte.c6000.098319.

[4] Jayalekshmi, M. B., and Dr. S.H. Krishnaveni. “Secure Data Replication Management Scheme with Better Data Accessibility.” *International Journal of Innovative Technology and Exploring Engineering* 8, no. 12 (2019): 4484–88. https://doi.org/10.35940/ijitee.l3517.1081219.

[5] Ganiga, Raghavendra, Radhika M. Pai, Manohara Pai M. M., and Rajesh Kumar Sinha. “Security Framework for Cloud Based Electronic Health Record (EHR) System.” *International Journal of Electrical and Computer Engineering (IJECE)* 10, no. 1 (2020): 455. https://doi.org/10.11591/ijece.v10i1.pp455-466.

[6] Campêlo, Robson A., Marco A. Casanova, Dorgival O. Guedes, and Alberto H. Laender. “A Brief Survey on Replica Consistency in Cloud Environments.” *Journal of Internet Services and Applications* 11, no. 1 (2020). https://doi.org/10.1186/s13174-020-0122-y.

[7] Rafique, Ansar, Dimitri Van Landuyt, Emad Heydari Beni, Bert Lagaisse, and Wouter Joosen. “CryptDICE: Distributed Data Protection System for Secure Cloud Data Storage and Computation.” *Information Systems* 96 (2021): 101671. https://doi.org/10.1016/j.is.2020.101671.

[8] Sumathi, M., and S. Sangeetha. “A Group-Key-Based Sensitive Attribute Protection in Cloud Storage Using Modified Random Fibonacci Cryptography.” *Complex & Intelligent Systems* 7, no. 4 (2020): 1733–47. https://doi.org/10.1007/s40747-020-00162-3.

[9] Mishra, Shailendra, Sunil Kumar Sharma, and Majed A. Alowaidi. “Multilayer Self-Defense System to Protect Enterprise Cloud.” *Computers, Materials & Continua* 66, no. 1 (2020): 71–85. https://doi.org/10.32604/cmc.2020.012475.

[10] Punithavathi, R., M. Kowsigan, R. Shanthakumari, Miodrag Zivkovic, Nebojsa Bacanin, and Marko Sarac. “Protecting Data Mobility in Cloud Networks Using Metadata Security.” *Computer Systems Science and Engineering* 42, no. 1 (2022): 105–20. https://doi.org/10.32604/csse.2022.020486.

[11] Ali, Munwar, 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 (2023): 749–60. https://doi.org/10.1016/j.aej.2022.10.056.

[12] Alturfi, Sabah M., Dena Kadhim Muhsen, Mohammed A. Mohammed, Israa T. Aziz, and Mustafa Aljshamee. “A Combination Techniques of Intrusion Prevention and Detection for Cloud Computing.” *Journal of Physics* 1804, no. 1 (February 1, 2021): 012121. https://doi.org/10.1088/1742-6596/1804/1/012121.

[13] Mujawar, Tabassum N., and Lokesh B. Bhajantri. “An Attribute-Based Encryption Method Using Outsourced Decryption and Hierarchical Access Structure.” *Journal of Telecommunications and Information Technology* 2, no. 2022 (2022): 75–81. https://doi.org/10.26636/jtit.2022.158421.

[14] Ramachandra, Mohan Naik, Madala Srinivasa Rao, Wen Cheng Lai, Bidare Divakarachari Parameshachari, Jayachandra Ananda Babu, and Kivudujogappa Lingappa Hemalatha. “An Efficient and Secure Big Data Storage in Cloud Environment by Using Triple Data Encryption Standard.” *Big Data and Cognitive Computing* 6, no. 4 (2022): 101. https://doi.org/10.3390/bdcc6040101.

[17] S, Malathi., Dr.L. Malathi, and N.Nasurdeen Ahamed. “Hybrid Cloud Storage For Secure Authorization And Information Hiding.” *International Journal of Engineering and Advanced Technology* 8, no. 6s (2019): 388–93. https://doi.org/10.35940/ijeat.f1082.0886s19.

[19] A. Syed, K. Purushotham and G. Shidaganti, "Cloud Storage Security Risks, Practices and Measures: A Review," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298281.

[20] A. Patel, N. Shah, D. Ramoliya and A. Nayak, "A detailed review of Cloud Security: Issues, Threats & Attacks," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 758-764, doi: 10.1109/ICECA49313.2020.9297572.

[21] R. Kumar and M. P. S. Bhatia, "A Systematic Review of the Security in Cloud Computing: Data Integrity, Confidentiality and Availability," 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020, pp. 334-337, doi: 10.1109/GUCON48875.2020.9231255.

[22] Nagarajan, G. and Kumar, K.S. (2019). A Security Risk on Data Storage in Cloud based System –Survey. International Journal of Emerging Technologies, 10(2): 195–199.

[25] G. Nagarajan and K. Sampath Kumar, "Security Threats and Challenges in Public Cloud Storage," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 97-100, doi: 10.1109/ICACITE51222.2021.9404641.

[26] Venkatesan, B., and S. Chitra. “Retracted: An Enhance the Data Security Performance Using an Optimal Cloud Network Security for Big Data Cloud Framework.” *International Journal of Communication Systems* 35, no. 16 (2021). https://doi.org/10.1002/dac.4854.

[28] Alghofaili, Yara, Albatul Albattah, Noura Alrajeh, Murad A. Rassam, and Bander Ali Al-rimy. “Secure Cloud Infrastructure: A Survey on Issues, Current Solutions, and Open Challenges.” *Applied Sciences* 11, no. 19 (2021): 9005. https://doi.org/10.3390/app11199005.

[29] Y. Sharma, H. Gupta and S. K. Khatri, "A Security Model for the Enhancement of Data Privacy in Cloud Computing," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 898-902, doi: 10.1109/AICAI.2019.8701398.

[30] Khashan, Osama Ahmed. “Secure Outsourcing and Sharing of Cloud Data Using a User-Side Encrypted File System.” *IEEE Access* 8 (2020): 210855–67. https://doi.org/10.1109/access.2020.3039163.

[31] P. J. Rani and M. Akkalakshmi, "An Optimized Sucddes to Control Access in Cloud," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 2019, pp. 1-5, doi: 10.1109/ICAC347590.2019.9036829.

[32] The NIST definition of cloud computing. Accessed September 19, 2023. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf.

[33] Recommandations pour les entreprises qui envisagent de souscrire ... - CNIL. (n.d.). https://www.cnil.fr/sites/cnil/files/typo/document/Recommandations_pour_les_entreprises_qui_envisagent_de_souscrire_a_des_services_de_Cloud.pdf

[34] CRYPTOGRAPHIC ALGORITHM METRICS - NIST Computer Security Resource Center. Accessed September 19, 2023. https://csrc.nist.gov/files/pubs/conference/1997/10/10/proceedings-of-the-20th-nissc-1997/final/docs/128.pdf.

[35] Ali, M., Jung, L. T., Abdel-Aty, A. H., Abubakar, M. Y., Elhoseny, M., & Ali, I. (2020). Semantic-k-NN algorithm: An enhanced version of traditional k-NN algorithm. *Expert Systems with Applications*, *151*, 113374.

[36] Rethlefsen, Melissa & Page, Matthew. (2022). PRISMA 2020 and PRISMA-S: common questions on tracking records and the flow diagram. Journal of the Medical Library Association : JMLA. 110. 253-257. 10.5195/jmla.2022.1449.

[37] Hu, Vincent C., and Karen Scarfone. *Guidelines for Access Control System Evaluation Metrics*, 2012. https://doi.org/10.6028/nist.ir.7874.

[40] Randell, Beverley. "Fr." Amazon, 1996. https://aws.amazon.com/fr/what-is-cloud-computing/.

[41] Premier Ministre - Agence nationale de la Sécurité des Systèmes D ... (n.d.-a). https://www.ssi.gouv.fr/uploads/2014/12/secnumcloud-referentiel-exigences-v3.2.pdf

[42] *Le règlement général sur la protection des Données - RGPD*. CNIL. (n.d.). https://www.cnil.fr/fr/reglement-europeen-protection-donnees

[43] Kitchenham, Barbara & Charters, Stuart. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. 2.

9. Table des figures

10. Liste des tableaux