ROYAUME DU MAROC
Université Mohammed V - RABAT -
جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم
CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre 3590

THESE

En vue de l'obtention du **DOCTORAT**

Structure de Recherche : Intelligent Processing Systems & Security (IPSS)
Discipline : Informatique
Spécialité : Blockchain et Sécurité Informatique

Présentée et soutenue le 22/01/2022 par :

Salima TRICHNI

Nouvelles Contributions Dans Le Contexte de La Sécurité Informatique Dédiées Aux Technologies Disruptives

JURY

Faouzia BENABBOU	PES,	FS Ben M'Sik de Casablanca, Université Hassan II	Présidente
Soumia ZITI	PES,	Université Mohammed V, Rabat. Faculté des Sciences	Rapporteur/Examinateur
Abderrahim TRAGHA	PES,	Université Hassan II, Casablanca. Faculté des Sciences Ben M'Sik	Rapporteur/Examinateur
Mouad BENMAMOUN	PES,	Université Mohammed V, Rabat. Faculté des Sciences	Rapporteur/Examinateur
Ahmed EL YAHYAOUI	PA,	Université Mohammed V, Rabat. Faculté des Sciences	Invité
Fouzia OMARY	PES,	Université Mohammed V, Rabat. Faculté des Sciences	Directrice de thèse

Année Universitaire : 2021/2022

Dédicaces

À Allah Seigneur de l'univers

À son Prophète Mohammed (bénédiction et paix sur lui)

À l'homme et la femme de ma vie, ma mère et mon père, mon exemple éternel, ceux qui se sont toujours sacrifiés pour me voir réussir, mes premier enseignants les plus influents, eux qui m'ont doté d'une éducation digne qui m'ont appris les valeurs de la vie, qui m'encouragent et que leur amour a fait de moi ce que je suis aujourd'hui. À mes parents : que ce rapport soit un meilleur cadeau que je puisse vous offrir.

À mon soutien et mon compagnon qui m'a chaleureusement supporté et encouragé tout au long de ce parcours. Mon mari, ma source d'amour et de force, celui qui a partagé avec moi les moments d'émotion et m'a aidé à surmonter toutes les difficultés.

À la lumière de mes jours, la flamme de mon cœur, ma source de motivation et d'inspiration. Mes enfants « Yahya » et « Ahmed» que Dieu vous protège.

À mon très cher frère Youssef et ma très chère sœur Nabila, pour leur amour, leur rigueur et leur soutien permanent. Je vous souhaite ainsi que vos chaleureuses petites familles, une vie pleine de bonheur et de succès.

À ma très chère professeur OMARY Fouzia et son aimable famille particulièrement la chère Ghita, pour leurs précieux conseils, leur disponibilité, leur goût, leur aide permanent à la fois scientifique et moral. J'ai énormément l'honneur de faire votre connaissance

À la mémoire de mon beau père que le rebond de ses prières m'accompagnait toujours, que Dieu le garde dans son vaste paradis.

À ma belle-famille, de plus petit à ma belle-mère, pour l'amour qu'ils me réservent.

À mes amis, au nom de l'amitié qui nous réunit pour Allah seul.

À mes collègues du travail pour leur esprit de Solidarité.

À tous mes proches et ceux qui me sont chers.

Je dédie ce modeste travail

Remerciements

Tout d'abord, Louange et Reconnaissance à Allah Seigneur de l'univers, le Très-Haut et le Tout Puissant, de nous avoir donné courage, patience et volonté, ainsi nous a guidés pour accomplir ce présent travail. Certes, le chemin ne manque pas d'obstacles et de défis mais comme a dit le prophète Mohammed ﷺ : « Sache que la victoire vient après la patience, et le soulagement après l'angoisse ; et qu'avec la difficulté il y a aisance ».

Arrivé au terme, Je tiens à remercier toutes les personnes qui, par leurs conseils et leurs encouragements, ont contribué à la réussite de ce travail.

La présente thèse est le fruit d'un grand effort effectué au sein du laboratoire Intelligent Processing Systems & Security (IPSS) de la Faculté des Sciences de Rabat - Université Mohammed V, Maroc. Sous la direction et l'encadrement de **Mme. Fouzia OMARY**, Professeur d'Enseignement Supérieur à la Faculté des Sciences de Rabat - Université Mohammed V et que je qualifie d'une dame vraiment exceptionnelle à qui je souhaitais toujours arriver à ce stade, pour graver dans ce mémoire, mes sincères expressions et sentiments de gratitude et de respect envers leur personne. Merci chère professeur de m'avoir inspiré à donner le meilleur de moi-même. Merci de m'avoir appris que l'émotion et la logique existent mieux ensemble et que la compétence et la connaissance vient après la diligence et la patience. Merci pour votre confiance, votre écoute et vos précieux conseils. Je me sens si chanceuse de vous avoir sur mon chemin de la vie tant sur le plan scientifique que sur le plan humain et pédagogique.

Mes profonds remerciements s'adressent également à **Mr. Idrissi Abdellah**, Professeur de l'Enseignement Supérieur à la Faculté des Sciences de Rabat - Université Mohammed V, pour son attention, son soutien, son esprit de collaboration, son expertise et ses idées pertinentes qui ont contribué dans l'accomplissement de ce travail.

Un grand Merci s'adresse également à **Mme. Soumia ZITI**, Professeur de l'Enseignement Supérieur à la Faculté des Sciences de Rabat - Université Mohammed V, d'avoir accepté d'examiner et de rapporter ce présent mémoire. Merci pour votre disponibilité, vos encouragements, vos conseils et votre énergie professionnelle que m'a grandement motivée.

Ma gratitude et mes respects s'adressent également à **Mme. Faouzia BENABBOU**, Professeur d'Enseignement Supérieur à l'Université Hassan II-Mohammedia, Ben M'Sik de Casablanca, d'avoir accepté de présider le Jury de la soutenance de cette thèse.

Toute ma reconnaissance à **Mr. Abderrahim TRAGHA,** Professeur d'Enseignement Supérieur à l'Université Hassan II-Mohammedia, Ben M'Sik de Casablanca, d'avoir accepté d'examiner et de rapporter le présent mémoire ainsi de juger les travaux réalisés au cours de cette thèse.

Je tiens à exprimer mes sincères remerciements à **Mr. Mouad BENMAMOUN,** Professeur de l'Enseignement Supérieur à la Faculté des Sciences de Rabat - Université Mohammed V, d'avoir accepté d'être examinateur et rapporteur de ce mémoire.

Mes remerciements s'adressent également à **Mr. Ahmed EL YAHYAOUI,** Professeur Assistant à la Faculté des Sciences de Rabat - Université Mohammed V, d'avoir accepté de participer dans l'évaluation de mes travaux et d'être parmi les membres de jury de cette soutenance.

Merci à tous les membres du laboratoire IPSS (professeurs et étudiants) ainsi que tout le corps administratif de la Faculté des Sciences de Rabat pour leurs conseils, encouragements et pour la sympathie qu'ils m'ont témoignée.

Aucun mot ne peut exprimer ma gratitude que j'éprouve pour ma famille qui a été d'un soutien inestimable durant toute cette aventure. Infiniment merci. Sans vous, cette aventure ne serait pas possible.

Résumé

Avec l'évolution immense du monde numérique et l'explosion de la quantité de données qui l'en découle, tirer profit des données de manière éthique et sécurisée est actuellement une préoccupation de premier ordre afin de pouvoir accompagner cette révolution technologique que les industrielles considèrent la 4ème de son genre. Cette thèse s'articule autour de cette problématique. L'un des grands défis de ce siècle : la confidentialité et le respect de la vie privée. Malgré son importance, ce domaine évolue très difficilement et manque de nouvelles stratégies plus récentes et adaptées aux exigences technologiques actuelles. Dans cette optique, notre première contribution tourne autour d'une nouvelle stratégie de chiffrement de données dite intelligente. Cette stratégie vise à utiliser un système à base de connaissance afin de décider le système de chiffrement le plus adéquat aux échanges réalisés. Pour se faire, Il se base sur une base de connaissances de cas de chiffrements très variés et un moteur de référence utilisant l'algorithme BNL de la recherche Skyline et retourne la méthode de chiffrement qui répond mieux aux contraintes exigées par la communication en cours. Les résultats de ce travail étaient très concluants par rapport aux décisions de chiffrements proposées. A travers la deuxième contribution de cette thèse, nous avons proposé d'améliorer les performances de ce système encore plus en y intégrant une couche supplémentaire permettant de brancher le moteur d'inférence avec l'algorithme Skyline le plus adéquat aux contraintes réellement exploitées. Ceci dit, au lieu de se contenter de l'algorithme BNL, nous avons étudié les performances des différents algorithmes de cette famille et nous avons pu ressortir avec paramétrage de notre application en fonction du nombre et du choix des dimensions dans notre contexte. Finalement, dans la dernière contribution, nous avons abordé la problématique de la confidentialité dans un autre champ d'application qui est la Blockchain. A travers ce travail, nous avons pu trouver un compromis entre les deux exigences Confidentialité & Transparence au sein d'un même protocole de validation. Nommé « Protocol for Partial Confidentiality & Transparency » PPCT, ce nouveau protocole consiste à appliquer une confidentialité partielle sur les transactions échangées tout en assurant le processus de leurs validations par les différents nœuds de la Blockchain. Nous avons montré la résistance et l'efficacité de ce protocole pour répondre largement aux objectifs fixés tout au long de cette thèse qui est la confidentialité et le respect de la vie privée pour les technologies disruptives telles que l'IA et la Blockchain.

Mots-clefs : Sécurité, Cryptographie, Confidentialité, Chiffrement Intelligent, Système à base de connaissances, Blockchain.

Abstract

With the immense evolution of the digital world and the explosion in the amount of data that results from it, taking advantage of data in an ethical and secure manner is currently a prime concern in order to be able to support this technological revolution that manufacturers consider as the 4th of its kind. This thesis revolves around this issue. One of the great challenges of this century: confidentiality and respect for private life. Despite its importance, this field is very difficult to evolve and lacks new, more recent strategies adapted to current technological requirements. With this in mind, our first contribution revolves around a new so-called intelligent data encryption strategy. This strategy aims to use a knowledge-based system in order to decide on the most suitable encryption system for the exchanges carried out. To do so, it is based on a knowledge base of very varied encryption cases and a reference engine using the Skyline BNL search algorithm and returns the algorithm that better meets the constraints required by the communication in progress. The results of this work were very conclusive in relation to the proposed encryption decisions. However, during the experiments, we found that the performance of the inference engine decreases in the case of several dimensions. This is why, during the second contribution of this thesis, we proposed to improve the performance of this system even more by integrating an additional layer allowing to connect the inference engine with the Skyline algorithm most suitable for constraints actually exploited during each communication. Finally, in the last contribution, we addressed the issue of confidentiality in another field of application, which is the Blockchain. Through this work, we were able to find a compromise between the two Confidentiality & Transparency requirements within a single validation protocol. Named "Protocol for Partial Confidentiality & Transparency" PPCT, this new protocol consists of applying partial confidentiality to the transactions exchanged while ensuring the process of their validation by the various nodes of the Blockchain. We have shown the resistance and effectiveness of this protocol to broadly meet the objectives defined throughout this thesis which is confidentiality and privacy for disruptive technologies such as AI and Blockchain

Keywords: Security, Cryptography, Confidentiality, Smart Encryption, Knowledge Based System, Blockchain.

Table Des Matières

Liste de Figures

Liste des Tables

I- Motivation

Depuis toujours, les codes secrets ont contribué à surmonter les revers et les coups du sort des peuples, des communautés et des nations. C'est grâce à eux que les militaires ont pu piloter leurs guerres dans l'ombre, les malfaiteurs et les bandits protéger leurs trésors, les amateurs camoufler leur passion. Ainsi, les Grecs ont été sauvés des Perses, César a pu poursuivre ses conquêtes, Mary Stuart a été arrêtée et décapitée, un puzzle du Masque de fer a été résolu, Wilson a rejoint les alliés et des milliers de vies ont été sauvées pendant la deuxième Guerre mondiale.

L'histoire met en scène des scientifiques non connus par obligation du secret, des linguistes, physiciens, mathématiciens, joueurs d'échecs, et même les amateurs de mots croisés. Une lutte constante passant de l'Antiquité au XXI^{ème} siècle, de ceux qui inventent les codes pour garder des secrets à ceux qui cherchent à les casser pour apprendre ces secrets. Des confrontations dures ont été menées entre les code-makers et les code-breakers.

Un merveilleux portrait de l'esprit humain se dévoile au carrefour du jeu, de la guerre, de la science, de l'espionnage et finalement mis à jour avec une nouvelle édition sur la cryptographie et la cryptanalyse informatique. Pour former ce qu'on appelle la Cryptologie. S'agit-il d'un art ! Plutôt, une science ! C'est l'histoire secrète de la civilisation. Une histoire qui a vu son printemps fleurit à la fin de XIX^{ème} siècle à l'aube de la cryptographie militaire. Formalisé par Auguste Kerckhoffs en 1883, dans son article intitulé « la cryptographie militaire » [1]. Le principe de Kerckhoffs énonce les règles stratégiques qu'un système cryptographique doit respecter pour assurer une communication confidentielle. Des principes qui regroupent l'aspect technique et l'aspect humain pour le fondement des premiers cryptosystèmes au sens large du terme[2].

Autrefois monopole du gouvernement, la cryptologie touche aujourd'hui tout le monde. En effet, grâce aux « Principes de Kerckhoffs » puis aux « Théories de l'information » fondées par Claude Shannon[3], la cryptographie moderne s'est instaurée et continue à se développer jusqu'à nos jours. Elle sécurise Internet, préserve la confidentialité des e-mails, maintient

l'intégrité des transactions, sécurise le e-commerce et les paiements en ligne, brouille les signaux de télévision sur les chaînes non payantes et bien beaucoup d'autres interventions.

Or, l'histoire magnifique et inégalée des codes et des chiffrements persiste-t-elle toujours devant la nouvelle vague révolutionnaire des technologies ? Une question qui se pose constamment, mais la réponse est timide. La cryptologie commence à stagner alors que la puissance des ordinateurs se multiplie d'une manière exponentielle d'une année à une autre[4]. Les prédictions de la loi de Moore deviennent une réalité et les composants microélectroniques sont de plus en plus fins et atteints déjà leurs limites en arrivant à la taille de l'atome 5nm en 2020, le défis de passer à 3 nm est toujours levé[5]. Heureusement, les géants de l'électronique comme IBM, Toshiba, Sony, Intel, Samsung, Apple, commencent à déclarer forfait dans leurs courses derrière la miniaturisation [5]. Bonne nouvelle, certes ! Cependant, à cette puissance déjà, le risque de casser certains algorithmes cryptographiques reste très probable.

En outre, mise à part l'évolution de la puissance des ordinateurs actuels, d'autres menaces de sécurité sont sujettes d'actualité pour le moment, il s'agit de la mise en place des ordinateurs Quantiques. Un chantier en pleine construction dont l'avenir est aussi prometteur en termes de vitesse de traitement qu'en termes de risque sur la sécurité des systèmes actuels[6]. Il faut d'ores et déjà penser à renforcer la sécurité de l'existant avant même de passer à d'autres solutions pour un domaine toujours en ambiguïté[7].

Nous optons dans cette thèse pour des solutions à remodeler pour englober les nouvelles exigences. Une nouvelle voix à franchir tantôt vers l'intelligence artificielle et la protection de Big Data, ces quantités astronomiques de données émises quotidiennement par les capteurs et les objets connectés, stockées ou gérées par le Cloud. Et tantôt vers la Blockchain et les réseaux distribués. Un nouvel esprit de libération basé sur le partage, la confiance collective et la transparence[8].

Si les bénéfices de ces technologies sont réels et à portée de main, cette transformation pose également des challenges concrets aux entreprises, développeurs, gouvernements et employeurs. Une partie de la force de travail devra être requalifiée. Des questions éthiques et légales doivent également être adressées et régulées afin de favoriser leur adoption dans les meilleures conditions.

II- Démarche et Objectifs

Le présent mémoire permet de rassembler les différents travaux réalisés dans le cadre de cette thèse sous le titre : « Nouvelles Contributions Dans Le Contexte de La Sécurité Informatique Dédiées Aux Technologies Disruptives ». Un axe de recherche plein de défis qui exige la maitrise d'un ensemble très important de techniques et de concepts dans différents domaines. D'une part, la sécurité et la complexité des outils cryptographiques et d'autre part, ces technologies de veille, leurs valeurs ajoutées et leurs pistes d'amélioration.

Afin de pouvoir s'y mettre, il fallait répondre à plusieurs questions. Tout d'abord, des questions dans le domaine de la sécurité informatique : sur quoi elle se base ? Comment elle est assurée ? Comment elle évolue ? Et quelles sont ses limites actuellement ? D'autres part, des questions concernant les technologies disruptives : quelles sont ces technologies ? Et pourquoi elles le sont ? Comment la sécurité est-elle abordée et traitée dans ces domaines ? Puis, quels sont les enjeux et les défis qui restent à franchir ?

Toutes ces questions et bien d'autres étaient notre premier challenge durant cette thèse. Cependant, le deuxième challenge était de partir d'un sujet aussi vague que le nôtre, à des objectifs plus spécifiques, pour en sortir des travaux accentués sur l'une des problématiques la plus exigée dans cette dernière décennie.

C'est dans cette optique que notre thèse s'inscrit. Il s'agit d'un focus sur l'un des grands défis de ce siècle : la confidentialité et le respect de la vie privée pour une meilleure exploitation des technologies dites disruptives. Ceci dit, creuser sur deux pistes de recherche différentes, l'une se situe dans la voix d'utiliser ces technologies pour le profit de la confidentialité et l'autre cherche à renforcer cette dernière pour le profit des technologies. Sur la première voix de recherche, nous nous sommes concentrés sur l'intelligence artificielle et nous l'avons utilisée comme outil de base pour augmenter le niveau de sécurité des échanges. Tandis que sur la deuxième voix, nous nous sommes intéressés plus à la technologie Blockchain afin de proposer un nouveau protocole permettant de sécuriser les échanges au sein de cette nouvelle architecture de communication.

Partons du général au plus spécifique, nous avons pu cerner notre périmètre de recherche entre : la Confidentialité en tant qu'une problématique d'ordre prioritaire dans le domaine de la « Sécurité Informatique ». L'Intelligence Artificielle en tant qu'une technologie disruptive qui a contribué à la mutation de l'industrie et du paradigme business étant donnée sa richesse en méthodologies intelligentes nourris par la puissance de données et des connaissances

potentielles. Finalement, la Blockchain dans laquelle la confidentialité fait partie des défis qui entrave largement sa mise en place.

A ce stade, nos objectifs deviennent plus clairs, il s'agit de :

1. Proposer un nouveau système de chiffrement basé sur l'intelligence artificielle
2. Proposer une solution pour renforcer la confidentialité dans la technologie Blockchain

III- Contributions

Après avoir décortiqué notre besoin initial en des objectifs bien fins, nos premiers pas dans le domaine de la recherche commencent à donner ses fruits par la concrétisation de notre premier travail regroupant la confidentialité et l'Intelligence artificielle. Suivi d'une autre publication dans le même axe de recherche et qui représente une autre variante de la première. Finalement, une dernière contribution pour assurer la confidentialité des transactions au sein d'une infrastructure Blockchain.

2.1 Première contribution

Notre première contribution touche le domaine de l'intelligence artificielle. Un domaine en perpétuelle évolution et qui englobe une variété de concepts et de techniques en vue de réaliser des machines capables d'imiter l'intelligence humaine. A traves cette contribution, nous cherchons à bénéficier de ces paradigmes afin de concevoir un nouveau modèle de sécurité, en mariant data, IA et expertise humaine. Il s'agit d'une nouvelle stratégie décisionnelle pour un chiffrement intelligent. Fondée sur le principe des Systèmes à Base de Connaissances, cette stratégie repose sur trois briques techniques[9] fondamentales à savoir :

- Une base de connaissances : permet de stocker les connaissances sur deux parties, la première représente une condition de déclenchement de l'action qu'on appelle Prémisses et la 2ème partie : représente l'effet du lancement de cette action et qu'on appelle Conclusion.
- Une base de Faits : contenant l'ensemble des faits qui décrivent les différentes situations possibles (vrai) pour une problématique précise.

- Le moteur d'inférence : exploite la base de règles et fait enchainé les instructions afin de conclure de nouveaux faits pour répondre à un but prédéfini.

Pour faire fonctionner cette stratégie, nous partons sur 4 étapes, la première est celle de l'acquisition des connaissances dans laquelle on réalise une analyse multidimensionnelle des messages en entrée, on effectue une classification de ces entrées, puis on les transforme en des connaissances sous forme de dimensions (critères) et de conclusions/indicateurs. La deuxième étape est une étape de Représentation et d'alimentation de ces connaissances. Elle consiste à alimenter une base de connaissance en s'appuyant sur une modélisation décisionnelle en mode étoile. L'alimentation de la base de connaissance s'effectue sur deux phases différentes : la première phase consiste à faire trainer le système en réalisant plusieurs expériences afin d'engendrer la plupart des cas possibles. Tandis que la deuxième phase s'effectue à la fin du processus décisionnel sur la base de l'étude en cours. Il s'agit d'une étape d'enrichissement permettant de remplir davantage cette base avec les nouveaux cas réellement produits.

La troisième étape de ce processus est une étape de Traitement des Connaissances. Elle consiste à exploiter les connaissances précédemment stockées et celles issues de l'analyse multidimensionnelle de la nouvelle entrée. Une analyse qui distingue entre les critères fixes de cette entrée (taille, type, langue, source, protocole réseau, etc.) et des critères à optimiser (entropie, temps de chiffrement, etc.). Cette étape fait tourner le moteur d'inférence de ce SBC et extrait l'expérience la plus adéquate qui correspond au mieux à notre cas. Ceci est réalisé par le biais de l'algorithme BNL qui permet de remonter l'ensemble des points Skyline les plus dominants par rapport aux critères spécifiés au préalable.

Une fois avoir la liste des lignes Skyline, nous passons à la dernière étape de ce système qui consiste à conclure l'algorithme de chiffrement le plus robuste par rapport à ce type d'entrée et procède à l'application de ce chiffrement puis à l'enrichissement dans la base de connaissance à la fin de cette expérience. Le prototype réalisé pour ce système retourne de très bons résultats qui coïncident parfaitement avec les décisions préalablement prisent par les experts dans le domaine.

2.2 Deuxième contribution :

La deuxième contribution représente une continuité du premier travail dans lequel nous proposons d'ajouter une autre brique à notre système de chiffrement intelligent. Cette brique met le focus sur les algorithmes Skyline et ajoute un paramétrage préalable permettant de modifier à chaque fois cet algorithme en fonction des critères de sélection de chaque entrée. En effet, nous avons remarqué au cours de nos expériences que malgré la simplicité et la performance de l'algorithme BNL, des lignes Skyline non attendues ont été glissées parmi les résultats remontés par cet algorithme. Nous avons procédé à une étude expérimentale des différents algorithmes Skyline. Les résultats de cette étude étaient très bénéfiques et nous ont permis de définir le paramétrage à considérer pour chaque cas de figure afin d'avoir les meilleurs résultats. Une nouvelle couche de configuration a été ajoutée à notre système prenant en compte le paramétrage qu'on vient de définir au cours de ce travail. Les résultats de ce travail étaient très concluants peu importe le nombre de dimensions, le type des valeurs, et leurs tailles.

2.3 Troisième contribution :

La troisième contribution vient pour répondre au deuxième objectif définit au début de notre thèse. Il consiste à viser la problématique de la confidentialité au sein de la technologie Blockchain.

En effet, étant donnée la carence constatée dans cette technologie concernant l'aspect de la confidentialité, nous avons décidé de franchir ce volet en espérant pouvoir arriver à une solution assurant cette confidentialité tout en préservant les autres propriétés exigées par la Blockchain. Un souhait qui a été concrétisé à la fin de cette recherche par notre travail intitulé : Nouveau Protocole pour une Partielle Confidentialité et Transparence au sein d'une infrastructure Blockchain (PPCT). Ce dernier introduit une nouvelle notion dans le domaine de la sécurité qui est la « confidentialité partielle ». Il permet, à travers un ensemble de concepts, de protéger les données sensibles d'une transaction tout en assurant la validation de cette transaction entre deux groupes de validateurs, un publiaue et un autre privé qui change en fonction de son rôle dans chaque transaction. Le protocole passe par plusieurs étapes qu'on peut récapituler dans les points suivants:

- ✓ L'initialisation de la transaction,
- ✓ Le chiffrement partiel de la transaction,
- ✓ La diffusion de la transaction,

- ✓ La première validation par le Groupe de Confiance de cette Transaction (TTG) qui sont en mesure de déchiffrer et valider l'ensemble du contenu en clair puis diffuser leurs décisions signées et chiffrées par une clé qu'on nomme clé de validation (Kv) générée à partir des données sensibles,
- ✓ La deuxième validation par l'ensemble du réseau permettant de vérifier que les validations du TTG ont été bien effectuées sur un texte clair via la Clé Kv et finalement appliquer le reste du processus de validation pour accepter ou refuser cette transaction en fonction du smart contrat.
- ✓ La construction du bloc et la mise à jour des registres.

IV- Organisation du document

Pour expliquer davantage les travaux réalisés dans le cadre de cette thèse, le présent document suivra le plan suivant :

Après avoir introduire le contexte général de cette thèse. Le **Chapitre 1** sera consacré à une vue détaillée sur le domaine de la cryptologie afin de parcourir les différents aspects et l'ensemble des techniques concernant ce domaine et qui sont en liaison avec la problématique de la Confidentialité des données qu'on essaie de couvrir dans cette thèse.

Le reste de ce mémoire sera organisé sur deux parties, une première partie qui aborde notre premier axe de recherche concernant l'Intelligence Artificielle ainsi que les contributions réalisées dans ce contexte. Tandis que la deuxième partie sera consacrée à l'autre volet de recherche de cette thèse concernant la technologie Blockchain et la contribution dans ce domaine.

La **Partie I** commence par un premier chapitre (**Chapitre 2**) sur le domaine de l'Intelligence Artificielle depuis sa naissance jusqu'à sa concrétisation dans différents domaines d'application. Ensuite, il décrit les différentes approches de cette technologie tout en détaillant les concepts des Systèmes à Base de Connaissances et les autres techniques les plus utilisées aussi au sein de cette science concernant les Machines Learning.

Dans cette même partie, un deuxième chapitre (**Chapitre 3**) a été élaboré afin d'aborder en détaille notre première contribution dans le domaine de l'intelligence artificielle intitulée « Nouvelle stratégie décisionnelle de chiffrement basée sur l'intelligence artificielle ». Dans ce

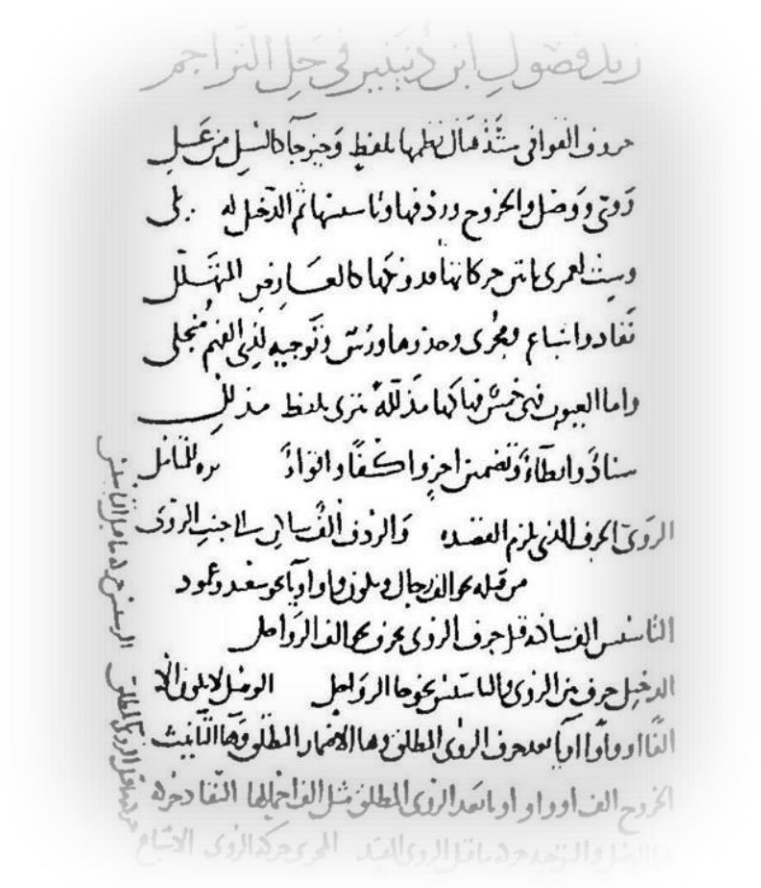
chapitre, nous présentons la motivation derrière ce travail, les travaux en relation avec notre proposition, les différentes étapes de cette nouvelle stratégie et les briques techniques ainsi que les algorithmes développés afin de mettre en place ce concept de chiffrement intelligent. Finalement, ce chapitre présentera les différentes expérimentations réalisées et les résultats obtenues et prouvons l'efficacité de notre système.

Le dernier chapitre de cette partie (**Chapitre 4**) est dédié à la 2ème contribution intitulée : « Nouvelle approche de chiffrement intelligent basée sur une étude des outils d'analyse multidimensionnelle ». Ce chapitre présente la problématique et la motivation pour ce travail, puis explique les différents algorithmes utilisés dans cette contribution concernant la recherche les points Skyline. Finalement, il démontre par une étude expérimentale les facteurs principaux impactant les performances du système et par conséquent, feront objet à une nouvelle brique du paramétrage au sein de notre stratégie de chiffrement intelligent.

Le présent mémoire traite dans sa troisième partie l'autre axe de recherche de cette thèse concernant la technologie Blockchain. Le premier chapitre de cette partie (**Chapitre 5**) est dédié à cette nouvelle technologie. D'un côté, Il donne un aperçu sur les concepts fondamentaux à propos de la Blockchain, son fonctionnement, ses types et le concept du consensus. Et d'un autre côté il parcourt le volet de la sécurité au sein de cette technologie.

Vient après, Le dernier chapitre (**Chapitre 6**) qui met le focus sur notre dernière contribution dans le domaine de la Blockchain intitulée : « Nouveau Protocole pour une Partielle Confidentialité & Transparence (PPCT) dédié à la Blockchain ». Ce chapitre discute la problématique qu'on cherche à résoudre, notre motivation, les défis et les enjeux en termes de confidentialité, les travaux antérieurs et finalement, il présente le nouveau protocole PPCT : ses nouveaux concepts, ses étapes, son application dans le domaine de la santé, puis une discussion concernant les apports et les points d'amélioration prévus dans le futur.

Ce mémoire est achevé par une **Conclusion générale** sur ce qui a été réalisé tout au long de cette thèse et les perspectives qu'on souhaite toujours aborder dans le futur.



Ibn Dunaynir (1187-1229)

La première trace d'un procédé de chiffrement faisant explicitement appel à un calcul

Chapitre 1 |

Cryptologie

Sommaire

I- Introduction

La cryptologie est l'une des plus anciennes disciplines. Des traces de son utilisation remontent à 2.000 avant J.C [10]. Cette science appelée aussi l'art du secret, continue toujours à se développer et se moderniser afin d'accompagner les changements culturels, économiques, sociales et aussi technologiques. Elle s'appuie sur un ensemble de méthodes et techniques astucieuses permettant à deux entités de communiquer en toute sécurité même à travers un canal peu sûr. En effet, un ennemi peut être à l'écoute du canal de communication, mais il n'est pas capable d'intercepter ni de comprendre ce qui a été échangé[2].
Cette problématique peut être schématisée comme suite :

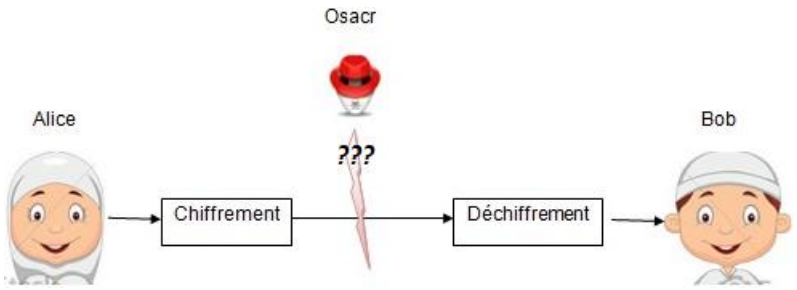


Figure 1 : Principe de communication secrète

Pour garantir son rôle, la cryptologie est une discipline qui fait développer deux domaines opposés mais qui se complètent. On parle ici de la cryptographie et la cryptanalyse. Ainsi la cryptographie cherche à mettre en jeu les techniques nécessaires pour assurer la sécurité des échanges tandis que la cryptanalyse part dans le sens contraire et cherche à analyser ces méthodes cryptographiques, détecter leurs faiblesses et démontrer leur niveau de sécurité et leur capacité de résistance face aux différentes tentatives d'attaques et d'interception[2].

Protéger une information ne revient pas seulement à cacher son contenu et garder sa confidentialité mais aussi assurer la provenance de cette information et garantir qu'elle n'était pas falsifiée lors de sa transmission, d'où proviennent les principaux piliers de la cryptographie regroupés dans l'acronyme CID et qui sont clarifiés comme suite[2] [10] :

Confidentialité : ensemble de méthodes permettant de garantir l'inintelligibilité des échanges d'informations à toute entité externe à cet échange. Cet objectif est assuré par le biais des techniques de chiffrement permettant de cacher l'information d'une façon à ce que seules les personnes autorisées peuvent la déchiffrer et puis la lire.

Intégrité : ce principe consiste à assurer que les données reçues n'ont pas été altérées au cours de leur transfert. Ce principe est assuré par le biais des techniques de signature. Une fois le document est signé ne peut plus être modifié car la signature est créée en fonction du contenu à échanger.

Disponibilité : Malgré que toutes les données et les informations sont bien sécurisées, ils doivent cependant rester accessible à tout moment afin de maintenir le bon fonctionnement des services les utilisant.

D'autres piliers ont été identifiés au fur à mesure que le domaine gagnait en maturité [11] tels que :

Authentification : c'est le fait de protéger l'accès aux données ou aux ressources en question. Seules les personnes dûment autorisées peuvent y accéder. Ce principe est assuré à travers une procédure d'identification unique basée sur les fonctions de hachage, les PKI, Biométrie, etc.

Non répudiation : il s'agit d'un besoin aussi bien intéressant que les autres objectifs de ce domaine. En effet, à travers ce concept les entités d'une communication ne peuvent pas nier qu'ils sont bien à l'origine des données envoyées.

II- Définitions

2.4 Cryptosystème

Un système cryptographique est un quintuple (P, C, K, E, D) [4] telle que :

P : est l'ensemble de textes clairs possibles

C : est l'ensemble de textes chiffrés possibles

K : est l'ensemble des clés possibles

E : est l'ensemble de règles de chiffrement

D : est l'ensemble de règles de déchiffrement

Ainsi, pour tout $k \in K$, il existe :

- Une règle de chiffrement $e_k \in$ / (e_k : P → C)
- Une règle de déchiffrement $d_k \in D$ d_k : C → P correspondante à e_k
- Telles que : $d_k (e_k (x)) = x$ pour tout texte clair $x \in P$

Pour mieux assimiler cette reformulation mathématique, revenons au schéma précédent afin de faire la projection de la définition mathématique qu'on vient de donner sur ce schéma représentant les concepts réels d'un système cryptographique. La figure suivante représente cette analogie d'une manière plus simple :

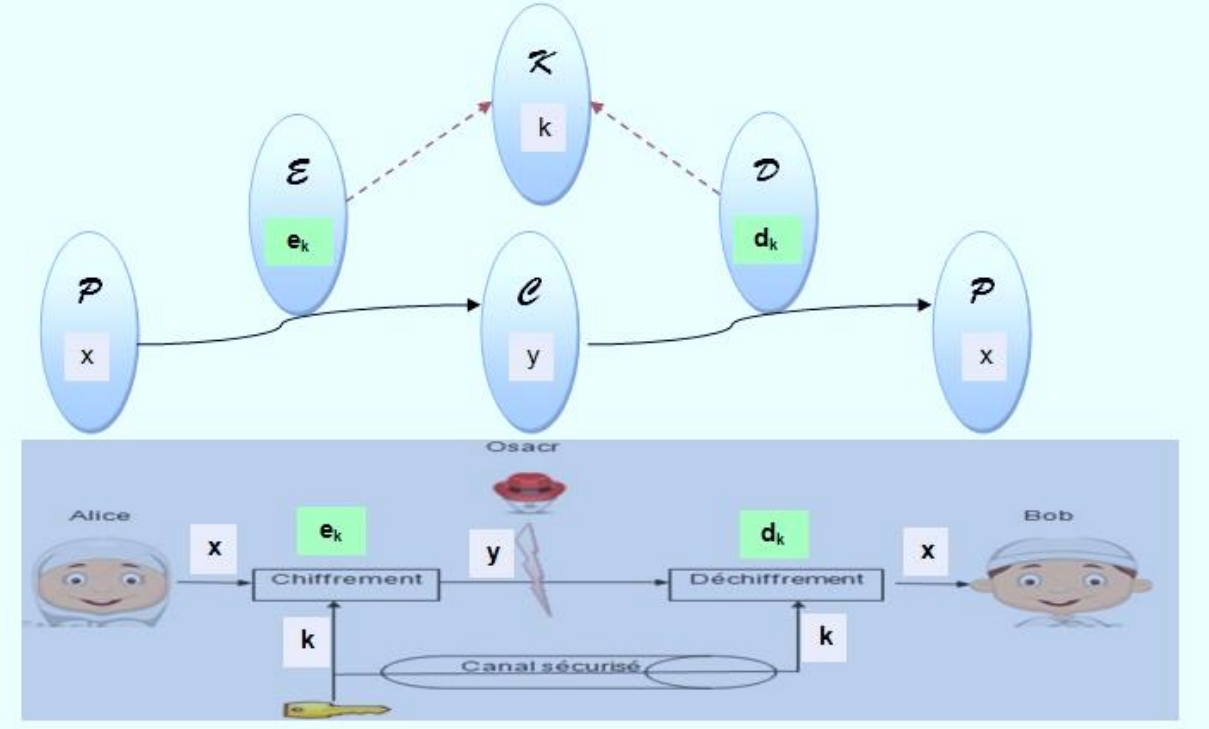


Figure 2 : Analogie mathématique d'un système cryptographique

Ce schéma montre aussi que si un texte clair x chiffré à l'aide de la fonction e_k, et si le texte chiffré y obtenu est ensuite déchiffré en utilisant la fonction d_k, alors on se retrouve avec le même message x de départ. Elle s'agit alors, de la propriété principale d'un système cryptographique.

Une autre opération qu'on utilise fréquemment dans ce domaine, c'est la notion de déchiffrement. En fait, il ne faut surtout pas confondre entre les deux mots 'déchiffrer' et 'décrypter', ils représentent deux notions différentes dans leur signification malgré qu'ils aient la même finalité[10]. Décrypter un message chiffré c'est le fait de retrouver le message clair correspondant sans avoir connaissance de la clé de déchiffrement. Cette opération entre dans le cadre de la cryptanalyse, l'autre discipline de la cryptologie et la science opposée de la cryptographie[11]. En effet, l'objectif de cette science est de casser les différentes primitives des systèmes cryptographiques les plus connus et retrouver les messages clairs où la clé utilisée en se contentant, dans la plupart des cas, sur l'analyse et l'observation du comportement de ces systèmes et les données transmises [12].

2.5 Cryptanalyse

L'objectif principal de la cryptographie est de garder le texte en clair secret des espions qui tentent d'obtenir des informations sur le texte en clair. Tel que discuté auparavant, des adversaires peuvent également être actifs et essayer de modifier le message. Les adversaires sont supposés avoir un accès complet au canal de communication.

La cryptanalyse est la science qui étudie les attaques contre la cryptographie. A travers ces attaques réussies, ils peuvent, par exemple, récupérer le texte en clair (ou des parties du texte en clair) à partir du texte chiffré, substituer des parties du message d'origine ou falsifier des signatures numériques [11].

La cryptographie et la cryptanalyse sont souvent subsumés par le terme plus général de cryptologie.

Une hypothèse fondamentale en cryptanalyse a été énoncée pour la première fois par A. Kerckhoffs au XIXe siècle. Il est généralement appelé Principe de Kerckhoffs. Ce principe indique que l'adversaire connaît tous les détails du cryptosystème, y compris ses algorithmes et leurs implémentations[1]. Selon ce principe, la sécurité d'un cryptosystème doit reposer entièrement sur le secret des clés. Les attaques contre le secret d'un schéma de chiffrement tentent de récupérer des textes en clair à partir de textes chiffrés ou, encore plus drastiquement, pour récupérer la clé secrète. L'enquête qui suit est limitée aux attaques passives. L'adversaire – comme d'habitude, nous l'appelons Eve - n'essaye pas de modifier les messages. Elle surveille le canal de communication et les extrémités du canal. Alors non seulement, il peut intercepter le texte chiffré, mais aussi (au moins de temps en temps) elle peut être en mesure d'observer le chiffrement et le déchiffrement des messages. Il n'a aucune information à propos de la clé. Par exemple, Eve peut être un opérateur d'un ordinateur bancaire. Il voit les textes chiffrés entrants et parfois aussi les textes clairs correspondants. Ou il observe les textes clairs sortants et les textes chiffrés générés[12].

Peut-être qu'il parvient à laisser chiffrer des textes en clair ou à déchiffrer des textes chiffrés de son propre choix.

Une tentative de cryptanalyse d'un système est appelée une attaque, et elle peut conduire à différents résultats [11] :

- ✓ Cassage complet : la cryptanalyse retrouve la clef de déchiffrement.

- ✓ Obtention globale : la cryptanalyse trouve un algorithme équivalent à l'algorithme de déchiffrement, mais qui ne nécessite pas la connaissance de la clef de déchiffrement.
- ✓ Obtention locale : la cryptanalyse retrouve le message en clair correspondant à un message chiffrer.
- ✓ Obtention d'information : la cryptanalyse obtient quelque indication sur le message en clair ou la clef (certains bits de la clef, un renseignement sur la forme du message en clair). D'une manière générale, on suppose toujours que le cryptanalyste connait le détail des algorithmes, fonctions mathématiques ou protocoles employés. Même si ce n'est pas toujours le cas en pratique, il serait risqué de se baser sur le secret des mécanismes utilisés pour assurer la sécurité d'un système, d'autant plus que l'usage grandissant de l'informatique rend de plus en plus facile la reconstitution de l'algorithme à partir du programme [10].

Les attaques possibles dépendent des ressources réelles de l'adversaire Eve. Ils sont généralement classés comme suit [11]:

1. Attaque par texte chiffré uniquement : Eve a la capacité d'obtenir des textes chiffrés. Cet est susceptible d'être le cas dans n'importe quelle situation de chiffrement. Même si Eve ne peut pas effectuer les attaques plus sophistiquées décrites ci-dessous, il faut supposer qu'elle peut accéder aux messages chiffrés. Une méthode de chiffrement qui ne peut pas résister à une attaque par texte chiffré est totalement non sécurisée.

2. Attaque en texte clair connu : Eve a la capacité d'obtenir des pairs textes clair-texte chiffré. En utilisant les informations de ces paires, elle tente de déchiffrer un texte chiffré dont elle n'a pas le texte en clair. À première vue, il pourrait sembler que de telles informations ne seraient normalement pas disponibles pour un attaquant. Cependant, il est très souvent disponible. Les messages peuvent être envoyés dans des formats standards connus d'Eve.

3. Attaque en texte clair choisi : Eve a la capacité d'obtenir des textes chiffrés pour textes en clair de son choix. Puis elle tente de déchiffrer un texte chiffré dont elle n'a pas le texte en clair. Tandis qu'à nouveau cela peut sembler peu probable, il existe de nombreux cas dans lesquels Eve peut faire exactement cela. Par exemple, elle peut envoyer des informations intéressantes à sa victime, qu'elle est sûre qu'il cryptera et enverra. Ce type d'attaque suppose qu'Eve doit d'abord obtenir les pairs textes clair-texte chiffré qu'elle souhaite, puis effectuer

son analyse, sans aucune autre interaction. Cela signifie qu'elle n'a besoin d'accéder qu'une seule fois au dispositif de chiffrement.

4. Attaque en texte clair choisi de manière adaptative : C'est la même chose que le point d'amure précédent, sauf que maintenant Eve peut faire quelques analyses sur les pairs textes clair-texte chiffré, et par la suite obtenir plus de paires. Elle peut basculer entre la collecte de paires et l'exécution de l'analyse aussi souvent qu'elle le souhaite. Cela signifie qu'elle a soit un accès prolongé au dispositif de chiffrement, soit qu'elle peut en faire un usage répété.

5. Attaques par texte chiffré choisi et de manière adaptative : Ces deux attaques sont similaires aux attaques en clair ci-dessus. Eve peut choisir des textes chiffrés et obtenir les textes clairs correspondants. Elle a accès au dispositif de déchiffrement.

III- Art des codes secret : Cryptographie classique

Dans cette partie nous allons décrire quelques techniques de chiffrement utilisées à l'antiquité afin de sécuriser la communication des messages à contenu secret. En fait, on ne peut pas avancer dans le domaine de la cryptographie sans avoir un aperçu sur les algorithmes de la cryptographie classique. Ils sont les premiers outils de chiffrement utilisés et restent jusqu'à maintenant la base de la conception de la plupart des systèmes de chiffrement.

Ces algorithmes sont simples dans leur conception, cependant, leur confidentialité repose sur le secret de l'algorithme lui-même car il est facilement décryptable par des tierces intelligibles à connaitre les messages communiqués[13].

La plupart des chiffrements classiques se basent sur deux principes majeurs, à savoir : la substitution et la transposition. La substitution est le cas général de la transposition, elle consiste à remplacer une lettre du message principal par un code différent désignant une autre lettre, un chiffre où même un symbole. La transposition par contre consiste à permuter les positions entre les caractères de ce même texte[13].

Dans ce qui suit nous allons voir quelques exemples de ces types de chiffrement pour chaque catégorie.

3.1 Chiffrement par Transposition

Le chiffrement par transposition est tout simplement une permutation entre les positions des caractères du message clair. L'utilisation de ce type de méthodes remonte à 487 avant J-C quand les grecques inventaient un dispositif appelé la « Scytale » enroulée d'une bandelette de cuire sur laquelle inscrivait le message chiffré. La Scytale est ensuite envoyé par le messager au destinataire[14].

Pour déchiffrer le message, le destinataire devait avoir un bâton identique à celui de l'encodage. Il lui permettait alors d'enrouler le cuire autour de ce bâton pour reconstruire le message clair[15].

3.2 Chiffrement par Substitution

3.2.1 Chiffrement par substitution mono-alphabétique

Postérieur et simple, les méthodes de la substitution mono-alphabétique représentent l'idée de base de plusieurs systèmes de chiffrement récemment inventés. Cette technique repose sur un système de permutation mais cette fois ci d'une lettre du message clair à une autre lettre de l'alphabet[16].

Autrement dit :

> Soient $\mathcal{P}$ et $\mathcal{C}$ l'ensemble des 26 lettres de l'alphabet français, on définit une permutation $\pi \in \mathcal{K}$ telle que :
>
> $e_\pi(x) = \pi(x)$ et $d_\pi(y) = \pi^{-1}(y)$

Voici un exemple d'une substitution aléatoire qui peut être appliquée comme une règle de chiffrement :

Π	a	B	C	d	E	f	g	H	i	J	K	l	m	N	o	P	q	r	s	t	u	v	w	x	y	z
	O	W	H	T	R	P	D	E	N	M	J	A	V	B	F	L	S	C	G	Z	I	K	X	Q	U	Y	

Par conséquent, le chiffrement par substitution est mis en place en remplaçant chaque caractère du texte clair par le caractère correspondant dans le tableau de la permutation comme suite :

$$e_\pi(a) = \pi(a) = O, \quad e_\pi(b) = \pi(b) = W, \quad \ldots, \quad e_\pi(z) = \pi(z) = Y$$

Le nombre de clés possibles pour ce système est l'ensemble de toutes les permutations possibles des 26 caractères alphabétiques, ce qui représente un nombre supérieur à 4,0 x 1026. Essayer toutes ces clés pour décrypter le message reste impossible même en utilisant un ordinateur [16].

Toutefois, ce type de méthode est attaquable facilement en utilisant une analyse par fréquence d'apparition.

3.2.2 Chiffrement par décalage

Le chiffrement par décalage est un cas spécifique de la substitution mono-alphabétique, au lieu que la permutation de celle-ci soit définie d'une manière aléatoire, nous considérons un décalage de l'ordre normal des alphabets, à gauche où à droite, par un nombre prédéfinit de position qui représente la clé du système[17] .

Si on se réfère à l'arithmétique modulaire, ce chiffrement peut être présenté comme une équation de congruence sur z_{26}, si on considère que chaque caractère de l'alphabet correspond à un résidu modulo 26. Donc, tout d'abord nous commençons par établir cette correspondance pour tous les caractères de notre texte clair [18], ensuite nous calculons la valeur numérique du chiffré pour chaque caractère en utilisant la fonction de chiffrement suivante :

$$e_k(x) = (x + k) \bmod 26$$

Finalement, il faut convertir les valeurs numériques trouvées en caractères alphabétiques afin d'avoir le chiffré.

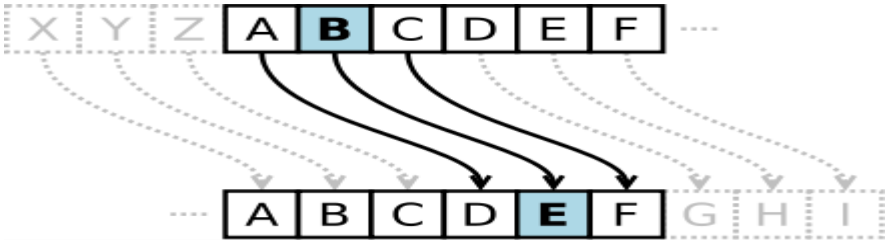
En contrepartie, le déchiffrement doit se réaliser en procédant de la même manière sauf que cette fois ci on doit considérer le texte chiffré et la règle de déchiffrement suivante :

$$d_k(y) = (y - k) \bmod 26$$

❖ **Exemple : Chiffrement de César**

Du spécifique au plus spécifique, le chiffrement de César est le chiffrement le plus ancien à utiliser les techniques de la substitution et plus précisément par décalage. Il se peut que d'autres systèmes plus anciens fussent présents à cette époque, mais au moins c'est le premier à avoir laissé ses traces dans l'histoire de la cryptographie [12].

Ce système est un chiffrement de décalage utilisant une clé égale à 3 (voir image ci-dessous) :



Cependant, Il n'y a que 26 façons différentes de chiffrer un message. Donc, ce type de chiffrement reste très fragile pour une attaque par recherche exhaustive en testant toutes les possibilités des clés.

3.2.3 Chiffrement affine

Un autre exemple plus particulier du chiffrement par décalage est le chiffrement affine. Dans ce système on limite les règles de chiffrement à une fonction affine de la forme :

$$e(x) = (ax + b) \bmod 26 \quad / \; a, b \in \mathbb{Z}_{26}$$

Ce qui nous ramène au chiffrement par décalage dans le cas où a = 1.

Cependant, cette fonction n'admet pas une solution unique, chose qui ne lui permet pas d'être une fonction de chiffrement valide car le déchiffrement peut générer plusieurs solutions.

Un théorème très connu dans le domaine de la théorie des nombres, le théorème de Bachet-Bézout, permet de résoudre cette problématique[19].

❖ **Théorème :**

Soit l'équation de congruence suivante : $y = a.x + b \; mod(m)$

L'équation admet une solution unique $x \in \mathbb{Z}_m$ pour tout $b \in \mathbb{Z}_m$ si et seulement si le $pgcd\ (a, m) = 1$.

D'où, la description complète du système de chiffrement affine sur l'ensemble z26 (en se contentant sur les 26 alphabets) :

Soit $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ *et soit* $\mathcal{K} = \{(a,b) \in \mathbb{Z}_{26}^* \mathbb{Z}_{26} : pgcd\ (a, 26) = 1\ \}$

Soit $k = (a, b) \in \mathcal{K}$ *et* $x, y \in \mathbb{Z}_{26}$, *on a :*

$$e_k(x) = (ax + b) \bmod 26$$

et

$$d_k(y) = a^{-1}(y - b) \bmod 26$$

3.2.4 Chiffrement par substitution poly-alphabétique

Début de 9ème siècle, Abu Yusuf Al-Kindi a pu révéler la technique de déchiffrement par analyse de fréquences d'apparition d'alphabet [20]. Depuis, d'autres méthodes ont apparu afin de remédier à ce problème. Des méthodes pour remplacer les alphabets par des points ou des croix et d'autres utilisant des bi-grammes pour les noms communs et les noms propres, ainsi nommées les nomenclateurs.

En 1466-1467, Leon Battista Alberti a publié le premier chiffrement poly-alphabétique utilisant un cadran chiffrant pour simplifier la réalisation de ce chiffrement[21]. Ce cadran utilise deux disques, le plus grand disque est fixe et contient l'alphabet ordonnés. L'autre disque est mobile et écrit l'alphabet dans un ordre aléatoire. Pour coder le message clair, on fait coïncider la première lettre de chaque disque puis on commence le chiffrement par une substitution simple en remplaçant le caractère du message clair, affiché sur le grand disque, par celui du petit disque. Un décalage du petit disque est ensuite effectué périodiquement afin de compliquer ce procédé. Ainsi, la substitution change d'alphabet au cours de chiffrement ce qui a donné naissance à l'idée de la substitution poly-alphabétique[16].

Figure 3 : Le cadran d'Alberti dans sa forme originale

Malgré que le procédé d'Alberti soit le premier à avoir inventé la substitution poly-alphabétique, les procédés de type nomenclateurs étaient les plus utilisés dans cette époque.

3.2.5 Chiffrement de Vigenère

En 1586, Blaise de Vigenère diplomate au service des ducs de Nevers et des rois de France, écrivain et historien à l'époque, a dévoilé son procédé de chiffrement poly-alphabétique basé sur le principe de chiffre de César avec plus de complication. En effet, ce chiffre utilise un décalage de 1 caractère pour chaque lettre en utilisant une table des suites du décalage pour les 26 alphabets[13].

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Figure 4 : le carré de Vigenère

En se servant de cette table et d'une clé sous forme d'une suite de caractère, on chiffre chaque caractère du message en claire par l'alphabet de son intersection avec le caractère de la clé.

Autrement dit :

Soit Ci , Pi et Ki la position du $i^{ème}$ caractère dans le chiffré , le texte clair et la clé.

Nous assignons aux caractères de A à Z les nombre de 0 à 25

L'équation de chiffrement par le procédé de Vignère est peut-être traduite comme suite :

$$C = E(K, P) = (P_i + K_i) \bmod 26$$

L'équation de déchiffrement est donc **:**

$$P = D(C, K) = (C_i - K_i) \bmod 26$$

Le chiffre de Vigenère a été réinventé de nombreuses fois au cours des siècles et il en existe plusieurs variantes.

La cryptanalyse du chiffre de Vigenère par la méthode de Kasiski ou par d'autres méthodes comme celles de l'indice de coïncidence demande un texte suffisamment long vis-à-vis de la clé[22]. Dans le cas extrême où la clé est de longueur égale à celle du message, et n'est utilisée qu'une seule fois, tous les textes de longueur égale à celle du message chiffré sont possibles : le chiffre ne peut être cassé ; c'est le chiffre de Vernam, ou masque jetable[23].

IV- Cryptographie Moderne

Depuis les années 1800, de nouvelles technologies de transmission des messages commencent à se développer, des moyens de transport plus rapide sont apparus comme le télégraphe, la télégraphie sans fil, etc., c'est l'essor des communications et la libération de la transmission des messages d'un messager à cheval.

Des inventions plus technique et plus rapide ont été concrétisées dont la plus connue est la machine ENIGMA. Inventée par l'Allemand Arthur Scherbius en 1918[24], la machine portative et électromécanique fut utilisée principalement par les Allemands avant et pendant la Seconde Guerre mondiale [16].

Cependant, grâce au déchiffrement des messages d'Enigma, Les informations dévoilées favorisèrent le camp des Alliés dans la guerre. En effet, les cryptanalystes dont le britannique Alan Turing ont pu perfectionner la cryptanalyse par les fameuses « bombes électromécaniques » initiées par le mathématicien polonais Marian Rejewski. Ils s'agissent des premiers ordinateurs du siècle capable de décrypter les messages de l'Enigma. L'informatique commence à se développer. L'ordinateur prend la relève et l'homme n'a plus besoin d'employer des opérations élémentaires pour réaliser ses calculs ni de tension d'esprit pour résoudre une problématique quelconque.

Une nouvelle période de l'histoire de l'humanité et surtout une nouvelle impulsion dans les notions définissant la science de la cryptographie.

Auguste Kerckhoffs a énoncé, pour la première fois, les principes fondamentaux de la sécurité des cryptosystèmes. Voici, texto, les six règles de Kerckhofs publiées en 1883 dans le Journal des sciences militaires, un article intitulé « La cryptographie militaire pour la conception du cryptosystèmes militaire »[1] :

1. Le système doit être matériellement, sinon mathématiquement, indéchiffrable ;
2. Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi ;
3. La clef doit pouvoir en être communiquée et retenue sans le secours de notes écrites, et être changée ou modifiée au gré des correspondants ;
4. Il faut qu'il soit applicable à la correspondance télégraphique ;
5. Il faut qu'il soit portatif, et que son maniement ou son fonctionnement n'exige pas le concours de plusieurs personnes ;

De ce fait, la sécurité d'un système cryptographique ne repose plus sur le secret qui entoure ce système. Des études publiques à son sujet fait que son niveau de sécurité soit très supérieur et le risque de faille est de plus en plus faible. En revanche, le secret doit dépendre d'un paramètre aisément modifiable : sa clé[18].

C'est à l'issue de cette période que le principe de la cryptographie Moderne commence à prendre sa forme. La cryptographie se détache peu à peu du domaine militaire pour envahir nos vies quotidiennes[14].

Des applications informatiques dans tous les domaines (industrie, commerce électronique, téléphonie fixe et mobile, banque, …) nous obligent à faire une confiance aveugle aux outils

cryptographiques. Ceci dit, la sécurité de ces outils doit reposer sur des techniques et des algorithmes bien plus complexes, plus sûres et surtout prouvées mathématiquement.

C'est ainsi que le concept de la Cryptographie Moderne a été conçu. Il consiste à concevoir des algorithmes publiques agissant sur des données numérisées (suite de 0 et 1) et exploitant le secret de la clé pour le chiffrement et le déchiffrement[25]. On distingue deux catégories de la cryptographie moderne :

- La cryptographie à clé secrète
- La cryptographie à clé publique

Dans ce qui suit, nous détaillons les principes et les algorithmes de chaque catégorie.

4.1 La cryptographie à clé secrète

Les algorithmes symétriques (ou à clé secrète) peuvent être considérés comme héritiers de la cryptographie classique. Ils transforment le texte clair en texte chiffré à l'aide d'une clé et le texte chiffré en texte clair en utilisant la même clé. Avec cette classe de chiffrement, la clé de chiffrement ou de déchiffrement doit être transmise de l'expéditeur au destinataire via un chemin distinct et sûr [18].

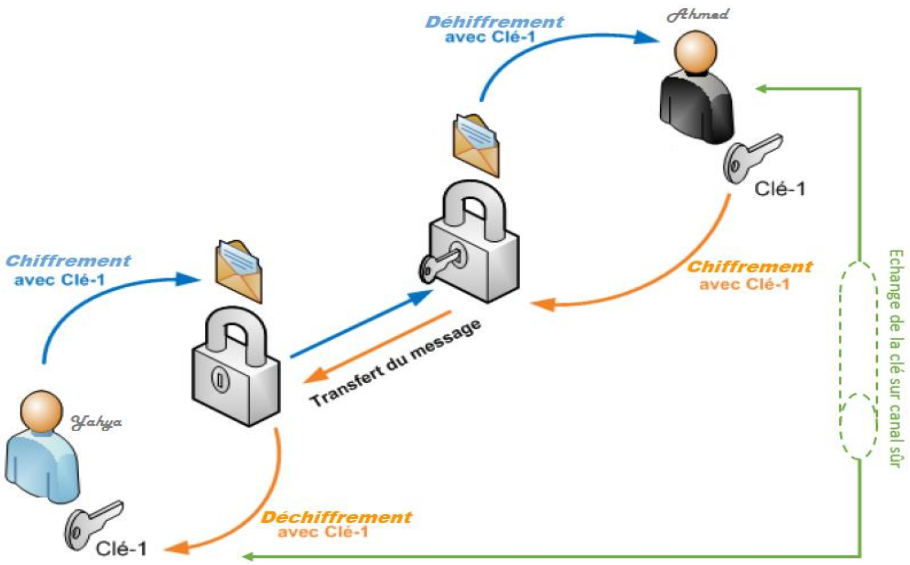


Figure 5 : Principe de chiffrement à clé secrète

Par exemple le chiffre de Vernam, inventé par l'ingénieur Américain Gilbert Vernam en 1926, consiste en un chiffrement par méthode de Vigenère en utilisant une clé générée d'une manière aléatoire et ayant la même taille que le texte clair[25]. Cette clé est unique pour chaque texte à chiffrer, d'où provient son nom du « masque jetable ». Le chiffre de Vernam a été prouvé

par le mathématicien Claude Shannon en tant qu'un chiffrement parfaitement sûr[26]. En effet, l'interception du message clair à partir du chiffré est impossible même en ayant une puissance de calcul infinie. Toutes les interceptions possibles du texte clair sont équiprobables [19].

Néanmoins, l'usage de ce chiffre reste très limité comme il s'appuie sur des conditions non faciles à mettre en place[27]. En fait :

- Se servir des clés extrêmement longues, revient à gérer leurs synchronisations. Chose qui n'est pas réalisable car si on arrive à sécuriser l'échange d'une clé du même taille que le texte clair vaut mieux échanger le texte clair du premier coup !
- De plus, la clé est unique pour chaque échange et si on envoie un message deux fois avec la même clé on risque à dévoiler les deux messages clairs :
- Soit $C_1=m_1 \oplus k$
  $C_2=m_2 \oplus k$.

  Alors $C_1 \oplus C_2=m_1 \oplus k \oplus m_2 \oplus k= m_1 \oplus m_2$

- En fin, la clé doit être aléatoire, cependant, on ne peut pas garantir qu'un processus informatique soit parfaitement aléatoire.

Pour remédier à ces problèmes, d'autres cryptosystème ont été inventés dans le but de contourner les fondements du chiffre de Vernam et les remplacer par des conditions plus réalisables tout en garantissant un bon niveau de sécurité.

On peut retenir à travers le chiffrement de Vernam que le niveau de sécurité d'un cryptosystème peut être définit comme suite :

- Plus la taille de la clé est grande plus la sécurité de la communication est élevée.
- Plus la génération de la clé est aléatoire plus la sécurité de la communication est élevée.
- Plus la clé est unique plus la sécurité de la communication est élevée.

Plusieurs systèmes de chiffrement symétrique ont été ensuite conçus, des systèmes basés sur un chiffrement par Bloc et d'autres sur un chiffrement par Flux. Dans ce qui suit, nous détaillons les algorithmes sur lesquels on va se référer lors de notre étude.

4.1.1 Chiffrement par Bloc

4.1.1.1 Schéma de Feistel

Les algorithmes de chiffrement par bloc, utilisent le réseau de Feistel, nommée d'après le cryptologue d'IBM, Horst Feistel. La première utilisation a été dans Lucifer et DES. Cette structure offre plusieurs avantages, le chiffrement et le déchiffrement ont une architecture identique dans certains cas. Un réseau de Feistel se base sur des principes simples dont des substitutions, des permutations, des échanges de blocs de données et une fonction prenant en entrée une clé intermédiaire à chaque étage [20].

Il est vraisemblable que Feistel ne soit pas le seul inventeur de cette architecture. Durant une conférence, *Don Coppersmith* a laissé entendre que Bill Notz et Lynn Smith (de l'équipe d'IBM travaillant sur DES) avaient été en grande partie à l'origine du réseau de Feistel tel que nous le connaissons [21].

Un réseau de Feistel est subdivisé en plusieurs tours ou étages. Dans sa version équilibrée, le réseau traite les données en deux parties de taille identique.

À chaque tour, les deux blocs sont échangés puis un des blocs est combiné avec une version transformée de l'autre bloc. Pour simplifier, la moitié des données sont encodées avec la clef, puis le résultat de cette opération est ajouté grâce à un xor (ou exclusif) à l'autre moitié des données. Puis au tour suivant, on inverse : c'est au tour de la dernière moitié d'être chiffrée puis d'être ajoutée avec un xor à la première moitié, sauf qu'on utilise les données chiffrées précédemment (sinon ça ne servirait à rien de faire plus de deux tours). Le schéma ci-contre montre le cheminement des données (le $\oplus$ représente le xor). Chaque tour utilise une clé intermédiaire, en général tirée de la clé principale via une génération appelée key schedule.

Les opérations effectuées pendant le chiffrement avec ces clefs intermédiaires sont spécifiques à chaque algorithme.

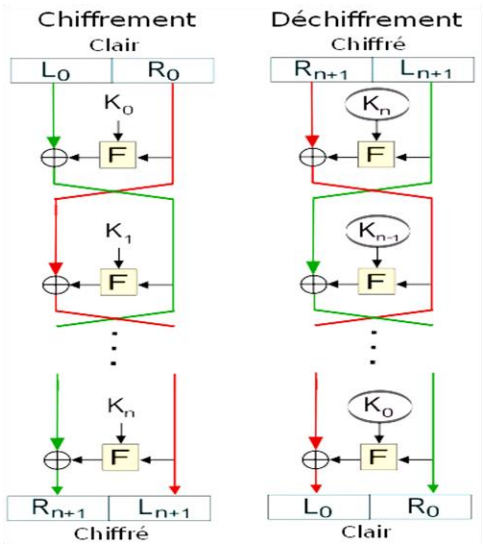


Figure 6: Schéma de Feistel

Dans le cas de DES, le réseau de Feistel possède 16 tours, chacun avec une sous-clé. Ces différentes clés permettent d'améliorer la robustesse d'un algorithme face à la cryptanalyse. Une variante, le réseau de Feistel non-équilibré coupe les données en deux parties de tailles différentes. Cette variante a été utilisée dans MacGuffin de Bruce Schneier, ou encore Skipjack, candidat pour AES [22] [23].

4.1.1.2 DES

Le Data Encryption Standard (DES) est un système de chiffrement symétrique par bloc, qui utilise des clés de 56 bits[10]. L'utilisation du DES n'est plus conseillé actuellement, du fait de sa lenteur à l'exécution et de son intervalle de clés trop petit qui peut permettre une attaque systématique en un temps raisonnable[28]. Mais il est toujours utilisé en Triple DES. DES a généralement été utilisé dans le système de mots de passe UNIX.

Le 1er standard DES est publié par FIPS le 15 janvier 1977 sous le nom FIPS PUB 46. La dernière version avant l'obsolescence date d'octobre 1999 [24].

DES fonctionne en trois étapes :

Etape 1 : Permutation initiale et fixation d'un bloc.

Etape 2 : Le résultat est soumis à 16 itérations d'une transformation, ces itérations dépendent à chaque tour d'une clé intermédiaire de 48 bits. La clé de chaque tour est calculée à partir de la clé initiale en appliquant un réseau de tables de substitution et d'opérateurs XOR. À partir de de chaque tour, le bloc de 64 bits est découpé en deux blocs de 32 bits, et ces blocs sont échangés l'un avec l'autre selon un schéma de Feistel. Le bloc de 32 bits ayant le poids le plus fort (celui qui s'étend du bit 32 au bit 64) portera une transformation.

Etape 3 : Le résultat du dernier tour est remplacé par la fonction inverse de la permutation initiale.

4.1.1.3 TripleDES

Le Triple DES appelé aussi 3DES ou TDES, est un système de chiffrement symétrique par bloc, séquencer trois applications successives de l'algorithme DES sur le même bloc de données de 64 bits [25], avec 2 ou 3 clés DES différentes.

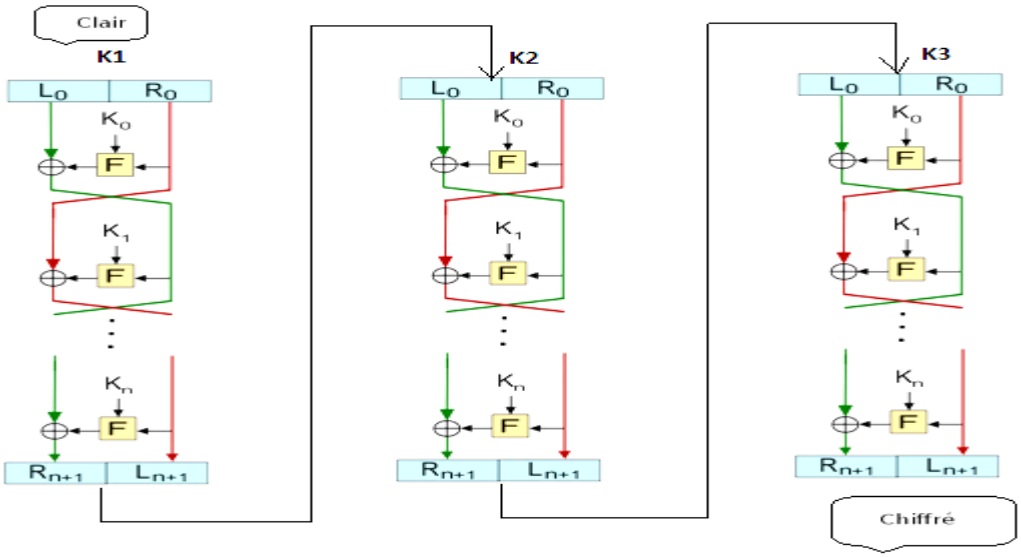


Figure 7 : Schéma Triple DES

K1 K2 K3, les clés différentes utilisés par Triple DES

La proposition d'utiliser les chiffrements DES trois fois a été développée par IBM, présenté en 1998 (sous la référence de ANSI X9.52) puis publié en 1999. Il existe en effet d'autres façons d'utiliser trois fois DES mais elles ne sont pas nécessairement sûres. Cette version utilise un chiffrement, suivi d'un déchiffrement pour se conclure à nouveau par un chiffrement.

Le Triple DES est souvent utilisé avec deux clés différentes seulement. Le mode d'usage standard est de l'utiliser en mode EDE (Encryption, Decryption, Encryption, c'est-à-dire Chiffrement, Déchiffrement, Chiffrement) ce qui le rend compatible avec DES quand on utilise trois fois la même clé[11]. Dans le cas d'une implémentation matérielle cela permet d'utiliser le même composant pour respecter le standard DES et le standard Triple DES.

Dans ce mode proposé, TDES s'écrit plus formellement de cette manière :

$$C = E_{DES}^{k3}\left(D_{DES}^{k2}\left(E_{DES}^{k1}(M)\right)\right)$$

4.1.1.4 Advanced Encryption Standard (AES)

Advanced Encryption Standard (AES) est l'algorithme de chiffrement symétrique le plus populaire et le plus largement adopté. Il est au moins six fois plus rapide que le triple DES [26].

Un remplacement pour DES était nécessaire car la taille de sa clé était trop petite. Avec une puissance de calcul croissante, il était considéré comme vulnérable aux attaques de recherche de clés exhaustives. Triple DES a été conçu pour surmonter cet inconvénient[22], mais il s'est avéré lent. Les caractéristiques d'AES sont les suivantes :

Type	Chiffrement par bloc symétrique à clé symétrique
Taille de données	Données 128 bits
Taille de la clé	clés 128/192/256 bits
Comparé à TDES	Plus fort et plus rapide que Triple-DES
Information	Fournir des spécifications complètes et des détails de conception
Code	Logiciel peut être implémenté en C et Java

Tableau 1 : Caractéristiques d'AES

Fonctionnement de l'AES

AES est un système de chiffrement itératif. Il est basé sur le « réseau de substitution-permutation ». AES comprend une suite d'opérations liées, dont certaines impliquent la substitution des entrées par des sorties spécifiques et d'autres impliquent le brassage des bits (permutations).

Fait intéressant, AES effectue tous ses calculs sur des octets plutôt que sur des bits. Par conséquent, AES traite les 128 bits d'un bloc de texte en clair comme 16 octets. Ces 16 octets sont disposés en quatre colonnes et quatre lignes pour être traités comme une matrice.

Contrairement à DES, le nombre de tours en AES est variable et dépend de la longueur de la clé. AES utilise 10 tours pour les clés 128 bits, 12 tours pour les clés 192 bits et 14 tours pour les clés 256 bits. Chacun de ces tours utilise une clé de tour de 128 bits différente, qui est calculée à partir de la clé AES d'origine.

Le schéma de la structure AES est donné dans l'illustration suivante :

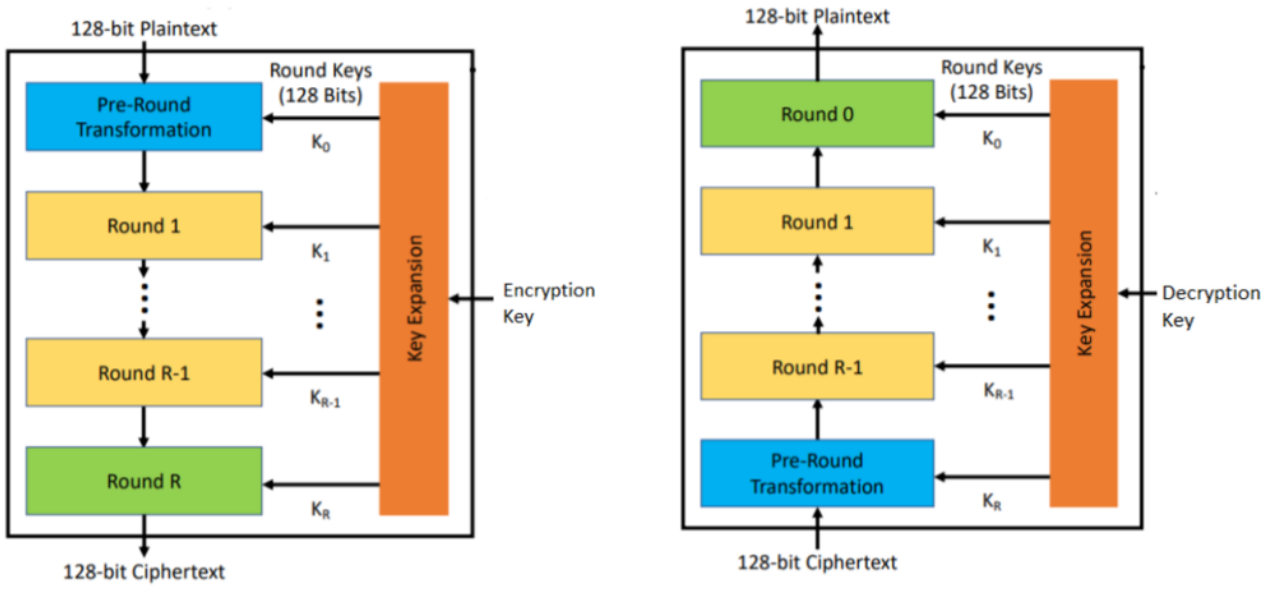


Figure 8 : Le schéma de la structure AES

Chacun des processus utilise essentiellement les mêmes opérations, outre le fait que l'un est l'inverse de l'autre. Les figures ci-dessous montrent chacune des transformations rondes pour le chiffrement et le déchiffrement respectivement. Ces transformations sont effectuées séquentiellement un bloc à la fois sur le tableau d'états. Les opérations sont :

1- Substitution des octets – Inverse de la substitution des Octets
2- Décaler les lignes- Inverse le décalage des lignes
3- Colonnes de mélange - Colonnes de mélange inversé
4- Ajouter une clé ronde

4.1.1.5 Blowfish

Blowfish est un chiffrement symétrique par bloc qui peut être utilisé en remplacement de DES ou IDEA. Il prend une clé de longueur variable, de 32 bits à 448 bits, ce qui le rend idéal pour un usage domestique et exportable. Blowfish a été conçu en 1993 par Bruce Schneier comme une alternative rapide et gratuite aux algorithmes de chiffrement existants. Depuis lors, il a été considérablement analysé et il est lentement accepté en tant qu'algorithme de chiffrement puissant. Blowfish est non breveté et sans licence, et est disponible gratuitement pour toutes les utilisations [27].

Quelques caractéristiques de ce système de chiffrement :

- Chiffrement par bloc : bloc de 64 bits
- Longueur de clé variable : 32 bits à 448 bits
- Beaucoup plus rapide que DES et IDEA
- Non breveté et libre de droits
- Aucune licence requise

4.1.2 Chiffrement par Flux

Les chiffrements par flots (on parle aussi de chiffrements à la volée) sont des chiffres qui cherchent à imiter le chiffre de Vernam, en agissant directement sur chaque bit du texte. Dans le chiffre de Vernam, le seul à garantir parfaitement la confidentialité, on ajoute à chaque bit du texte chiffré le bit correspondant d'une clé parfaitement aléatoire, aussi longue que le message à chiffrer. Dans la pratique, on ne peut pas échanger des clés parfaitement aléatoires arbitrairement longues. Les chiffrements par flots cherchent donc à imiter ceci en produisant, à partir d'une clé courte fixée, une clé arbitrairement longue qui semble parfaitement aléatoire. Ils peuvent donc utiliser par exemple des générateurs de nombres pseudo-aléatoires [28].

En pratique, si on souhaite chiffrer le message $m_1 \ldots m_n$ avec la clé K, un algorithme de chiffrement effectue les opérations suivantes :

La génération du flot de clés : à partir de K_K, on construit une clé $k_1 \ldots k_n$ de même longueur que le message.

Pour chaque caractère mimi du message, on calcule le caractère chiffré correspondant ci / $c_i=E(m_i,k_i)$ où E est une fonction qui prend en entrée un caractère mm et une clé k et retourne un caractère chiffré E(m,k).

Le message chiffré est $c_1 \ldots c_n$.

Le cas le plus simple de chiffrement par flots (et celui qui imite le mieux le chiffre de Vernam) est celui où le message est écrit comme une succession de bits $m_1 \ldots m_n$, où le flot de clés est aussi une succession de bits $k_1 \ldots k_n$ et où l'opération de chiffrement est simplement le ou exclusif, ci= mi⊕ki.

Le principal avantage des chiffrements à la volée est qu'ils permettent de chiffrer et déchiffrer un message en continu, sans avoir besoin de connaitre tout le message. Ceci justifie leur

utilisation dans les domaines où il faut du chiffrement et du déchiffrement en temps réel et simultané (communications téléphoniques, Bluetooth, …).

Parlons un peu sécurité. Comme pour le chiffre de Vernam, la clé ne doit être utilisée qu'une seule fois. Cela peut poser un problème si jamais une clé doit être impérativement être réutilisée. Pour éviter ceci, on n'utilise jamais la clé seule, mais on lui adjoint un nombre utilisé une seule fois, baptisé nunique, contraction de nombre et d'unique. Ce nunique peut par exemple être la valeur d'une horloge bien réglée (et commune à l'expéditeur et au récepteur), ou un compteur de message (on parle alors de chiffrement synchrone, car l'expéditeur et le récepteur n'ont pas besoin de se le communiquer), ou bien un nombre quelconque transmis en clair avec le message (on parle alors de chiffrement asynchrone). Le flot de clés $k_1 \ldots k_n$ est alors généré à partir de la clé secrète initiale K, et du nunique. En particulier, elle n'est utilisée qu'une seule fois.

Si les chiffrements par flots sont souvent plus rapides que les chiffrements par blocs, leur sécurité est souvent plus faible. Par exemple, le chiffrement par flots implémenté dans le protocole de communication des téléphones GSM, bien que non rendu publique, a été cassé quelques années après le début de son utilisation [bibmath.net]

4.2 La cryptographie à clé publique

Malgré que la cryptographie Symétrique assure un niveau de sécurité très élevé, l'échange des clés secrètes posait souvent un grand problème et entrave l'évolution de ces méthodes notamment avec le tournement de ce domaine vers un usage civil et très fréquent.

Pour contourner cet écueil, Whitfield Diffie et Martin Hellman ont découvert, en 1976, une nouvelle façon différente de toutes les propositions tout au long de l'histoire de cette science. Il s'agit d'une grande avancée théorique dont la sécurité ne repose plus sur des heuristiques mais plutôt sur des problèmes mathématiques difficile à résoudre [29]

Le principe de la cryptographie à clé publique peut être expliqué par un exemple très connu suivant les étapes ci-dessous :

> Etape 1 : pour que l'expéditeur puisse envoyer son message à son destinataire, le destinataire commence par envoyer à l'expéditeur un Cadenas ouvert et garde la Clé de ce dernier chez lui.

Etape 2 : l'expéditeur reçoit le Cadenas.

Etape 3 : l'expéditeur envoie la lettre confidentielle dans une boite fermée à l'aide du Cadenas du destinataire.

Etape 4 : le destinataire ouvre la boite fermée puisqu'il est le seul qui possède la clé du Cadenas.

Par analogie à cet exemple, dans la cryptographie à clé publique, nous utilisons deux types de clé :

- Clé publique qui correspond au Cadenas dans l'exemple
- Clé privée, qui représente la clé du Cadenas et permet de retrouver le message clair.
- Le chiffrement s'effectue en utilisant la clé publique
- Le déchiffrement passe par la clé privée.

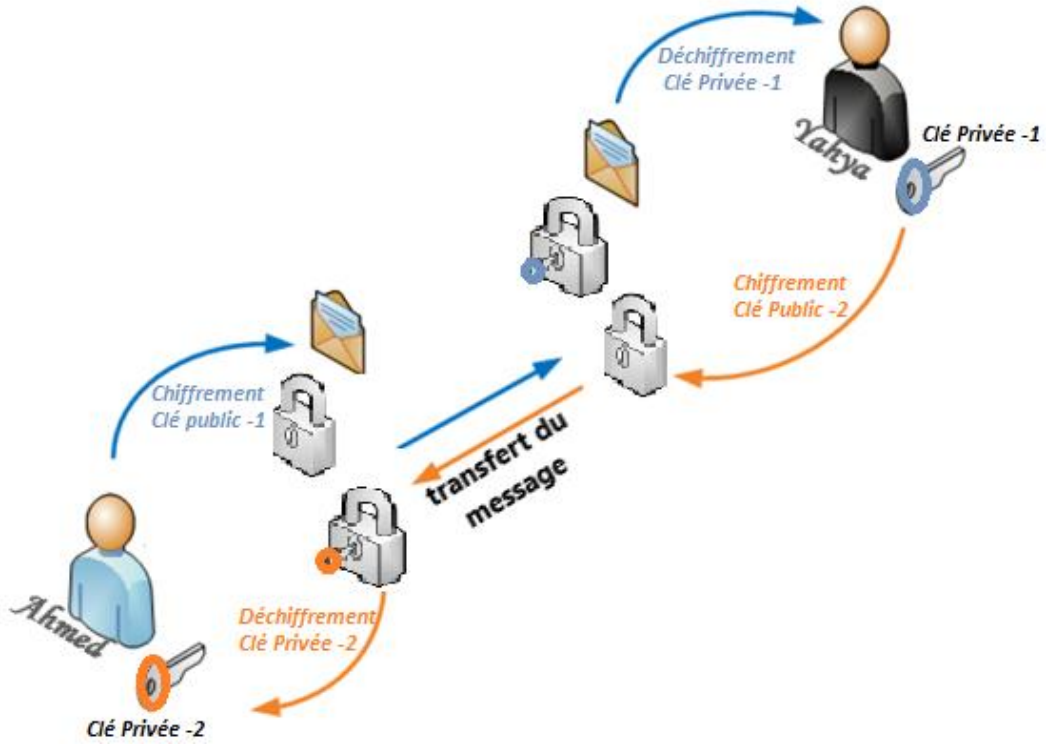


Figure 9 : Principe de chiffrement à clé secrète

4.2.1 Protocole d'échange de clés de Diffie et Hellman

Protocole d'échange de clés de Diffie et Hellman repose sur l'arithmétique modulaire [29], et sur le postulat suivant :

- Étant donné des entiers *p*, *a*, *x* avec p premier et $1 \leq a \leq p-1$:

- Il est facile de calculer l'entier $y = ax \ mod(p)$.

Si on connait $y = ax \ mod(p)$, a et p, il est très difficile de retrouver x, pourvu que p soit assez grand.

Retrouver x connaissant $ax \ mod(p)$, a et p s'appelle résoudre le problème du logarithme discret. Comme pour la factorisation d'entiers, c'est un problème pour lequel on ne dispose pas d'algorithme efficace.

À la fin du protocole, Alice et Bob sont donc en possession d'une même clé secrète K, qu'ils ne se sont pas échangés directement.

4.2.2 Chiffrement RSA

Diffie et Hellman n'ont pas eux-mêmes proposé de fonctions satisfaisantes, mais dès 1977, D. Rivest, A. Shamir et L. Adleman trouvent une solution possible, la meilleure et la plus utilisée à ce jour, la cryptographie RSA[30].

Le RSA repose sur la dichotomie suivante :

- ✓ Il est facile de fabriquer de grands nombres premiers p et q (pour fixer les idées, 500 chiffres).
- ✓ Étant donné un nombre entier $n=pq$ produit de 2 grands nombres premiers, il est très difficile de retrouver les facteurs p et q.
- ✓ La donnée de n est la clé publique : elle suffit pour chiffrer. Pour déchiffrer, il faut connaitre p et q, qui constituent la clé privée. Le problème de factorisation de grands entiers étant très difficile, la connaissance de la clé publique n ne permet pas de retrouver les entiers p et q, qui constituent la clé secrète[31].

4.3 Fonction de hachage

Une fonction de hachage est une primitive cryptographique assurant l'intégrité des données au moyen d'une fonction unidirectionnelle qu'on nomme aussi "fonction de hachage à sens unique" [32]. En effet, ce type de fonction doit être inversible, il s'applique sur un message de longueur arbitraire pour produire un condensé de message H(M) de n bits (qu'on appelle 'empreinte' ou simplement 'haché').

L'objectif principal d'une fonction de hachage cryptographique est de vérifier l'authenticité des données[33]. Cependant, on la retrouve aussi dans les signatures numériques et dans le stockage des données sensibles comme le cas des Mots de Passe. En fait, Le hachage permet

aux utilisateurs d'obtenir une autorisation de données sans connaître le contenu des données. Les mots de passe sont enregistrés sous la forme d'une valeur de hachage ou d'un mot de passe de hachage plutôt qu'en texte clair. La valeur de hachage rend les données plus sûres[34].

Une fonction de hachage doit vérifier les propriétés suivantes[35] :

- Déterministe : pour la même entrée, le résultat doit être le même.
- Non réversible : c'est une fonction inversible ne permettant pas de retrouver l'information d'origine.
- Résistant aux collisions : Deux entrées n'aboutissent pas à la même sortie.
- Non prévisible : Une fonction de hachage génère de manière aléatoire une valeur de hachage unique qui n'est pas prévisible.
- Compression : La sortie de la fonction de hachage est beaucoup plus petite que la taille de l'entrée.

Les propriétés de la fonction de hachage qu'on vient de lister ci-dessus permettent de grantir les caractéristiques suivantes :

- Sécurisé du fait que cette fonction est irréversible et à sens unique.
- Unique du fait qu'elle ne permet de produire la même empreinte pour deux ensembles de données différents.
- Taille fixe du fait que la fonction de hachage donne une empreinte de taille fixe.

4.4 Signature numérique

Les signatures numériques sont des méthodes cryptographiques utilisant les algorithmes à clé publique pour assurer l'intégrité des données. Le principe de ces méthodes est de permettre à l'utilisateur de signer des données en utilisant une signature numérique unique qui lui est appropriée. Ceci dit, quelqu'un d'autre peut vérifier la signature et peut prouver que les données proviennent de cet utilisateur n'ont pas été modifiées après qu'il les ait signées[36].

Le principe de la Signature Numérique repose sur deux types de primitives cryptographiques qui sont : les algorithmes de chiffrement à clé publique et les fonctions de hachage[37].

Le schéma ci-après explique comment la plupart des signatures numériques génèrent et vérifient leurs signatures[37] :

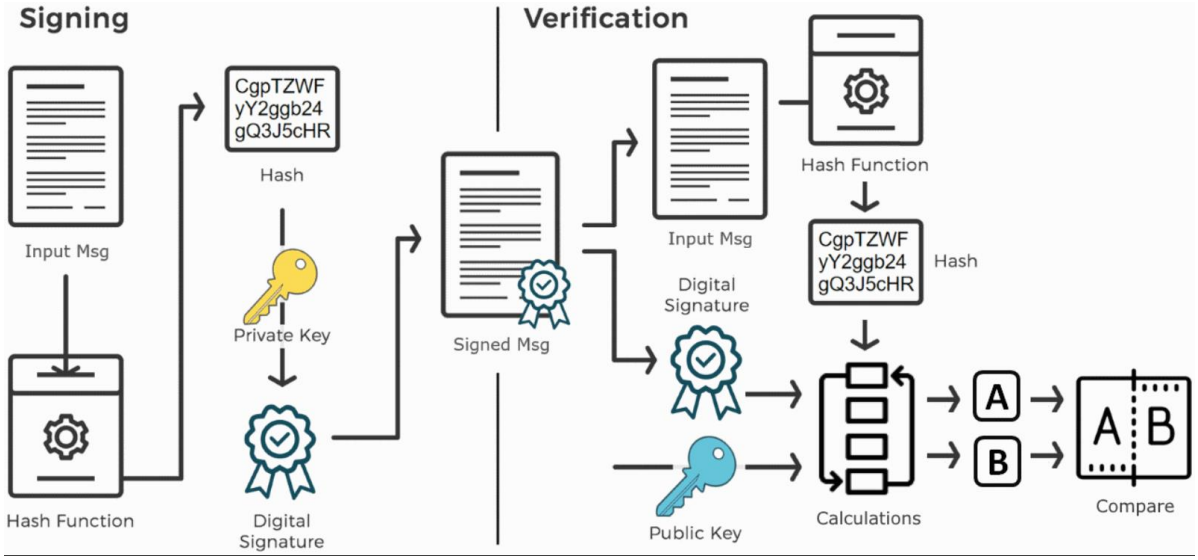


Figure 10 : Schéma de l'approche des signatures numériques

Contrairement aux chiffrements à clé publique qui chiffrent les données en utilisant la clé publique du destinataire, les signatures numériques utilisent le même principe sauf qu'elles exploitent la clé privée dans le processus de chiffrement. En effet, pour signer un document, l'utilisateur commence par appliquer une fonction de hachage sur le document à signer puis chiffre le haché en utilisant sa clé privée[38].

Lors de la vérification de la signature, on procède par le déchiffrement de cette signature en utilisant la clé publique de l'utilisateur en question pour avoir le haché du clair. Après avoir récupérer le message clair à vérifier, ce dernier est haché et enfin une comparaison entre les deux haché décide si la signature est valide ou non[38].

Partie I

Intelligence Artificielle

“

L’intelligence artificielle ne fait pas le poids face à la stupidité naturelle

Albert Einstein

Chapitre 2 |

A Propos de l'Intelligence Artificielle

Sommaire

I- Historique

1.1 L'aube de l'Intelligence artificielle

L'intelligence artificielle n'est pas une technologie fraichement née, c'est un paradigme de recherche que les scientifiques ont commencé à étudier depuis bien longtemps. En effet, cette technologie a été instaurée en faisant intervenir plusieurs disciplines scientifiques. De la Philosophie et la Neurophysiologie, au Mathématique et la Cybernétique, vers l'automatisme et l'informatique, puis l'Apprentissage automatique. Il a grandi petit à petit en passant à chaque fois par des défis qui l'ont fait reculer mais d'autres fois lui permettre de bien se développer, pour devenir la technologie révolutionnaire de ce siècle[39].

La grande problématique derrière la naissance de cette technologie était « comment peut-on imiter le comportement humain ? ». Pour répondre à cette question, il fallait passer par deux principales branches, l'une consiste à comprendre le mécanisme de l'intelligence humaine et l'autre cherche à concrétiser la notion de la pensée en un modèle formel capable d'être implémenté dans des machines. C'est sur cette deuxième branche que l'IA commençait son chemin. La définition de la pensée est le premier souci à confronter. La réponse la plus concrète sur le sujet était proposée par le philosophe grec Aristos qui inventait une première tentative de décrire la pensée humaine en un ensemble de règles formelles indépendantes du contenu[40].

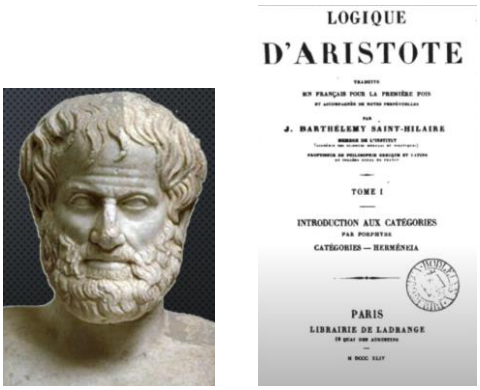


Figure : Aristos (-384,-322) fondateur de la logique

Aristos rejette ainsi la théorie des idées du Platon et considère la connaissance comme un résultat de l'expérience. Selon lui, c'est en observant un concept la première fois que nous

formons l'idée de ce concept et donc l'expérience précède la connaissance[40]. Dans sa théorie de la pensée, Aristos se base sur le syllogisme afin de mener un raisonnement juste. Il s'agit de la logique formelle. La tentative d'Aristos a été ensuite progressée par Leibniz pour donner lieu à une logique utilisant des termes artificiels symboliques.

De la philosophie aux mathématiques, Georges Boole (1815-1864) a initié la première tentative de la formulation mathématique de l'ensemble des lois de la pensée dans ce qu'on appelle l'algèbre de Boole[39]. D'autres mathématiciens continueraient dans cette même perspective.

Augustus de Morgan (1806-1871) proposa la logique des relations par ses fameuses lois de Morgan. Charles Sanders Peirce (1839-1914) développa la sémiotique (théorie des signes) et introduisit les tables de vérité et les quantificateurs. Vient ensuite Gottlob Frege (1879), en proposant dans son livre Begriffsschrift[41] le principe du calcul conceptuel comprenant le calcul des propositions et des prédicats. D'où la naissance de la logique moderne, à la fin du XIXe siècle, engagée également par d'autres mathématiciens de l'époque comme Russell et Whitehead.

$$\text{homme(Socrate)} \wedge (\forall x\ \text{homme}(x) \Rightarrow \text{mortel}(x)) \Rightarrow \text{mortel(Socrate)}$$

Loi de Morgan : $\vdash\ \forall x \forall y\ \neg(x \wedge y) \Leftrightarrow \neg x \vee \neg y$

Démonstration de non (non p) = p :

$$\dfrac{\dfrac{\dfrac{[(p \to \bot) \to \bot] \qquad [p \to \bot]}{\bot}(\to E)}{p}(RAA)}{((p \to \bot) \to \bot) \to p}(\to I)$$

La traduction du raisonnement humain en des modèles logiques et formels a fait évoluer plusieurs d'autres domaines, notamment, l'automatisme, la théorie de l'information et bien d'autres.

Ils ont concrétisé toutes ces notions mathématiques par la création des premières machines pensantes de l'histoire. Les automates, les calculatrices mécaniques, les cartes perforées, les machines analytiques et finalement la machine de Turing en 1936 [42]. Par cette Invention, Alan Turing a franchi l'ère de l'informatique et permettait de décider la calculabilité d'un énoncé. Turing a participé également à la réalisation du Manchester Mark 1, l'un des premiers ordinateurs apparus en 1949[43].

Cependant, la naissance officielle de l'IA a été mise en cause par le théorème de Gödel, qui illustre les limites de l'approche mécaniste et démontre que les machines ne peuvent

tout calculer. En revanche, Turing oppose cette théorie et considère que le futur des ordinateurs est très prometteur et peut imiter l'intelligence du cerveau. En 1950, ce dernier confirme son intuition et publie dans son article intitulé « Computing Machinery and Intelligence » un test pour mesurer l'intelligence d'une machine et la comparer à celle des hommes. C'est le fameux test de Turing ou encore le jeu de l'imitation[44].

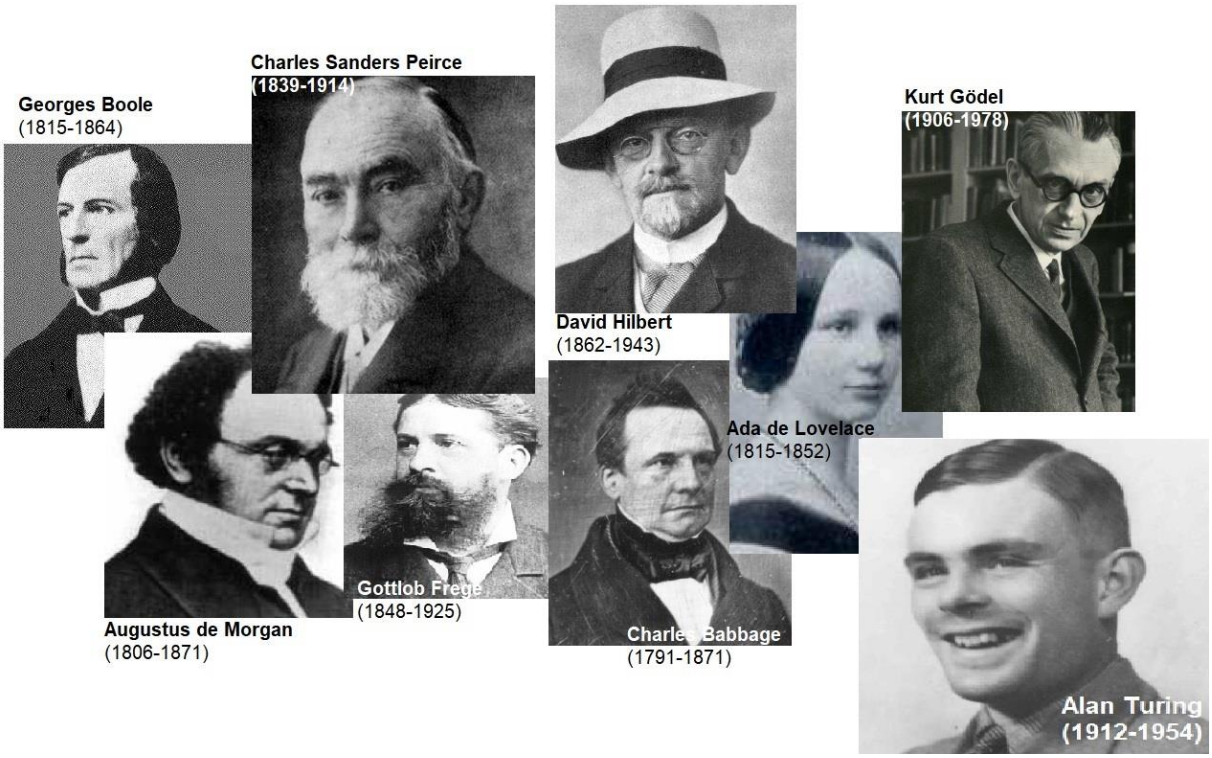


Figure 11 : les premiers mathématiciens participants dans la réflexion de l'IA

Un peu plus loin des mathématiques et bien avant l'apparition des ordinateurs. L'autre axe historique de l'aube de l'IA évoluait également au fur et à mesure, surtout après la perception de Santiago Ramon Y Cajal concernant le rôle du cerveau. En fait, dans l'antiquité, la liaison entre le cerveau et la pensée était obscure. Aristos, par exemple, pensait que : « le cœur est au centre des processus sensitifs, le cerveau ayant alors un simple rôle de la réfrigération de ce dernier »[45]. Ce n'est qu'à la fin du XIX siècle que le neuroscientifique Cajal a mis en lumière le fonctionnement de cet organe, et l'identifie comme le siège de la pensée[46]. Il considérait le cerveau comme un ensemble d'entités interconnectées entre elles et fonctionnant d'une manière autonome. Convaincus par cette vision, plusieurs chercheurs en médecine et neurologie continuaient sur cette piste scientifique. Notamment, l'anatomiste allemand Waldeyer que l'on doit le mot

« Neurone » en 1891. Puis en 1897, Charles Scott Sherrington qui a créé le mot Synapse pour désigner le point de contact entre un neurone et un autre neurone[47]. Petit à petit, les chercheurs ont pu résoudre toutes les ambiguïtés sur ce système complexe, et décider finalement la structure des neurones et leur physiologie[48].

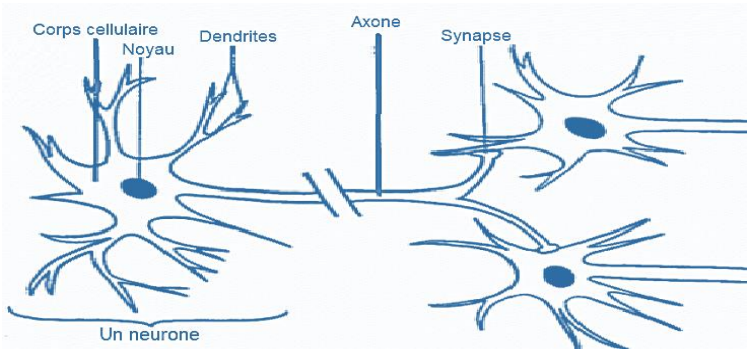


Figure 12 : Structure d'un neurone biologique

Comme présenté dans le schéma ci-dessus, le Neurone est donc un corps cellulaire qui reçoit des signes en entrée via des milliers Dendrites et envoi ses signaux à travers l'Axone. La communication entre les neurones passe par le biais des Synapses.

Cette représentation biologique du cerveau, inspire les mathématiciens encore une autre fois! La collaboration entre le neurologue américain Warren McCulloch et le mathématicien Walter Harry Pitts a été couronnée par la création du premier neurone artificiel en 1943[49].

En 1949, Hebb a mis en évidence l'importance du couplage synaptique dans l'apprentissage par renforcement ou dégénérescence des liaisons inter-neuronales lors de l'interaction du cerveau avec le milieu extérieur.

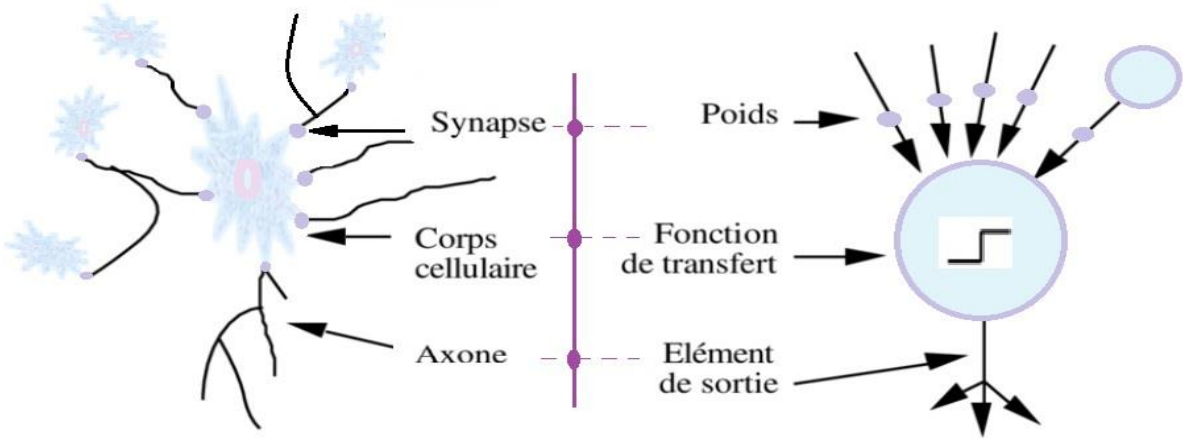


Figure 13 : Analogie entre neurone biologique et neurone artificiel

Tout simplement, cette première reformulation du neurone artificiel a été définie par une fonction mathématique à plusieurs variables (Figure13). Des variables en entrée X_i pour lesquels on peut attribuer des poids W_j (les coefficients synaptiques) représentant la force de la connexion. Ensuite, on calcule leur sommation pondérée tout en respectant un certain seuil ou encore une fonction d'activation (g) qui définit le potentiel de sortie d'un neurone en fonction des niveaux d'activité de ses entrées. Finalement, nous retrouvons la valeur de l'élément de sortie S_j[50] .

Voici, un petit exemple simple schématisant ceci :

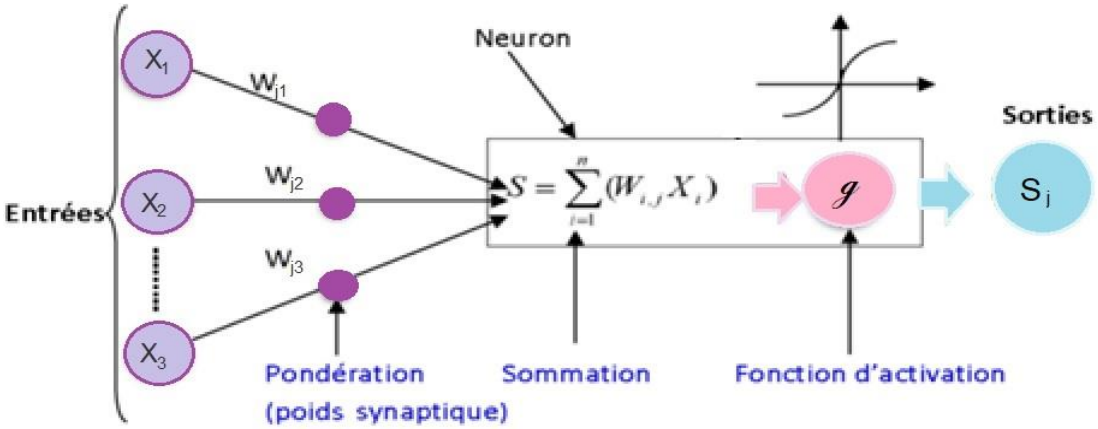


Figure 14 : principe du calcul d'un neurone artificiel simple

Le contexte intellectuel devient de plus en plus favorable. L'intersection des deux axes de l'émergence de l'IA prend son sens. La plupart des scientifiques affirme que le corps n'est qu'une « machine » et les phénomènes physiques suffisent à expliquer la vie, et la pensée[51].

À l'issu de ce rassemblement, le rêve de l'IA commence à prendre corps. Les chercheurs ont conventionné officiellement et pour la première fois, lors de la Conférence Dartmouth en 1956, le terme de l'Intelligence Artificielle (IA)[42] [40].

Figure 15 : les fondateurs officiels de l'Intelligence Artificiel (Conférence Darmouth 1956)

Depuis, la machine ne cesse d'apprendre, d'auto-apprendre et de développer ses mécanismes, ses capacités techniques et ses réseaux de neurones.

1.2 Les premières applications IA

Les applications intelligentes fleurissent dans plusieurs domaines, d'ailleurs dans la même année de la fameuse conférence Dartmouth, Alan Newell et Herbrt Simon ont développé le premier logiciel d'IA « Logic Theorist » [52] capable de démontrer des théorèmes mathématiques en toute autonomie.

En 1957, le psychologue Frank Rosenblatt invente le premier programme d'apprentissage grâce à un réseau de neurone simple appelé le perceptron[39].

Noam Chomsky[50], chercheur Linguiste américain, invente des modèles mathématiques pour différents langage afin de les rendre compréhensible aux machines.

Plutard en 1965, Joseph Weizenbaum développe le premier programme informatique « Eliza » réussissant les tests de Turing[53]. Il remplace les psychothérapeutes pendant les entretiens thérapeutiques.

1.3 « L'Hiver de l'IA »

Néanmoins, l'IA suscite autant d'espoir que de craintes. Malgré l'éclaircissement de tous les concepts scientifiques sur cette discipline. Les années 1970 marquaient une période de crise pour l'intelligence artificielle. On parle de « l'Hiver de l'IA »[39]. La recherche s'effondre suite à des illusions scientifiques. Les premières questions éthiques émergent et séduisent les premiers reproches.

En 1965, le philosophe américain Husbert Dreyfus publie une première approche critique sur l'intelligence artificielle. Il conteste le fait que l'intelligence soit réduite à de simple calcul et il souligne l'importance des émotions et du ressenti du cerveau humain[54].

En 1973, le mathématicien Michael James Lighthil a mis aussi des critiques dans un rapport commandé par le parlement britannique avec pour conséquence le gel des financements européens pour l'IA [39].

Le blocage est aussi lié à la puissance des ordinateurs de cette époque. D'après la loi de Moore[15], la capacité de ces machines est toujours limitée et reste incapable de mener à bien des projets de l'intelligence artificielle. Les projets en IA reculent et les investissements s'orientent plus sur des objectifs réalistes de court à moyen terme.

1.4 Vers la Renaissance

Les opportunités géantes de cette discipline ne laissaient pas durer ses années noires. 6 ans plus tard dans les années 1980, l'IA reprend ses forces et les investissements repartent à la hausse grâce à la l'invention de ce qu'on appelle les « Systèmes Experts »[54]. Ils s'agissent des programmes qui agissent sur un domaine spécifique et résolvent l'ensemble de ses problématiques en se basant sur les mêmes analyses que l'expert humain.

C'est aussi le moment où se développent les algorithmes d'apprentissage. Des programmes plus complexes mais avec une approche plus performante et optimiste.

Cependant, les industriels sont toujours retissant pour investir dans des projets IA. Les ordinateurs personnels ont participé aussi à cette résistance. On se contente d'exploiter plus de programmes classiques pour répondre aux différents besoins dans la mesure du possible. Face à cette situation, le développement de l'IA reste encore otage des laboratoires de recherches.

Stimulé par le progrès Informatique surtout le tout début du BigData dans les années 90, l'IA a finalement pu casser ces entraves et l'espoir revient avec sa nouvelle formule[50]. Des inventions très puissantes capables de concurrencer l'intelligence humaine, commencent à voir le jour. On cite notamment :

- 1997 : l'IA Deep Blue batte Garry Kasparov (le champion du monde d'échéc de l'époque)
- 1997 : Développement d'un logiciel de reconnaissance vocale et son installation sur un système Windows
- L'essor du Deep Learning par Yann leCun et développement des applications de reconnaissance d'écritures et d'images

1.5 L'ère de la maturité

Au cours de cette dernière décennie, la puissance de calcul des ordinateurs, la capacité de stockage et l'accumulation des données, augmente d'une façon exponentielle. Grâce à cette immense évolution technique et l'émergence des technologies BigData, l'IA vie un nouveau tournant dans son histoire[55]. Les grandes boites s'y plongent très profondément pour augmenter leur compétitivité dans le marché. Notamment, Facebook par la création de l'algorithme de connaissance faciale DeepFace. Puis, Google par la création du l'IA AlphaGo qui a battu le champion européen du GO en 2015.

Dans tous les domaines, l'IA a pris sa place et émerge notre vie quotidienne par ses inventions :

- Les voitures autonomes,
- La traduction simultanée d'une conversation,
- Les suggestions des moteurs de recherche
- Analyse et modification artistique des images par Deep Dream.
- Détection des Fraudes
- Diagnostic médical et analyse des Radio
- Et bien beaucoup d'autres applications très pertinentes

Aujourd'hui, L'IA trace les grandes lignes de la 4ème révolution industrielle. Elle est au cœur des préoccupations scientifiques à l'échelle mondial et dans tous les secteurs.

Néanmoins, ce progrès technologique est loin de faire l'unanimité et il soulève encore des questions sur le futur de cette discipline et son impact sur la société[56]. Des inquiétudes qui touchent surtout l'aspect social et humain et essaye de répondre à des questions telles que : La machine remplacera elle bientôt l'être humain ? Peut-elle dépasser l'intelligence humaine ? Est ce qu'on saura toujours contrôler cette intelligence ? Quel sera le statut juridique sur ces créations artificielles ?

D'autres problématiques éthiques sont également omniprésentes. La protection des données est l'un des inconvénients majeurs qui interrompe les chercheurs et les scientifiques. En effet, la performance des systèmes intelligents est liée principalement à leur capacité d'apprentissage qui, à son tour, impose une masse de données immense pour l'extraction des connaissances et l'enrichissement de ce type de système.

Une alimentation permanente des données issues de notre monde réel est d'une importance stratégique pour la durabilité d'un logiciel IA. Or, la très grande majorité de ces richesses est centralisée chez des organismes privés. Ceci donne à réfléchir sur comment peut-on sécuriser la fuite de ces données et contrôler leur usage ? Puis, comment peut-on les extraire tout en respectant la vie privée des citoyens qui en sont à l'origine ?

Pour conclure, on peut dire que l'IA devient de plus en plus très répandue et populaire, mais ne voit pas encore son printemps arrive. Est-ce qu'il s'agit d'un voyage éternel ou sera bientôt aboutit ?

II- Les Approches de l'IA

L'IA intervient aujourd'hui dans divers secteurs d'activité et vise à aboutir à des solutions plus intelligentes que les systèmes classiques de façon à inclure l'aspect cognitif dans leur conception. Ceci-dit, augmenter la capacité cognitive des entreprises telles que percevoir, raisonner, apprendre, dialoguer, décider, et agir logiquement, efficacement et à temps réel. Pour ce faire, les chercheurs se sont plongés dans une variété de procédures techniques et analytiques qu'on peut placer sur deux principales approches : l'IA connexionniste et l'IA Symbolique [50].

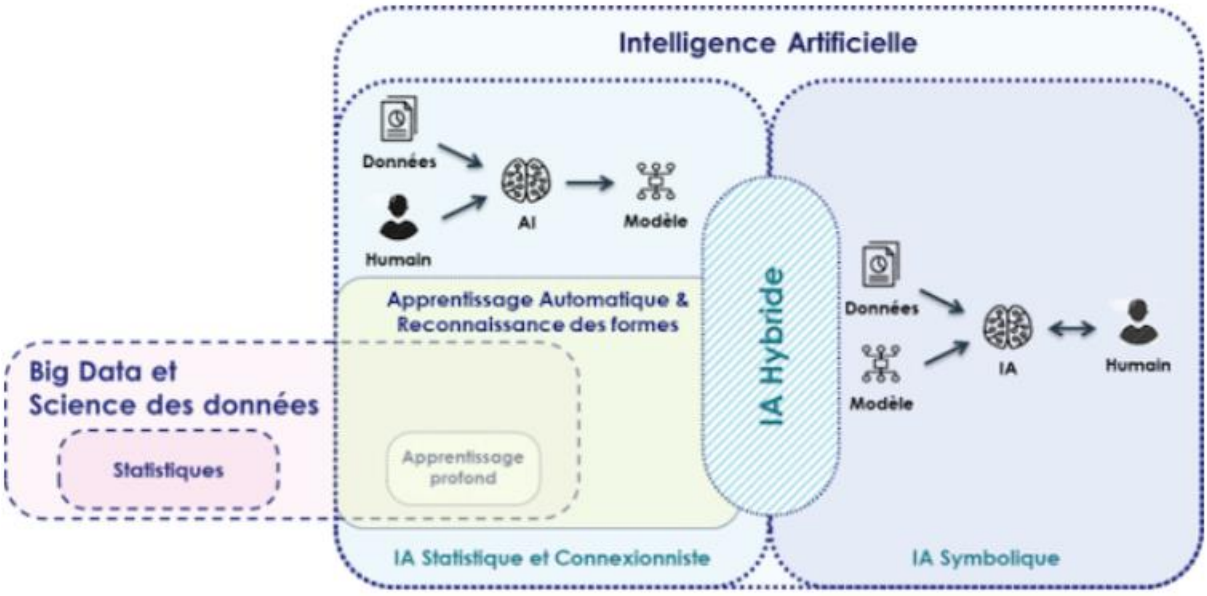


Figure 16 : Les deux grandes approches de l'IA, IA connexionniste et IA symbolique

IA Symbolique part plus dans l'approche philosophique qui vise à « penser comme l'homme » et elle suppose l'existence d'un « langage de la pensée ». Autrement dit, une représentation mentale de type propositionnel[57]. De ce fait, l'apprentissage symbolique considère les connaissances et le raisonnement comme des recombinaisons logiques de symboles élémentaires. Elle traite chaque problématique dans un espace continu des points isolés sous forme d'une modélisation discrète[58]. Ce mode de description possède un contenu sémantique avec une structure linguistique compréhensible et proche du langage humain, ce qui constitue un avantage majeur pour sa mise en œuvre. Cette IA reste pertinente pour la résolution de problème complexe dans un contexte d'incertitude comme la décision sous contraintes. Néanmoins, elle donne toujours à réfléchir sur comment peut-on décrire et reconnaitre les représentations approximatives ?

L'IA connexionniste[58], quant à elle, ne se focalise pas sur la formalisation de la pensée humaine mais il part plus sur l'approche Socratienne qui vise à « Acter comme l'homme ». Est donc, peu importe le raisonnement derrière un comportement spécifique, formalisé ou non, logique ou pas, ce type d'IA s'intéresse plus à copier l'action finale et faire trainer le système pour arriver à cette même conséquence. C'est une sorte de régularisation qui s'adapte au fur et à mesure aux nouvelles données pour faire émerger des représentations internes. La définition des règles passe d'une manière discrète et n'est pas très maitrisés. C'est comme une boite noire qui cherche à retrouver son monde via

des procédures d'apprentissage incrémentales changeant leur composition en fonction des d'exemples d'informations à sa disposition. Et donc, l'efficacité de ces systèmes connexionnistes est très relative à la richesse et la masse des données en entrée[58].

Ainsi ce type d'IA exploite de nouvel espace de modélisation plus vaste qu'auparavant et permettent plus de flexibilité et de souplesse. Chose qui représente un avantage pour lui. Cependant, deux problématiques majeures le serre toujours. La première est due à la difficulté de pouvoir comprendre, interpréter et expliquer le comportement des couches intermédiaires qui sont dans la plupart des fois très flous et difficile à décrire. La seconde, est dû au fait que cette IA est très mal adaptée à la résolution de besoins complexes et se contente des besoins en termes de la perception.

Plusieurs axes de recherches sont toujours ouverts dans cette discipline et de nouvelles pistes très prometteuses sont en train de se développer et s'oriente plus vers une Hybridation de ces deux approches qui peuvent se compléter l'une avec l'autre.

III- Domaines de l'IA

La variété des sujets traités par l'IA est assez étonnante : Démonstration de théorème, Systèmes experts, Simulation des jeux, Reconnaissance de formes, Apprentissage, Induction, robotique[56].

Connexionniste ou symbolique, l'IA intègre une ou plusieurs techniques dans sa démarche pour aboutir à ces besoins. Ces outils techniques sont très nombreux et s'améliorent continument pour inclure les différentes propriétés de l'intelligence.

En général, la structuration de ce domaine en sous domaines n'est pas facile et font sujet de plusieurs débats. Si on fait une catégorisation suite aux besoins déjà traités en termes d'IA, on se retrouve avec une interaction emmêlée entre différents mots clés et domaines techniques. INRIA, par exemple, dans son livre blanc [51] propose d'organiser son chantier de recherche en IA en se basant sur huit sous-domaines classés dans la hiérarchie mentionnée ci-dessous. :

	Sous-Domaines fonctionnels d'IA	Domaines techniques
1	CONNAISSANCES	Bases de connaissances Extraction & nettoyage de connaissances Inférence Web sémantique

		Ontologies
2	TRAITEMENT DU LANGAGE NATUREL	
3	APPRENTISSAGE AUTOMATIQUE	Apprentissage supervisé Apprentissage (partiellement) non-supervisé Apprentissage séquentiel et par renforcement Optimisation pour l'apprentissage Méthodes bayésiennes Réseaux de neurones ou neuronaux Méthodes à noyau Apprentissage profond Fouille de données Analyse de données massives
4	TRAITEMENT DES SIGNAUX : Parole et Vision	Reconnaissance d'objets Reconnaissance d'activités Recherche dans des banques d'images et de vidéos Reconstruction 3D et spatio-temporelle Suivi d'objets et analyse des mouvements Localisation d'objets Asservissement visuel
5	ROBOTIQUE	Flottes de robots Apprentissage des robots Cognition pour la robotique et les systèmes
6	NEUROSCIENCES, SCIENCES COGNITIVES	Compréhension et stimulation du cerveau et du système nerveux Sciences cognitives
7	ALGORITHMIQUE DE L'IA	Programmation logique et ASP Déduction, preuve Théories SAT Raisonnement causal, temporel, incertain Programmation par contraintes Recherche heuristique Planification et ordonnancement
8	AIDE À LA DÉCISION	Business intelligence

Tableau 2 : Hiérarchie de mots-clés pour l'IA établie par INRIA

Pour répondre à un besoin en termes d'intelligence artificielle on peut faire intervenir plusieurs axes de réflexion au même temps[51]. Dans le cadre de cette thèse par exemple, et suivant la catégorisation d'INRIA, nous allons traiter notre problématique en faisant appel aux deux premiers sous-domaines, à savoir; le sous-domaine des Connaissances et celui d'apprentissage automatique.

Dans la suite de ce chapitre, nous allons se concentrer plus sur ces deux sous-domaines pour étudier en près le fonctionnement des procédures utilisées dans notre contexte.

IV- Système à Base de Connaissances

Un système à base de connaissances est en grande partie entre dans le cadre de l'IA symbolique dont la cognition, l'inférence et le raisonnement représentent ses trois piliers principaux[9].

- La cognition dans le sens où il interagit sur la base des connaissances issues du monde réel. Il reflète ainsi, tout un ensemble d'activité mentale en termes de réflexion, perception, conscience, raisonnement, émotion, langage …
- L'inférence, quant à elle, assure l'autonomie et la continuité du système dans le cas des nouvelles informations et donc sa capacité de reproduire de nouvelles connaissances logiques et compréhensibles.
- Finalement, le pilier raisonnement qui touche la finalité du système et sert à répondre à la problématique métier pour laquelle le SBC était mise en place. Et donc, il essai de monter une succession d'inférence capable de répondre à une question bien définie.

En gros, un SBC cherche à monter une expertise artificielle sur un domaine spécifique ou une pratique bien définie. Il se base sur la manipulation des connaissances issues d'une expertise humaine ou autres et essaie de produire de nouvelles connaissances en suivant un raisonnement formel de règles et d'axiomes. Et finalement, il exploite ces savoirs ainsi « conservés » pour répondre aux différentes situations demandées[59].

4.1Prérequis de la réalisation d'un SBC

Avant de présenter l'architecture technique d'un SBC, il faut satisfaire un certain nombre de prérequis concernant [60]:

- Le domaine d'application de ce système
- Le type de problèmes à régler.
- Les attentes sur chaque sujet
- Puis, Le type d'informations à notre disposition

4.1.1 Prérequis liés au domaine d'application

Le domaine d'application d'un SBC joue un rôle très important dans la faisabilité de ce type de système. En général, on ne peut pas partir dans cette piste dans un domaine dont le savoir-faire n'est pas bien maitrisé et dont l'origine de ses connaissances n'est ni expérimentale ni heuristique. En revanche, un SBC est plus utile quand les connaissances sont très éparses, très consistantes et augmentent ou changent avec le temps. En effet, l'utilisation des algorithmes classiques dans ces cas de figure n'est pas recommandée du fait que leur complexité devient exponentielle au fil de l'eau. En outre, un algorithme classique nécessite, pour sa réalisation, des entrées bien structurées qu'on peut coder facilement, chose qui ne se réalise pas dans le cas d'un SBC, on se base généralement sur des savoir-faire peu structurés et difficilement codables.

4.1.2 Prérequis liés au type des problèmes

Résoudre un problème en utilisant un SBC est un chantier très couteux que ça soit en termes de coût ou d'effort. Proposer une solution basée sur un SBC n'est pas une décision évidente pour une entreprise. Les problèmes traités par le biais de ce type de système doivent être d'un poids considérable afin de pouvoir justifier sa construction, et par conséquence, justifier la non faisabilité ou la difficulté de la résolution de ce type de problème en utilisant les méthodes de calcul traditionnelles.

D'autre part, Il faut aussi déterminer si le type de problème en question est adapté au contexte d'un SBC. C.à.d., voir s'il s'agit d'un problème cognitif dont le besoin peut être définit suivant un enchaînement de raisonnements, de la perception, la conscience et l'apprentissage.

4.1.3 Prérequis liés aux attentes

La définition des attentes est aussi un facteur clé pour la réussite d'un SBC. Déjà, l'implémentation technique de ce type de système offre un certain nombre d'avantage que peuvent satisfaire par défaut des attentes comme :

- ✓ Stocker les connaissances et leurs sémantiques dans le contexte d'une pratique bien définie
- ✓ Conserver le savoir et le savoir-faire dans le domaine
- ✓ Faciliter l'exploitation et la manipulation de ce savoir-faire
- ✓ Sinon, produire de nouvelles connaissances pour répondre aux autres demandes dans le même contexte.

Côté fonctionnel, les attentes qu'un SBC peut satisfaire doivent en général s'inscrire dans l'un des contextes suivants :

- ✓ Prédiction
- ✓ Prise de décision
- ✓ Contrôle comportemental
- ✓ Conception
- ✓ Diagnostic
- ✓ Planification
- ✓ Réparation d'un disfonctionnement
- ✓ Interprétation de données abstraites
- ✓ Monitoring
- ✓ Simulation
- ✓ Sélection

4.1.4 Prérequis liés au type d'information

Une fois on est dans le contexte de la réalisation d'un SBC, il faut étudier le type d'informations que l'on dispose.

En effet, avoir une information tout court n'est pas suffisant pour un SBC car il faut absolument convertir cette information à une connaissance. Autrement dit, il faut disposer pour chaque information ou un groupe d'information un mode d'emploi ou une logique de fonctionnement permettant d'en extraire de la connaissance. Et donc, réaliser un SBC

nécessite une étude préalable sur notre capacité de construire les connaissances nécessaires à travers les informations à notre disposition.

4.2 Architecture d'un SBC :

Comme évoqué précédemment, un SBC se fonde sur trois objectifs principaux qui sont : la cognition, l'inférence et le raisonnement. De même, l'architecture la plus simple pour ce type de système repose essentiellement sur trois composants techniques dont chacun rempli un des objectifs prédéfinis[9]. Dans cette architecture nous trouverons :

- ✓ Une base de connaissances
- ✓ Une base des Faits
- ✓ Un moteur d'inférence

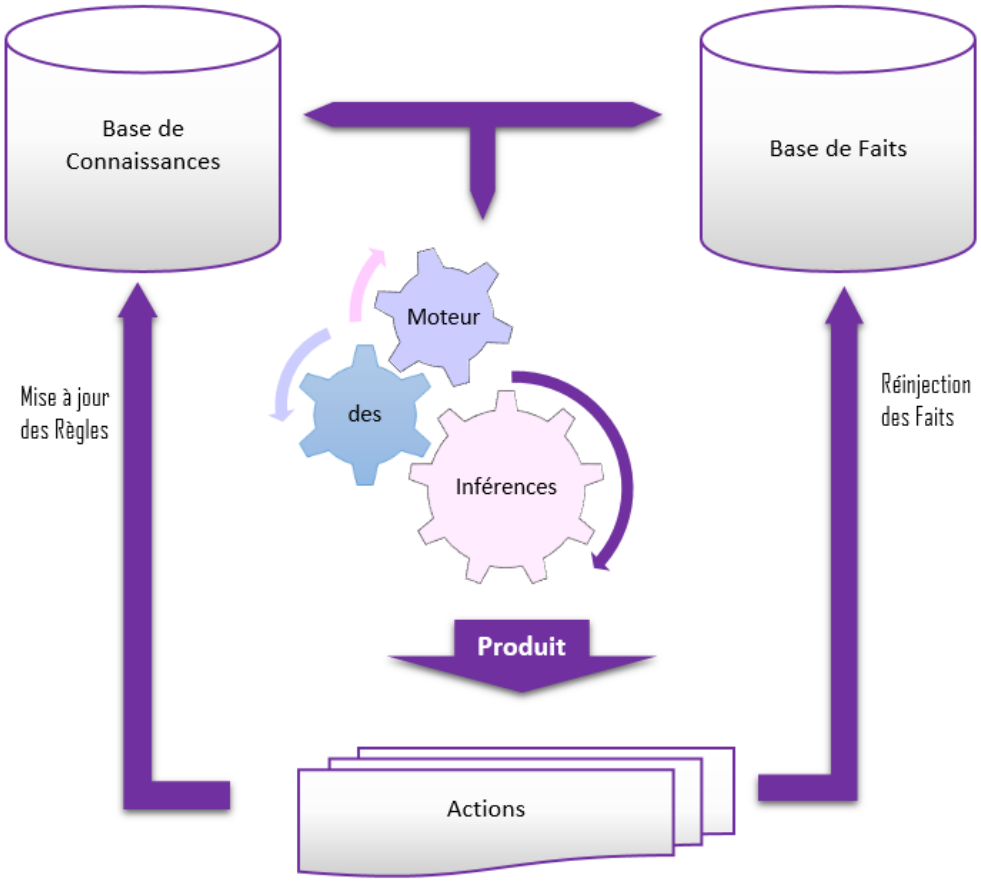


Figure 17 : Les composants principaux d'un SBC

Dans cette architecture, la séparation entre les connaissances et les inférences représente un facteur clef pour la réussite du SBC[61]. En effet, ces trois composants sont le cœur d'un SBC et doivent être implémentés en utilisant un codage différent pour chaque entité. Cette séparation est d'une grande utilité et favorise l'indépendance des traitements malgré les changements et les évolutions que la base de connaissances peut subir. En outre, les modifications au niveau du moteur d'inférence sont aussi tolérables et on pourra même les exécuter et les tester sur la même base de connaissance sans impacter son évolution. Cela dit, la maintenabilité et l'évolution de ces composants sont plus flexibles et très fiables[61].

Néanmoins, la réalisation de ce système nécessite d'autres composants pour sa mise en production, notamment des composants pour :

- L'acquisition des connaissances
- Et l'utilisation finale de ces connaissances

La construction d'un tel système doit passer par les étapes suivantes :

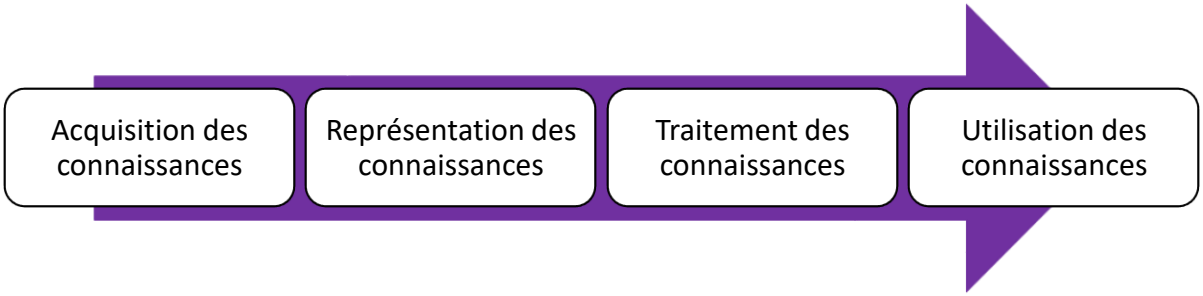


Figure 18 : Les étapes de la construction d'un SBC

4.2.1 Acquisition des connaissances

Dans la première étape de l'acquisition des connaissances, on doit définir les différentes sources possibles pour la collecte des connaissances en liaison avec la problématique demandée. Cette acquisition peut passer par la coopération entre experts des domaines en question et/ou par le biais d'une interaction avec les bases de données et les systèmes existants. Sinon d'autres possibilités sont aussi valables comme l'utilisation des

techniques heuristiques en se basant sur un historique précédemment instauré par la pratique.

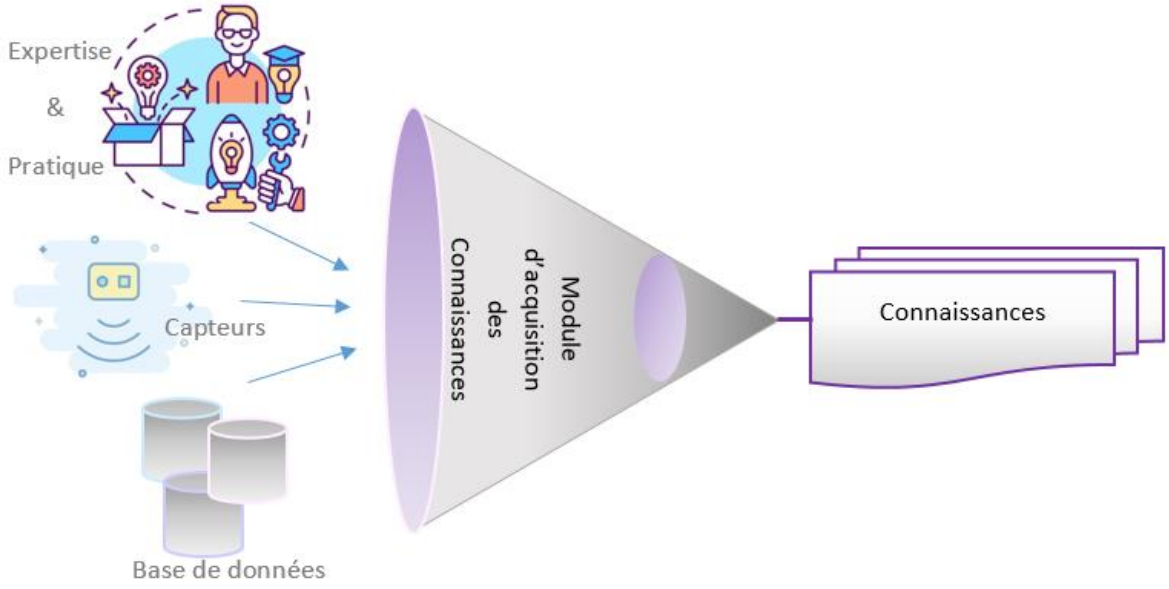


Figure 19 : Module de l'Acquisition des connaissances

4.2.2 Représentation des connaissances

4.2.2.1 Définition de la connaissance

Définition : *Une connaissance c'est une mobilisation d'une ou plusieurs informations dans un contexte bien définit afin de déclencher une action ou produire une autre connaissance*[60].

4.2.2.2 Donnée, Information et connaissance

La plupart des gens confuse la notion de la connaissance avec celle de l'information sinon la donnée même. En fait, ces trois notions abstraites sont très liées et d'un niveau d'abstraction superposé dont la donnée et la plus petite entité.

- La donnée est une représentation linguistique et symbolique d'une notion, d'un objet ou d'une réalité.

- L'information par contre, c'est une interprétation de cette donnée intelligible afin de lui affecter son sens informatif
- La connaissance est une notion très relative au sujet qu'il traite. Elle s'agit d'une exploitation d'un ensemble d'information dans un mode d'emploi sémantique (analytique, décisionnel, logique, analogique, …) en vue de leur conférer un sens plus large pour agir sur un but précis.

Par exemple :

Donnée brute : « SOS »

Information : reflète le service d'urgence médicale.

Connaissance : Si appel SOS Alors envoyé le médecin à l'adresse désignée dans l'appel.

4.2.2.3 Types de la représentation d'une connaissance

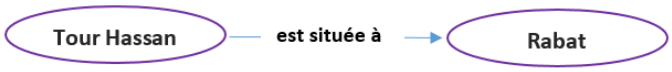
Il y'a plusieurs manières pour exprimer une connaissance, on trouve des connaissances déclaratives, d'autres procédurales et une troisième catégorie structurée qui regroupe les deux puis des connaissances sur les connaissances qu'on appelle méta-connaissance.

Pour une connaissance déclarative, la connaissance vient pour déclarer comment une action est faite (Si … Alors …) d'où la notion des « faits » qu'on retrouve dans l'architecture d'un SBC. Autrement dit, cette connaissance détient la logique du fonctionnement et définit les objets et les concepts induisant à un tel « fait ». La base des faits dans ce cas de figure est indépendante de la structure procédurale et permet d'inclure toutes sortes de faits sur chaque domaine, ce qui représente son point fort. Cependant cette architecture est plus lente car elle s'appuie sur l'interprétation des procédures à chaque exécution.

Par ailleurs, dans une connaissance procédurale on y retrouve déjà son mode d'emploi du fait qu'elle reflète la logique de l'action sous forme de « règles, procédures, stratégies, agendas ». L'avantage dans ce type d'architecture c'est que l'exécution des règles est plus rapide car la connaissance est codifiée dans des procédures compilés et prêtes à être exécutées. Néanmoins, l'accessibilité aux données devient plus compliquée car elles sont imbriquées dans le code compilé.

Maintenant, en IA, la représentation de la connaissance dans chaque catégorie est formalisée différemment. On peut la représenter sous forme de :

- Triplet <objet, attribut, valeur> : dont l'objet est le sujet à traiter, l'attribut est la propriété à considérer, valeur est la valeur de cette propriété
  Exemple :

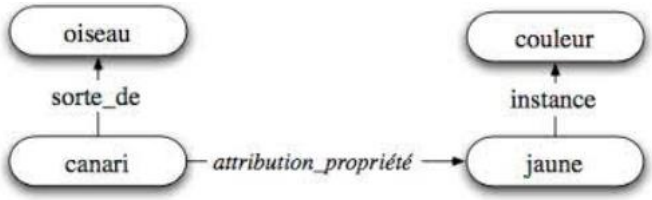


- Formule logique : à l'aide du calcul des prédicats et des propositions on arrive à présenter une situation
  Exemple :

Si une personne est femme et est un parent, alors cette personne est la mère de quelqu'un.

- F(x): "**x est une femme**", P(x): "**x est un parent**"
- M(x,y): "**x est la mère de y**"
- Pour chaque personne x, Si x est un parent, alors x est la mère de quelqu'un.
- $\forall x ((F(x) \wedge P(x)) \rightarrow \exists y M(x,y))$

- Réseau sémantique : dont les nœuds sont les concepts et les arcs représentent les relations.
  Exemple :



- Règle : liaison entre informations pour déduire d'autre informations dans le but de conclure une relation, une stratégie, une directive, une heuristique …
  Exemple :
  - Si <Fleur, Couleur, Rose> Alors « J'aime la Fleur »
  - Si « J'aime la Fleur » alors « J'achète la Fleur »
  - Si « J'aime la Fleur » et <Emballage, Prix, Gratuit> Alors « J'achète bouquet de Fleurs »

4.2.2.4 Base de Connaissances

En général, chaque unité de connaissance peut être divisée en deux parties, la première partie représente une condition de déclenchement de l'action qu'on appelle Prémisses et

la deuxième représente l'effet du lancement de cette action et qu'on appelle Conclusion. Cette conclusion correspond à un résultat qui à son tour peut déclencher une autre action ou se statuer dans une situation finale qu'on désigne par un fait.

De ce fait, une base de connaissance permet de stocker ces différents éléments dans des structures indépendantes. On trouve alors :

Une base de règles : contenant des connaissances sous forme de règles de type « Si Prémisses Alors Conclusion »

Une base de Faits : contenant l'ensemble des faits qui décrivent les différentes situations possibles (vrai) pour une problématique précise.

4.3 Traitements des connaissances

4.3.1 Moteur d'inférence

Le moteur d'inférence est un ensemble d'instructions informatiques permettant de conduire un raisonnement en fonction des savoirs et savoir-faire inscrits dans la base de connaissances. Il exploite la base de règles, fait enchainer des inférences puis conclut de nouveaux faits pour répondre à un but prédéfini[61].

Pour accomplir sa tâche efficacement, le moteur d'inférence doit être capable de détecter et gérer les cas suivants :

1. Désigner l'ensemble des règles de la BR à comparer avec les faits de la BF
2. Spécifier l'ordonnancement de ces règles en adéquation avec le besoin demandé.
3. Déclencher l'exécution des règles choisies suivant la stratégie d'enchainement précédemment précisé.
4. Détecter les mises à jour de la base de faits et les appliquer
5. Gérer les règles en double et éliminer les règles qui sont déjà utilisée
6. Détecter les règles qui peuvent provoquer des confusions et éliminer les contradictions

BF étant la base des Faits et BR est la base des Règles, ci-dessous un exemple de l'algorithme d'un moteur d'inférence :

```
Modif ← vrai
Tant que Modif
        Modif ← faux
        Pour chaque règle R Faire :
                Si (¬appartient (R, BF)
                        Et Si appartient (prémisses(R), BF))
                Alors
                        Pour chaque conclusion C de R Faire:
                                Si (¬ appartient(C, BF))
                                Alors message d'erreur
                                Sinon Ajouter(C, BF)
                                FinSi
                        FinBoucle
                        Modif ← vrai
                        Marquer R
                FinSi
        FinBoucle
FinTantQue
```

Figure 20 : Algorithme d'un moteur d'inférence

4.4 Utilisation des connaissances

Comme, dans le module d'acquisition, le SBC dispose d'un système de dialogue avec les experts, il doit également avoir un moyen de communication avec les utilisateurs finaux. Il s'agit d'une interface graphique permettant d'assurer la compréhension du raisonnement réalisé par le système et capable d'expliquer chaque Fait généré. Pour se faire, le SBC met en place un système de gestions des logs afin d'exploiter ses traces et expliquer les différentes actions effectuées par le moteur d'inférence.

Ci-après un schéma regroupant les composants d'un système à base de connaissance :

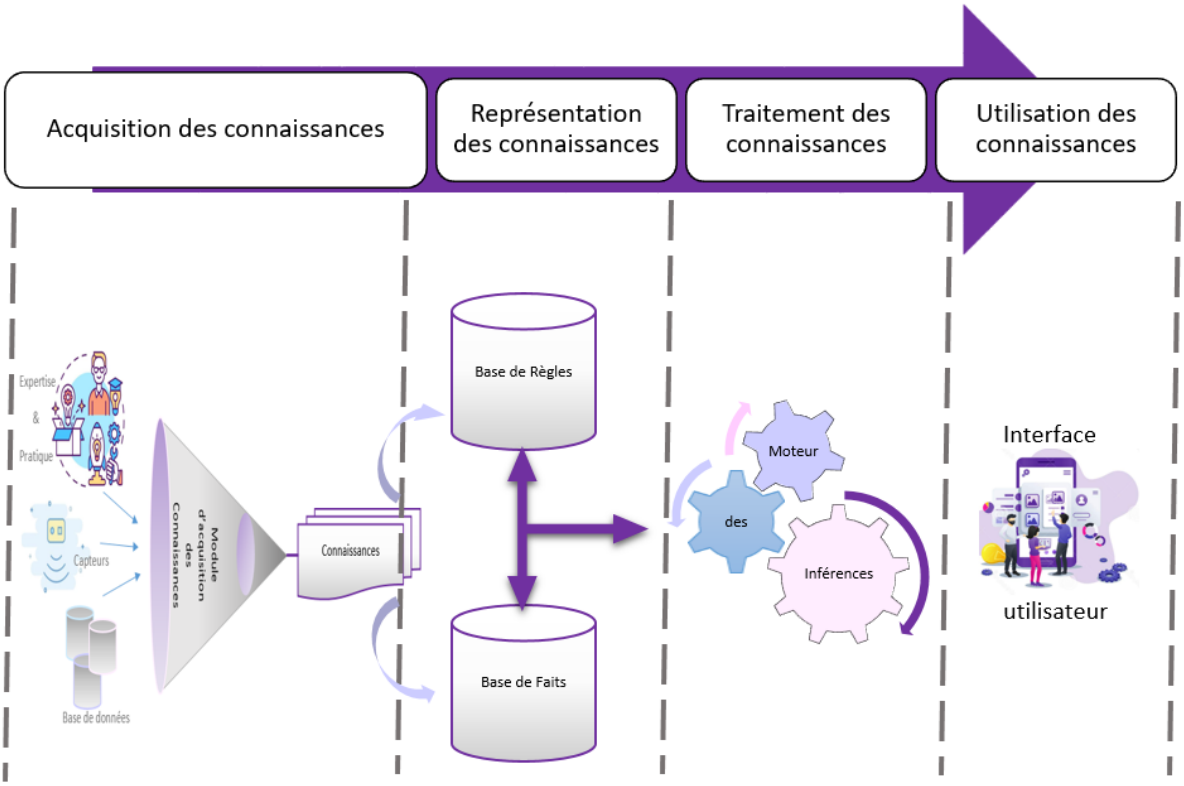


Figure 21 : Architecture d'un SBC

Cette architecture représente la base de tout SBC. Néanmoins, elle reste très basique et manque d'outils et techniques d'optimisation dans les cas complexes. En effet, plus le sujet est compliqué, plus les connaissances sont nombreuses et variées, plus la complexité des calculs augmente et devient exponentielle. A l'encontre de l'être humain qui devient plus efficace quand il augmente ses connaissances.

Pour remédier à ces problèmes, d'autres architectures peuvent être proposées en se basant sur des méta-connaissances superposées dans des couches supplémentaires permettant de superviser et contrôler les accès aux connaissances[9].

D'autres problématiques sont aussi omniprésentes lors de la mise en place d'un SBC, mise à part la définition de son architecture et ses mécanismes d'inférence, il faut également disposer des jeux d'essai très puissants et pertinents qui couvrent la plupart des cas possibles y compris les cas limites et extrêmes. De plus, comme tout autre système, Il faut savoir détecter et gérer toutes les lacunes du système : ses incompatibilités entre prémisses, les redondances, les contradictions, etc[40].

V- Apprentissage automatique : Machine Learning

La machine Learning c'est tout simplement donner à la machine la capacité d'apprendre et de s'autocorriger sans la programmer d'une manière explicite. Contrairement au programme informatique classique qui ne sait faire que des calculs, l'américain Thom Michel en 1998 a défini « un programme qui apprend »[62] comme suite :

> *Définition : On dit qu'un programme informatique **apprend** de l'expérience **E** à réaliser des taches d'une classe **T** et à une mesure de performance **P**, lorsque sa performance P pour faire une tache T s'améliore grâce à la nouvelle Expérience E.*

En effet, grâce aux algorithmes d'apprentissage, un ordinateur peut apprendre à conduire des voitures, faire de la reconnaissance vocale et faciale, reconnaitre un concert et plein d'autres applications. La plupart de ces algorithmes s'inspirent du comportement humain et en général, ils sont classés en trois familles d'apprentissage[42] :

- Apprentissage supervisé
- Apprentissage non supervisé
- Et apprentissage par renforcement

5.1 Apprentissage supervisé

Le principe le plus simple d'une machine Learning et celui d'apprendre à partir des exemples. Comme pour les êtres humains, le plus souvent nous apprenons à partir des exemples. Imaginons qu'on commence à apprendre la langue chinoise ! On va sûrement acheter un bouquin dans lequel on va trouver des exemples de traduction de la langue chinoise vers la nôtre, une autre solution est de faire payer un professeur particulier qui va superviser notre apprentissage en nous fournissant des exemples de traductions prêts et que nous devrons mémoriser. En machine Learning, la technique d'apprentissage la plus courante s'inspire directement de ce mode de fonctionnement. C'est ce qu'on appelle l'apprentissage supervisé. Dans ce type d'apprentissage, c'est bien nous qui jouons le rôle du professeur.

Dans l'apprentissage supervisé, on donne à la machine des exemples qu'elle doit étudier pour créer ce qu'on appelle un modèle. Ces exemples, en général, sont regroupés dans ce qu'on appelle un DataSet qui ressemble à un tableau de données[44].

Imaginons qu'on dispose d'un DataSet avec deux colonnes X et Y et on veut que la machine apprenne la relation qui relie X à Y : f(X) = Y

Notre tableau de données nous donnera un nuage de point à partir duquel la machine pourra apprendre un modèle linéaire qui ressemblera à (fig.22), ou bien peut être un modèle polynomial qui ressemblera à (fig.22)

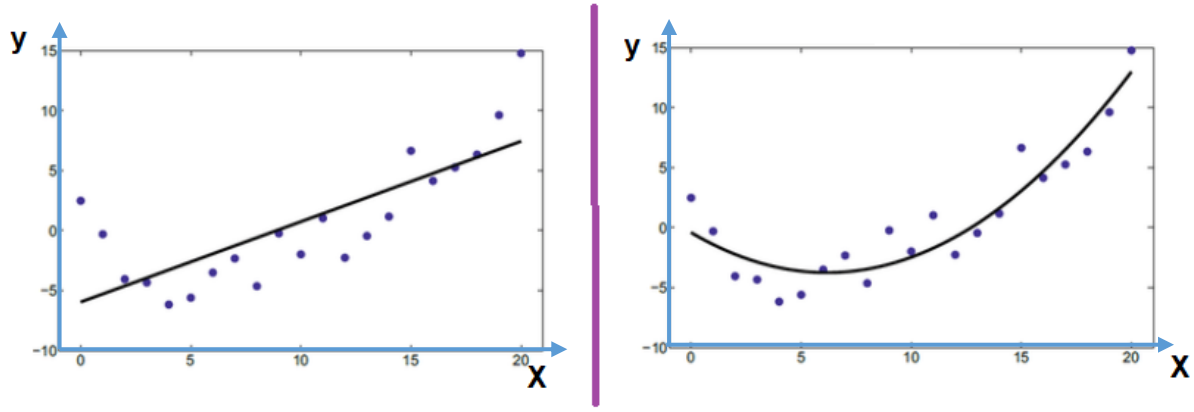


Figure 22 : Exemple de modèle d'apprentissage linéaire et polynomiale

La question qui se pose alors ; c'est comment choisir le meilleur modèle selon le type d'application qu'on cherche ?

Dans l'apprentissage supervisé on compte deux types de problèmes : d'un côté, on a les problèmes de régression dans lesquels on cherche à prédire la valeur d'une variable continue (avec une infinité de valeurs possibles) et d'un autre côté, on a des problèmes de classification dans lesquels on cherche à prédire la valeur d'une variable discrète (une variable qui prend un nombre de valeur fini)[63]. Par exemple, si on cherche à développer un filtre anti spam on est face à deux situations : Soit l'email est un spam et donc sa valeur est 1, soit-il ne s'agit pas d'un spam et sa valeur est 0 :

$$\text{Donc soit : } \boldsymbol{f(x)} = \begin{cases} \mathbf{1} & \boldsymbol{si\ x\ est\ un\ spam} \\ \mathbf{0} & \boldsymbol{sinon} \end{cases}$$

Lors de la construction de ce DataSet, on remarque que certains facteurs peuvent déterminer la nature de l'émail comme par exemple le nombre des liens, le nombre des fautes, etc.

On suppose que la représentation de ce DataSet est la suivante :

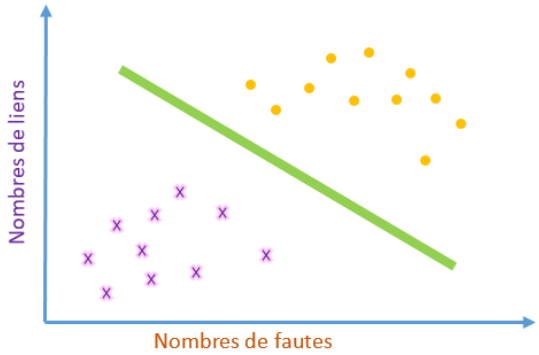


Figure 23 : Exemple d'un problème de classification des emails spam

On peut constater qu'on a plus de spams là où il y'a beaucoup de liens et de fautes d'orthographe, alors qu'il y'on n'a moins dans le cas contraire. A partir de cette représentation, dans les problèmes de classification, la machine crée ce qu'on appelle une frontière de décision qui divise notre espace de recherche entre les différentes classes de données.

Ce qu'il faut retenir, dans l'apprentissage supervisé que ça soit un problème de classification ou de régression, il y'a quatre notions fondamentales pour la mise en place de cette machine [63] :

5.1.1 Le Dataset

Il s'agit d'une structure de données qui contient en général deux types de variables :

- Les variables de type « Target » et ils représentent la valeur à prédire et la cible de notre apprentissage, comme le prix d'un ordinateur par exemple.
- Les variables de type « Features » et qui représentent les caractéristiques qui influencent sur la valeur et la qualité de notre prédiction, comme : la capacité mémoire, la capacité Disque dur, le type du processeur, etc.

De cette manière, le Dataset peut être représenté par une matrice d'ordre m*n dont n est le nombre de « features » et 'm'le nombre de lignes de ce Dataset :

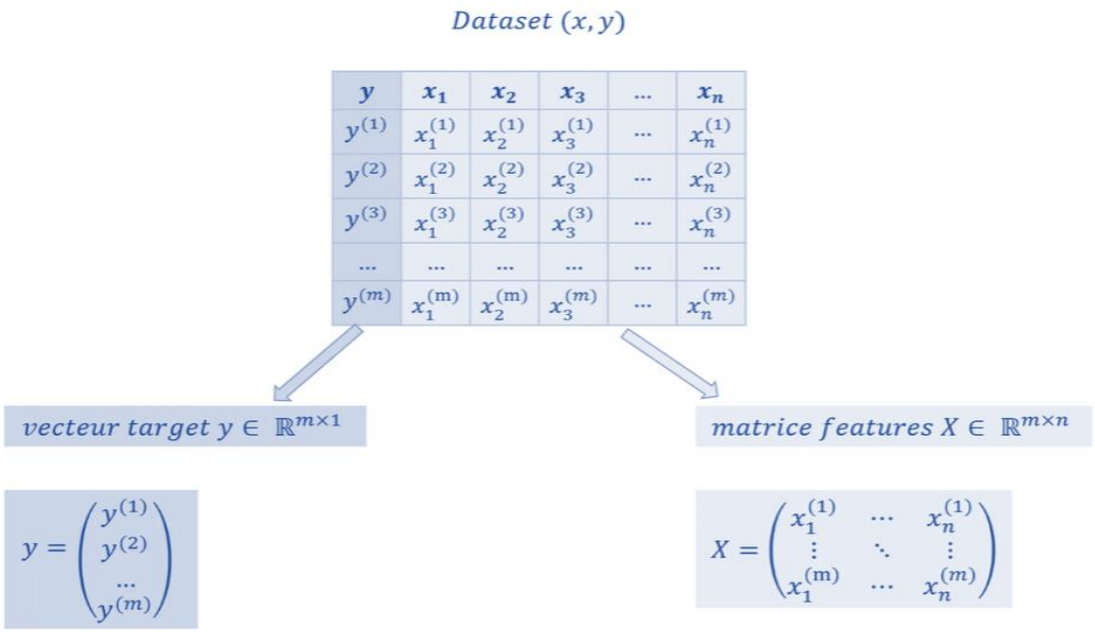


Figure 24 : Représentation matricielle d'un Dataset

5.1.2 Le Modèle d'apprentissage

A partir des Target du Dataset on peut schématiser une distribution de points en fonction des dimensions considérées. Cette visualisation permet de concrétiser cette distribution dans un modèle capable de distinguer les différentes catégories du problème à étudier. On suppose qu'à partir de notre Dataset on a pu constater qu'un modèle linéaire rentre mieux dans notre nuage de points et une simple fonction affine de type f(x) = ax + b peut bien représenter notre problème. Ce modèle a bien des coefficients tels que 'a' et 'b' qu'on appelle des paramètres. Donc, l'idée derrière la définition du modèle est de préciser à la machine Learning les paramètres à calculer pour apprendre les exemples à sa disposition[63].

5.1.3 La fonction du coût

Maintenant que la définition du modèle et de ses paramètres sont calculés, il faut s'assurer de l'efficacité de ce modèle. Pour se faire, la machine calcul une fonction du coût qui représente la sommation de toutes les erreurs entre le modèle et ses paramètres et les différents points du Dataset. Autrement dit, pour chaque point du Dataset, l'erreur

est la distance entre ce point et sa projection sur la courbe du modèle précédemment définis[40].

On suppose le modèle du graphe ci-dessous qui représente le Dataset des prix des appartements en fonction de leurs surfaces.

Soit le point *xi* du Dataset :

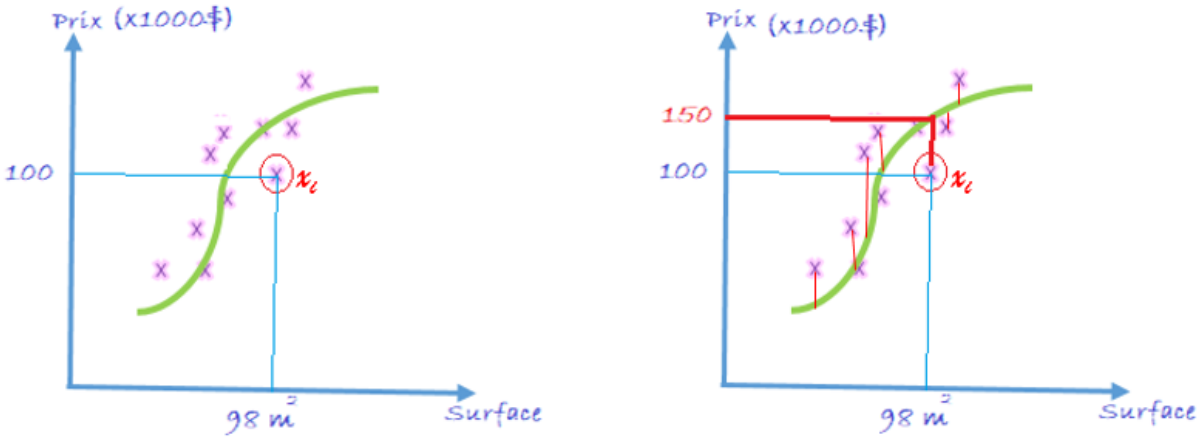


Figure 25: Exemple des erreurs calculées par la fonction coût d'un modèle d'apprentissage

A droite, le prix de l'appartement pour le point xi est de 100.000$. Cependant, en regardant la prédiction de notre modèle dans le graphe de gauche, on remarque que le prix est estimé à 150.000 $. Et donc, notre modèle réalise une erreur de 50.000$ de différence juste pour ce point. Ainsi, la fonction du coût somme l'ensemble des erreurs de prédictions estimées par notre modèle sur tous le Dataset.

5.1.4 L'algorithme de minimisation

Revenant à l'objectif principal de l'apprentissage supervisé et qui vise à trouver un bon modèle avec une meilleure configuration de ses paramètres. En fait, Un bon modèle est une représentation généralisée de l'ensemble des points du Dataset tout en garantissant de bonnes prédictions pour la plupart des cas possibles. De ce fait, pour atteindre cet objectif, il faut minimiser le taux des erreurs réalisées par le modèle. D'où la quatrième notion fondamentale de ce type d'apprentissage et qui consiste à définir un bon algorithme de minimisation pour optimiser le taux des erreurs du modèle[63].

L'algorithme de minimisation prend en entrée la fonction du modèle et doit pouvoir chercher les paramètres les plus optimaux qui minimisent la fonction coût.

On suppose l'exemple de classification des Emails entre Spam et non-Spam comme suite :

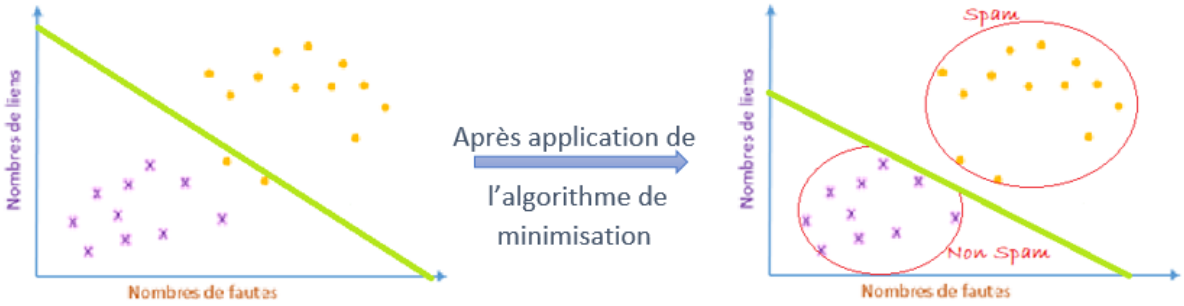


Figure 26: Exemple de classification avant et après l'application de l'algorithme de minimisation d'un apprentissage supervisé

Dans cette visualisation, la frontière de décision (ligne en vert) de notre premier modèle contient toujours des erreurs qui laissent échapper des emails « Spam » dans la classe des emails « Non-Spam ». Tandis que dans le deuxième graphe, on constate qu'après vérification des paramètres de notre modèle tous les points du Dataset entre dans l'ordre et la classification réalisée par cet apprentissage devient plus réaliste.

5.2 Apprentissage non-supervisé

Maintenant, imaginons cette fois-ci qu'au lieu d'avoir un bouquin ou un professeur qui supervise notre apprentissage, on se met directement dans un environnement qui ne parle que cette langue et on devrait se débrouiller pour apprendre cette langue. Donc on est face à des exemples qu'on arrive à interpréter petit à petit. Il s'agit de la deuxième famille des machines Learning et qu'on appelle l'apprentissage non-supervisé[63]. La différence entre les deux familles c'est que dans ce dernier on ne donne pas au programme les réponses des exemples fournis. Donc, le programme ne sait plus quoi faire et ne peut plus s'autoévaluer ni comparer ses résultats. Cependant, on lui demande d'apprendre à reconnaitre des structures dans des exemples non étiquetés. Autrement dit, détecter les différences et les ressemblances à partir de ces exemples. A la fin, on peut se servir des

informations que le programme a accumulées afin de les regrouper selon leurs ressemblances et leurs différences.

5.3 Apprentissage par Renforcement

Un autre type d'apprentissage est aussi valable mais qui ne nécessite plus d'exemples pour apprendre. En effet, Contrairement à l'apprentissage supervisé dans lequel on se sert des exemples avec des questions/réponses et l'apprentissage non-supervisé dans lequel on n'a que des questions, l'apprentissage par renforcement ne nécessite aucune donnée en entrée mais c'est à la machine elle-même de gérer sa propre expérience. Cette dernière joue le rôle d'un agent qu'on lui donne la liberté d'entreprendre des actions au sein d'un environnement. Suivant l'action entreprise, l'environnement modifie l'état de l'agent et donne une récompense positive ou négative associée à cet état. C'est ce qu'on appelle une expérience. Et donc l'objectif du programme et de développer une politique d'actions en maximisant le nombre de récompenses positives[57].

Pour illustrer cette différence, voici un schéma qui récapitule le principe d'apprentissage de chaque famille d'algorithmes :

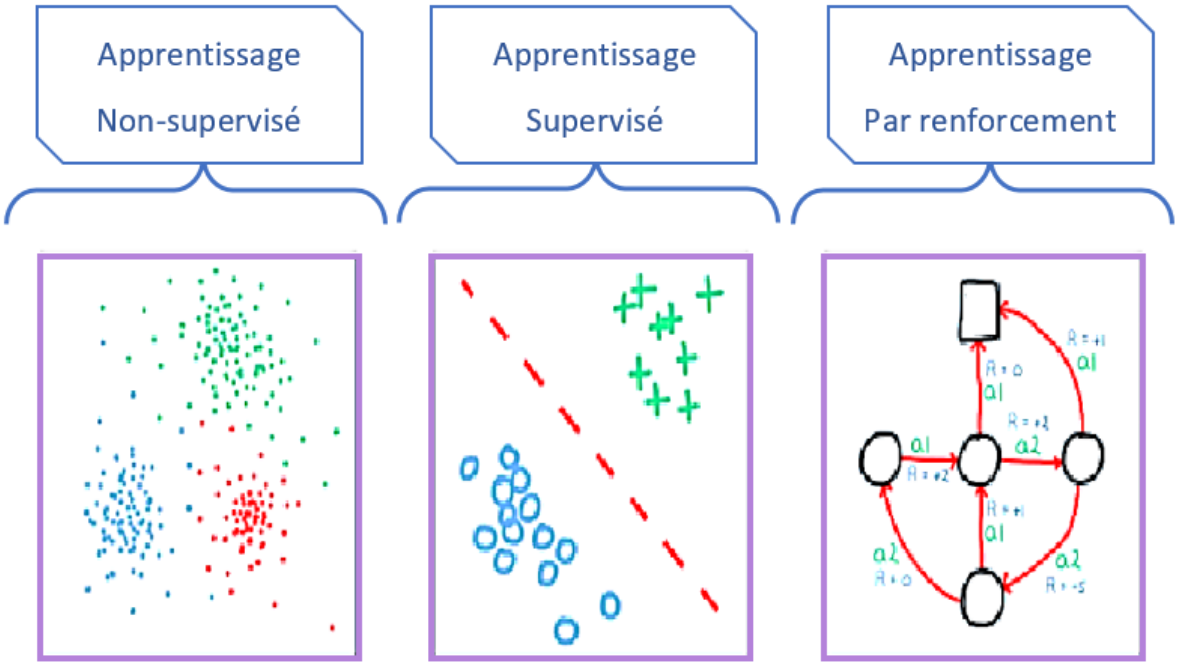


Figure 27 : Les trois familles d'algorithmes d'apprentissage par machine learning

Chapitre 3 |

Système d'Aide à la Décision Pour un Chiffrement Intelligent

Sommaire

I- Introduction

Dans ce chapitre nous allons discuter notre contribution qui vise à proposer une nouvelle stratégie de sécurité basée sur des techniques de l'intelligence artificielle afin de rendre cette finalité adaptable et change d'une plateforme à une autre tout en garantissant le meilleur niveau de confidentialité possible.

Comme présenté auparavant, sécuriser une information où une communication revient à garantir sa confidentialité, son intégrité, son authenticité et sa disponibilité. Dans ce travail, nous nous intéressons plus au volet confidentialité qui s'effectue en deux opérations : le chiffrement et le déchiffrement

Protocole de chiffrement :

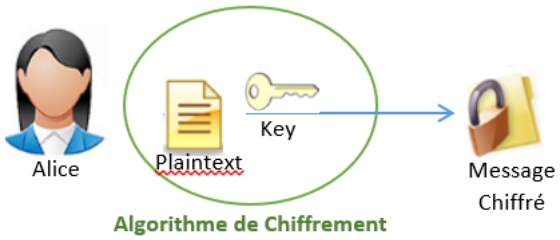


Figure 28 : Protocole de Chiffrement

Protocole de Déchiffrement :

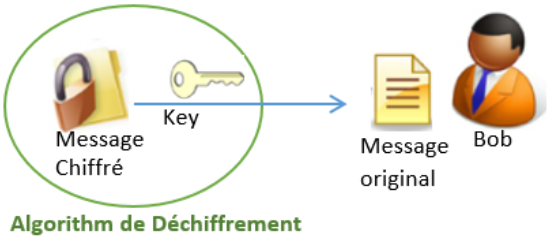


Figure 29 : Protocole de Déchiffrement

Notre objectif et d'effectuer ces deux opérations de la manière la plus efficace en prenant en compte les contraintes des environnements et des données sources à traiter. Pour ce, nous faisons appel à l'une des techniques de l'intelligence artificielle qui sont les systèmes à base de connaissance. Il s'agit d'un nouveau système qui essaie d'exploiter les connaissances et l'expertise dans le domaine des techniques de chiffrements afin de décider le meilleur dans

chaque situation. Dans ce qui suit, nous allons détailler le fonctionnement de cette approche comme suite :

Tout d'abord nous allons présenter notre motivation et les travaux réalisés dans ce même contexte. Ensuite nous discuterons la solution proposée, son principe de fonctionnement, et les différentes étapes la classification des données et de l'alimentation de la base de connaissance. Puis l'algorithme Block Nested Loop (BNL) utilisé au sein du moteur d'inférence de ce système. Finalement nous présenterons l'application et l'expérimentation de cette approche en se basant sur une dizaine d'algorithmes de chiffrements les plus connus.

II- Motivation

D'après la littérature et les différents travaux de recherche sur les algorithmes de chiffrements existants [64] et aussi notre propre expérience avec ces algorithmes [65] [66] [67], nous constatons que le niveau de sécurité de chaque système de chiffrement n'est pas stable et change d'un travail à un autre. Parfois, on opte pour AES comme meilleur algorithme parfois sur un autre [68]. En fait, le niveau de performance de chaque algorithme change en fonction du contexte et de l'environnement d'exécution.

Par exemple, dans [69] on trouve une étude comparative sur les performances des algorithmes de chiffrement symétrique DES, AES et Blowfish dans le contexte des images.

Plusieurs images ont été considérées dans cette étude et de différentes tailles en spécifiant à chaque fois leurs histogrammes. Ci-dessous un exemple représentant leurs résultats sur l'image de Lena.

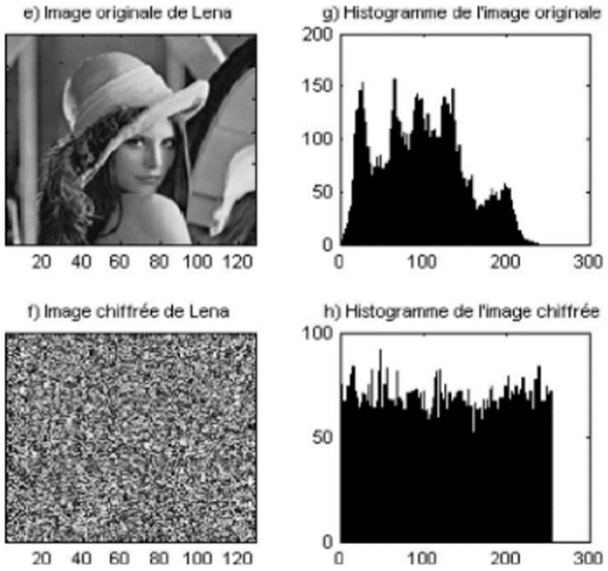


Figure 30 : Histogramme de chiffrement de l'image de Lina

Par rapport au temps de chiffrement et de déchiffrement de chacun de ces algorithmes, on se retrouve avec les diagrammes comparatifs suivants :

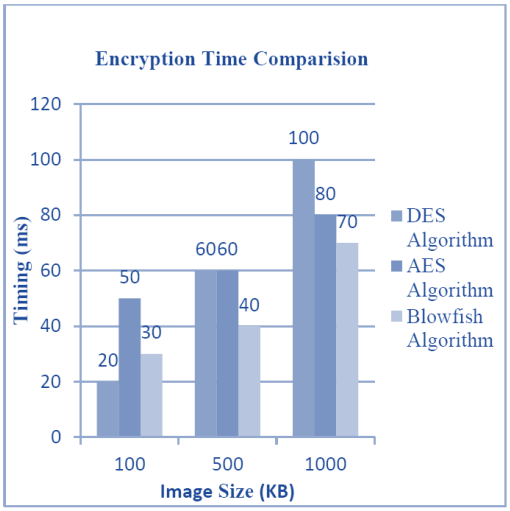


Figure 32 : Comparaison entre le temps de chiffrement des images de différentes tailles

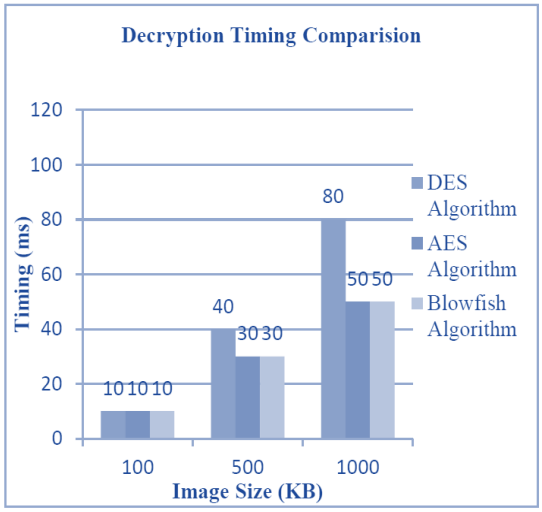


Figure 31: Comparaison entre le temps de déchiffrement des images de différentes tailles

Le constat qu'on peut déduire à partir de ces résultats est que :

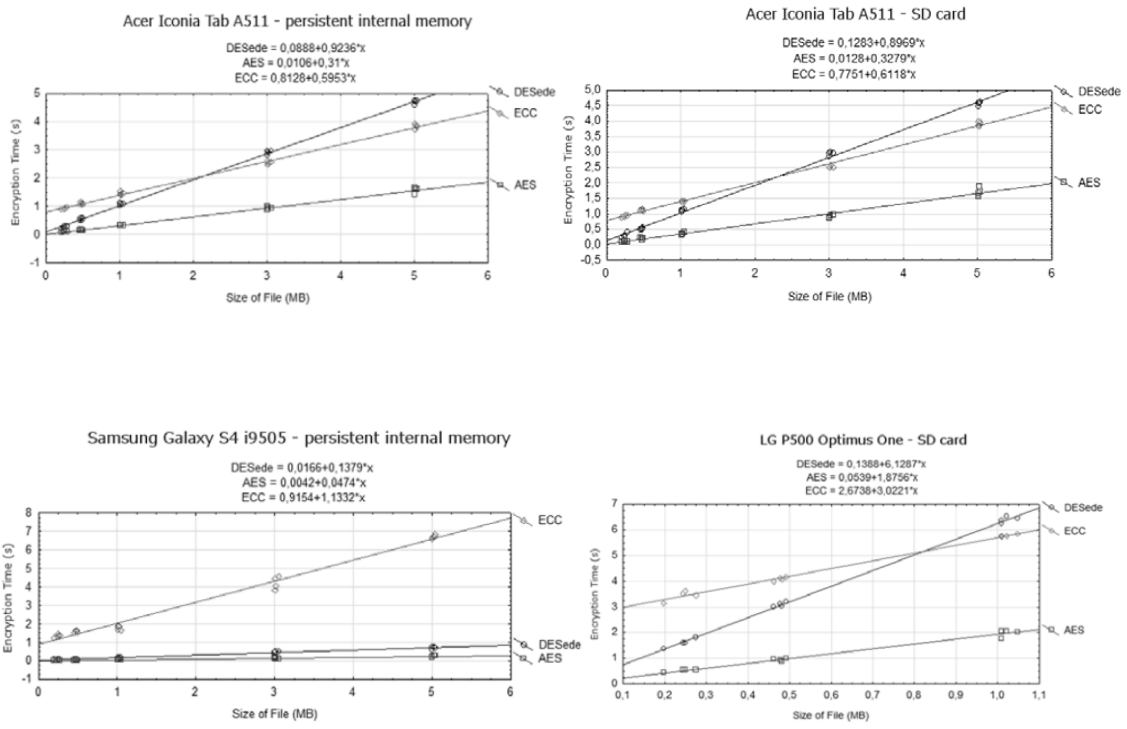
- DES est le plus performant pour les images de petite taille
- Blowfish est le meilleur par rapport au temps de chiffrement pour les images de grande taille
- Blowfish et AES performe de la même manière lors du déchiffrement

Indépendamment de la data, un autre travail [70] qui nous a aussi inspiré dans notre idée mais qui s'intéresse plus aux performances des algorithmes de chiffrement dans les plateformes Mobile. Ce travail met le focus sur les algorithmes Triple DES, AES et les méthodes basées sur l'arithmétique des courbes elliptiques(ECC). Puis, il cherche à comparer leurs performances dans différentes plateformes Android à savoir :

- ✓ Acer Iconia Tab A511
- ✓ Résultat de chiffrement sur Samsung Galaxy S4 i9505
- ✓ Résultat de chiffrement sur LG P500 Optimus One

Aussi, dans chaque environnement, il teste aussi les performances en fonction du type de stockage dans la carte mémoire (Carte SD, ou stockage interne).

Ci-dessous, un aperçu des résultats trouvés au cours de cette étude :



D'après ces diagrammes, les résultats de ce travail une vision différente du celle du premier comme :

- ✓ Sur Galaxy S4 i9505 , AES et DESede performe de la même manière, par contre les ECC sont déconseillés
- ✓ Sur LG P500, AES reste le meilleur

- ✓ Sur Acer Iconia Tab A511, ECC et AES performe de la même manière sur les petites tailles de fichiers en entrée, cependant, l'ECC est déconseillé sur les grands textes

Par conséquent, le fait de se baser sur un seul système de chiffrement dans toutes nos communications peut impacter le niveau de sécurité sinon les performances de notre système.

De plus, si on dispose d'un système qui permet d'envoyer des données aux différents destinataires avec différentes infrastructures, on se réfère à un expert dans le domaine de la sécurité, qui en fonction de son expertise et l'analyse du contexte de cette communication, décide le bon algorithme à utiliser.

Maintenant, si on suppose que notre système est hétérogène, interactif et capable de communiquer à chaque fois avec de nouveaux tiers mais qui n'ont pas forcément les mêmes exigences de sécurité que les autres partenaires, par exemple, ils sont installés sur un réseau Adhoc ou ils utilisent des données plus sensibles que les autres etc. Donc, ou bien on adopte la même politique de sécurité dans tous nos échanges chose qui pourra restreindre l'adhésion de plusieurs utilisateurs, sinon, évoluer le système à chaque intégration d'un nouveau client afin de prendre en compte leurs nouvelles exigences et sans pour autant impacter les anciens clients. Chose qui pourra être aussi couteuse.

III- Travaux antérieurs

Plusieurs travaux ont été réalisés dans ce contexte afin de renforcer et d'améliorer la sécurité des systèmes d'information.

En général, dans la littérature, on distingue entre deux orientations principales concernant les propositions qui combinent l'intelligence Artificielle et la sécurité informatique. On trouve des propositions d'aide à la décision dont l'objectif est de maitriser ou améliorer la politique de sécurité des entreprises. D'autres propositions par contre, visent à créer de nouveaux outils cryptographiques utilisant les techniques de l'intelligence artificielle, et là on trouve vraiment une carence par rapport à cet axe de recherche.

Par exemple, par rapport à la première orientation, différentes solutions ont été conçues pour dans le domaine du cyber-sécurité afin surmonter les failles et les anomalies des systèmes.

Dans [71] par exemple, on trouve une proposition d'un système de détection d'intrusion dans un environnement totalement non supervisé. Ce système se base sur l'algorithme d'Information mutuelle pour la sélection des features et sur le modèle Deep Learning PV-DM pour la lecture des paquets réseaux. Dans ce même domaine, d'autres solutions commencent déjà à prendre leurs places dans le marché comme la plate-forme AI2 [72] qui, à partir des lignes log, identifie les activités suspectes. AI2 applique des algorithmes de clustering sur les données d'entrée en utilisant les algorithmes d'apprentissage automatique non supervisés. Les résultats sont présentés ensuite aux experts en sécurité informatique, qui décident quels incidents sont de véritables attaques. Ce système est également capable de générer en continu de nouveaux modèles en quelques heures, ce qui peut considérablement améliorer la vitesse de sa capacité de détection des cyber-attaques.

Les articles [73] et [74] donnent un récapitulatif et un état de l'art détaillé sur d'autres systèmes de l'IA utilisés dans le domaine de la sécurité que ça soit au niveau de l'infrastructure, du réseau, du cloud, des terminaux, du mobile, des applications, de l'IoT, du Web et aussi au niveau de la Gestion des identités.

Toutes ces solutions viennent pour aider et faciliter la prédiction et/ou la détection des problèmes et des failles des systèmes. En revanche, dans l'autre catégorie, les solutions IA sont très innovantes et consistent à monter une intelligence artificielle afin d'assurer l'une des propriétés de la sécurité informatique. Pour assurer la confidentialité par exemple, Mart´ın Abadi and David G. Andersen de l'équipe Google [75] ont proposé une solution basée sur l'entrainement des Machines Learning pour le chiffrement et le déchiffrement des messages. En effet, les auteurs considèrent Alice et Bob comme des réseaux de neurones qui essaient de communiquer entre eux en toute confidentialité tout en empêchant le troisième réseau de neurones nommé Eve de déchiffrer leurs messages. Ce système ressemble aux principes des jeux électroniques. Dans les premières étapes, Alice et Bob commencent à s'entrainer pour découvrir un moyen de communication qui permet à Bob de décrypter au moins en partie du texte chiffré d'Alice, cependant, Eve peut commencer à casser ce code mais avec plus d'entrainement Alice et Bob raffine leurs transformations et exploitent plus la clé secrète en utilisant une formule quadratique dans le but de minimiser l'erreur de Eve dans son estimation du Chiffré(C) [75].

[65] Représente une autre manière pour concevoir un système de chiffrement en se basant sur les algorithmes évolutionnistes. SEC est la première variante de ce type d'algorithme qui réalise un encodage du texte sous forme de listes de positions, puis il applique un ensemble d'opérateurs génétiques (mutation et croisement) sur ces positions au niveau de chaque itération afin de reproduire de potentielles solutions. Finalement, et à travers une fonction d'évaluation bien définies, il évalue les individus et décide la solution la plus sûre.

Plusieurs extensions de ce système ont été développées afin d'augmenter le niveau de l'intelligence de ce système dans le choix de la solution la plus pertinente. Parfois en ajoutant des problèmes difficiles dans le processus de chiffrement comme le cas de [76], et dans d'autre cas en modifiant la fonction d'évaluation à appliquer sur des individus [77].

Notre proposition vient pour tirer avantage des deux axes de recherche qu'on vient de citer regroupant la sécurité et l'IA. Il s'agit d'une hybridation entre les deux approches, D'une part, il consiste en une solution d'aide à la décision et d'autre part, il cherche à assurer la confidentialité des données.

IV- Les méthodes de la Recherche du Skyline

4.1Le concept de dominance

Étant donné une relation de dominance dans un ensemble de données, une requête Skyline renvoie des objets qui ne peuvent être dominés par aucun autre objet[78].

Dans le cas d'un jeu de données composé d'objets multidimensionnels, un objet domine un autre objet s'il est également bon dans toutes les dimensions, et meilleur dans au moins une dimension[79].

Le calcul de Skyline était un problème algorithmique par nature, et toutes les données étaient supposées résider en mémoire.

Cependant, de nos jours, nous sommes confrontés à de grands ensembles de données qui sont stockés dans la mémoire secondaire. Ayant les données sur le disque, les algorithmes proposés pour le traitement des requêtes Skyline sont séparés en deux catégories[79] :

- Les algorithmes non basés sur des indexes

- Les algorithmes basés sur des indexes.

4.2 Algorithmes non basés sur des indexes

4.2.1 Block Nested Loop (BNL)

Un algorithme naïf pour calculer une requête Skyline consiste à comparer chaque objet avec tous les autres objets de l'ensemble de données à l'aide d'une boucle imbriquée. Cependant, la complexité quadratique O (n2) rend cet algorithme très inefficace (n est le nombre total d'objets dans l'ensemble de données) [78].

L'algorithme Block-Nested-Loop applique la même idée, mais utilise une fenêtre (bloc de mémoire à espace limité), qui contient un nombre limité d'objets de données. Tout objet candidat p est comparé aux objets de la fenêtre [78].

Trois cas peuvent se produire [78] :

1. p est dominé par un objet dans la fenêtre ! p est éliminé.
2. p domine un ou plusieurs objets dans la fenêtre ! Ces objets sont éliminés et p est placé dans la fenêtre.
3. Il n'y a aucun objet dans la fenêtre ! p est inséré directement dans la fenêtre. (Si la fenêtre est pleine, un fichier disque temporaire est utilisé pour contenir les objets candidats).

BNL peut nécessiter un grand nombre de passes jusqu'à ce que la ligne d'horizon complète soit calculée et éventuellement se terminer car à la fin de chaque passe, la taille du fichier temporaire sera réduite[80].

4.2.2 Diviser pour mieux régner (Divide & Conquer D&C)

L'algorithme D&C [81] calcule la valeur médiane d'une dimension, et divise l'espace en deux partitions P1, P2.

Ensuite, il calcule les Skylines S1, S2 de P1, P2, en divisant récursivement P1 et P2.

Le partitionnement récursif s'arrête lorsqu'il n'y a qu'un (ou quelques) objets. La ligne d'horizon globale est calculée en fusionnant S1 et S2 et en éliminant les objets de S2 dominés par n'importe quel objet de S1.

4.3Algorithmes basés sur les indexes

4.3.1 Index

L'index est un algorithme basé sur B-tree pour les données bidimensionnelles, où les données ont deux index ordonnés[82].

Par exemple, supposons qu'un arbre B et un arbre B + représentent chacun une dimension. L'algorithme en calcule un sur toute la ligne d'horizon en balayant les deux indices simultanément et s'arrête dès qu'un objet 'p' est trouvé dans les deux indices.

Tout objet qui n'a pas été inspecté dans les deux indices ne fait certainement pas partie de la ligne Skyline, car il est dominé par p. Par conséquent, les objets candidats sont ceux qui ont déjà été inspectés dans au moins un index. Ces objets sont conservés dans un ensemble séparé S (le sur-ensemble de la ligne Skyline).

4.3.2 Recherche du voisin le plus proche (Nearest Neighbor search NN)

Le premier algorithme Skyline basé sur un index spatial (comme R-tree), est l'algorithme Skyline NN[82]. Il est appelé NN en raison de sa pertinence pour la recherche du voisin le plus proche. Il identifie les objets Skyline par une recherche NN répétée, en utilisant une mesure de distance appropriée[83].

L'algorithme trouve de manière itérative l'objet le plus proche (NN) de l'origine dans une région donnée de l'espace sur la base d'une fonction de distance monotone, par ex. la distance euclidienne. Au cours du processus algorithmique, des régions entières dominées par un objet candidat sont ignorées et des régions qui ne peuvent pas être ignorées sont ajoutées à une liste de tâches pour un partitionnement supplémentaire de l'espace. Et ainsi de suite jusqu'à ce que la liste devienne vide et, ainsi, l'algorithme se termine[83].

4.3.3 Bitmap

L'algorithme Bitmap tel qu'il a été définit dans [82] est basé sur une représentation vectorielle de l'ensemble des données.

Pour décrire l'algorithme, soit p un point dans un espace à d dimensions représentées par un vecteur de m bits.

$$p = \{p._{d1}, ..., p._{dj}\},\ 1 \leq j \leq d,$$

A partir de ces m bits, chaque p.di est représenté par un nombre ki de tranche de bits. Chaque ki a autant de bits que le nombre de valeurs de coordonnées distinctes de tous les points de l'ensemble de données dans cette dimension.

Après avoir construit les tranches de bits, l'algorithme effectue 3 opérations entre 2 ensembles de parties de bits. Le premier ensemble contient les parties de bits Vx, Vy (un pour chaque dimension) où réside le dernier bit du point qui est égal à 1. Le deuxième ensemble contient les tranches de bits Vx + 1, Vy + 1. Dans le cas où les tranches de bits de l'étape précédente sont les dernières dans l'ordre, on utilise alors la tranche de bits zéro (tous les bits à zéro).

- La première opération A sera une opération ET entre Vx et Vy.
- La deuxième opération B sera une opération OU entre Vx+1 et Vy+1.
- La troisième opération C sera aussi une opération ET entre les résultats

Si le résultat de l'opération finale est nul, le point testé est un point Skyline.

V- Approche de chiffrement Intelligent

Notre contribution consiste à répondre à cette problématique et propose une nouvelle alternative afin d'industrialiser l'intelligence et l'expertise humaine dans le domaine de la sécurité. Il consiste à appeler les méthodes de l''intelligence artificielle pour garantir une meilleure confidentialité et l'adapter aux changements qui peuvent arriver dans les échanges entre les différentes entités du système interactif.

La figure suivante, illustre le schéma global de l'approche adoptée :

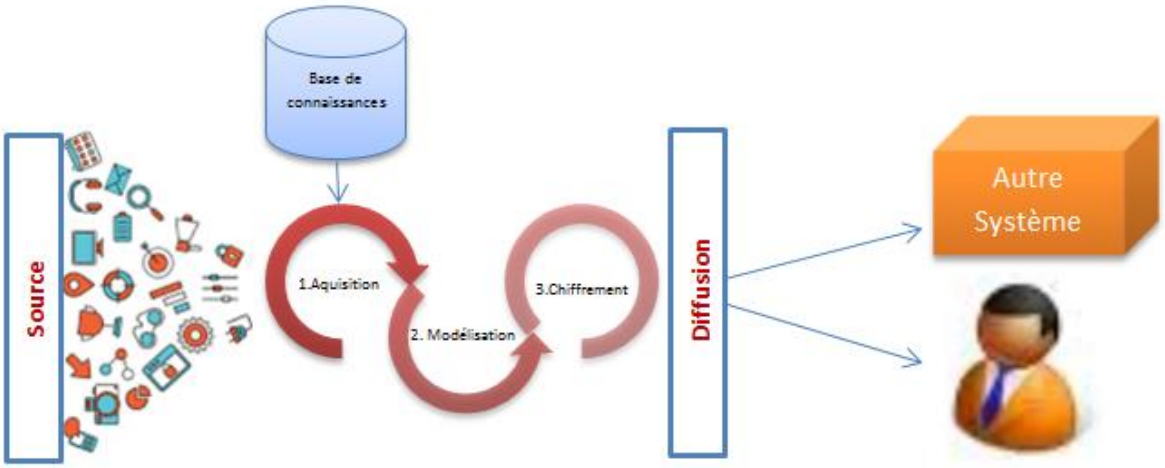


Figure 33 : le schéma global de l'approche adoptée

1.1SBC pour le Chiffrement

1.1.1 Architecture

La solution proposée repose sur l'une des premières méthodes de l'intelligence artificielle et qui a performée dans plusieurs domaines d'application. Il s'agit d'un Système à Base de Connaissance qu'on essaie de l'appliquer dans le contexte de chiffrement de données. Comme nous l'avons déjà présenté dans le chapitre 3 de ce mémoire, le Système à base de Connaissance se base sur quatre étapes fondamentales [84]dont chacune englobe un ensemble d'opérations pour le traitement et l'exploitation de connaissances. Ces étapes sont :

- Etape 1 : Acquisition des connaissances
- Etape 2 : Représentation des connaissances
- Etape 3 : Traitements des connaissances
- Etape 4 : Utilisation des connaissances

Au cours de ces étapes, on fait appel à trois composants techniques essentielles pour la modélisation et la conception de ce système[85] et qui sont :

- La base de connaissances
- La base de faits

- Le moteur d'inférence

Ces étapes ainsi que ces composants sont ainsi représentées dans le schéma suivant :

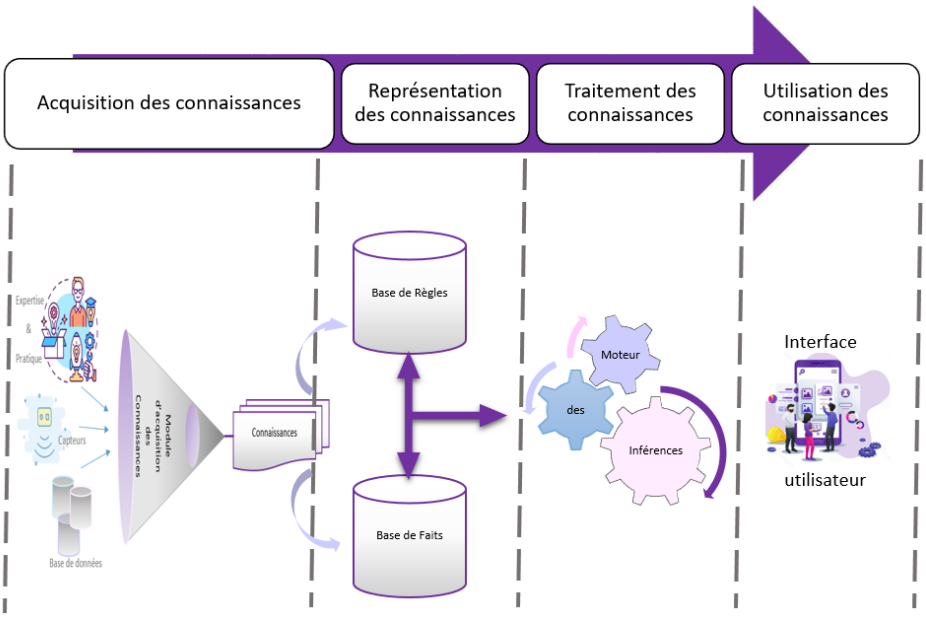


Figure 34 : Architecture du SBC pour le Chiffrement des données

1.1.2 Principe de fonctionnement

Le système de chiffrement intelligent peut être connecté à un ou plusieurs sources de données, comme il peut être connecté à un ou plusieurs destinataires, ce qui engendre une diversification des données à traiter et des exigences à prendre en considération dans le choix de l'algorithme de chiffrement à appliquer.

Le schéma suivant synthétise le principe de la classification et de la sélection du Skyline préféré

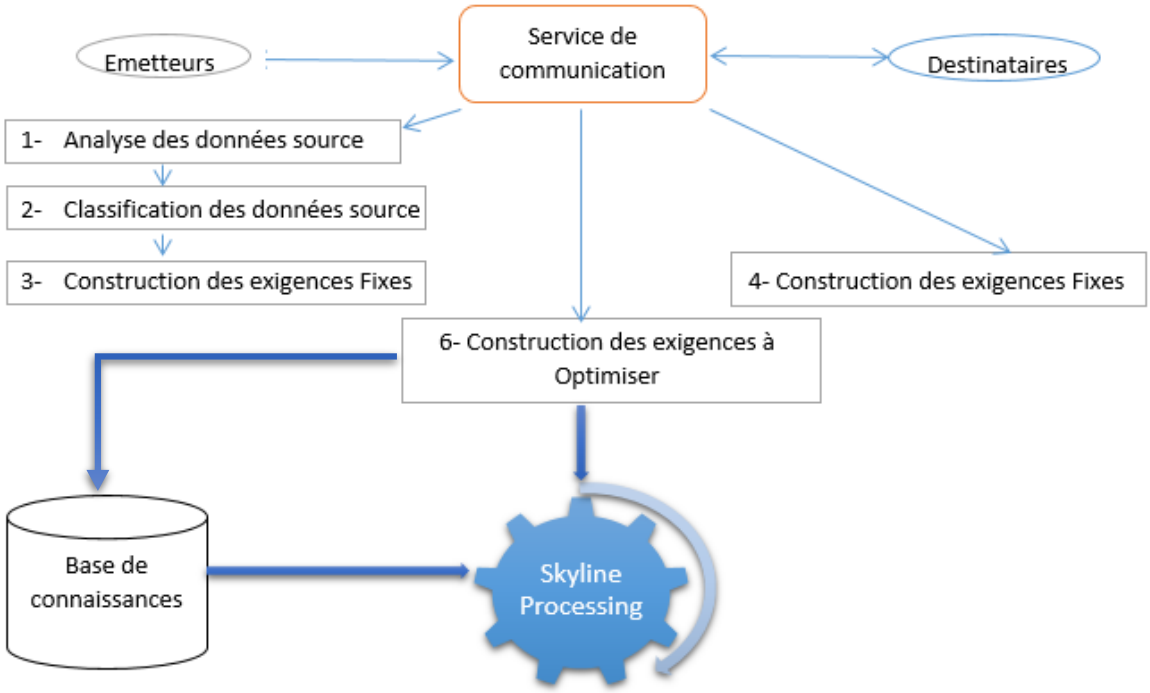


Figure 35 : synthétise le principe de la classification et de la sélection du Skyline préféré

1.1.2.1 Acquisition des connaissances

1.1.2.1.1 Analyse des données sources

Etant donné que les données à chiffrer proviennent de plusieurs sources, nous devons procéder en premier temps par une analyse de la structure et la nature de ces données afin d'en tirer les informations les plus importantes pour mener à bien cette procédure[9]. Les résultats issus de cette analyse seront ensuite centralisés dans une base de données tampon appelée Staging.

Dans cette contribution, la base de données Staging considérée comporte deux tables principales :

- Sta_analyse_date_source : pour l'analyse du message à chiffrer
- Sta_analyse_ciphering : pour l'analyse des messages chiffrés et l'enrichissement de la base de connaissance.

1.1.2.1.1.1 Alimentation de la table « Sta_analyse_date_source » :

L'usage de cette table vient après la première analyse réalisée sur les données sources. Il consiste à extraire les informations les plus intéressantes ayant un impact sur l'efficacité du système de chiffrement, à savoir :

- Le pourcentage des images par rapport à l'ensemble des données à chiffrer
- Le pourcentage de la vidéo à chiffrer par rapport à l'ensemble des données à chiffrer
- Le pourcentage du texte Littéral par rapport à l'ensemble des données à chiffrer
- La taille du message
- etc.

1.1.2.1.1.2 Alimentation de la table « Sta_analyse_ciphering» :

L'objectif de cette table est d'enrichir la base de connaissance par les résultats réels des chiffrements réalisés. C'est au moment de l'alimentation de cette table qu'on procède à la construction des connaissances nous permettant de servir dans la phase décisionnelle. Donc, on fait appel à cette table au niveau de deux périodes du cycle de vie de cette approche :

Dans la phase de Rodage : il s'agit du 1er lancement de cette stratégie afin de permettre l'alimentation de notre base de données décisionnelle

Dans la phase de l'Enrichissement : il s'agit de la dernière étape après chaque exécution de cette méthode en appliquant le chiffrement élu

Les connaissances qu'on souhaite conserver dans cette table sont :

- Temps d'exécution de chiffrement
- Temps d'exécution de déchiffrement
- Mémoire utilisée
- Entropie du message chiffré [86]

Ces informations représentent la base de calcul des indicateurs à optimiser pour chaque expérimentation.

1.1.2.2 Représentation des connaissances

1.1.2.2.1.1 Modélisation de la Base de connaissance

Dans cette étape, nous procédons à la consolidation, l'historisation et l'agrégation des données stockées dans les tables Staging vers la base de connaissances données qui prend la forme d'une Datawarehousse pointu sur les connaissances en liaison avec le sujet des chiffrements.

La modélisation de notre base suit le modèle en étoile. Elle contient, une seule table de Fait et un ensemble des tables de dimensions comme présenté dans le schéma ci-dessous :

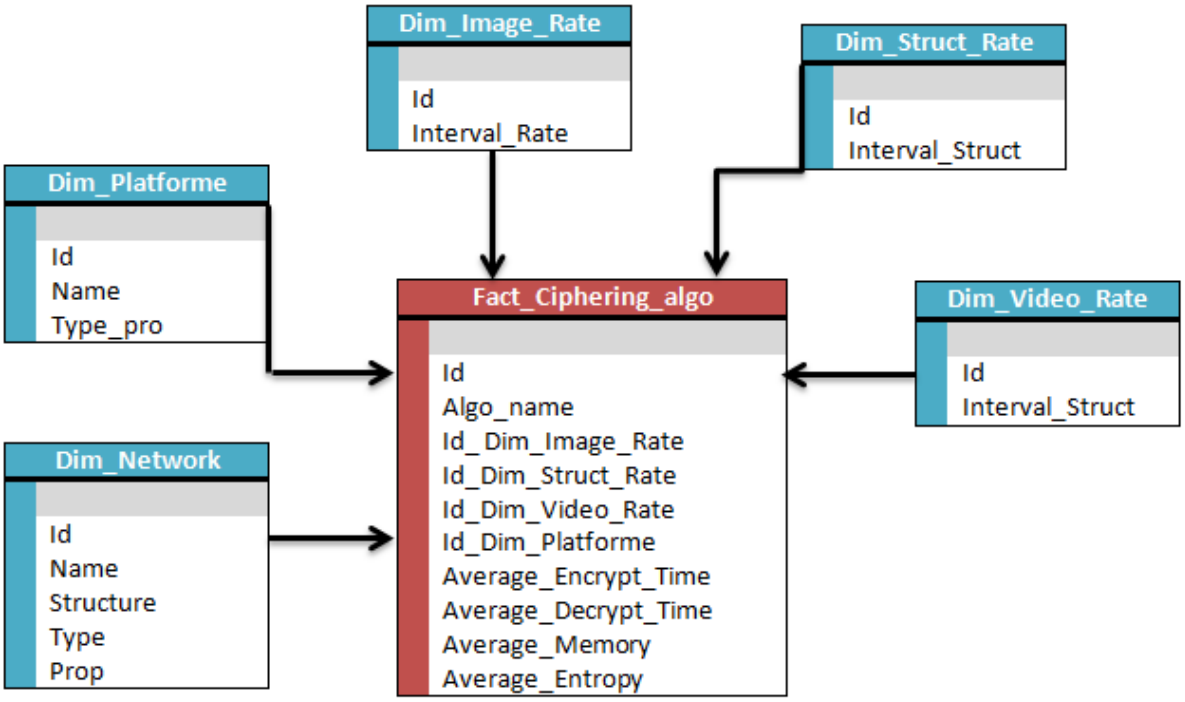


Figure 36 : Exemple de la modélisation de notre Base de Connaissance

1.1.2.2.1.2 Alimentation de la table de Fait « Fact_ciphering_algo» :

La table Fact_ciphering_algo située au centre de l'étoile représente la seule table des faits utilisée dans cette modélisation. Cette table contient un ensemble des mesures calculées pour chaque dimension et agrégées en utilisant la fonction « Moyen ».

1.1.2.2.1.3 Alimentation des dimensions

Les tables situées aux extrémités de l'étoile représentent les différentes dimensions utilisées dans notre système. On distingue alors entre deux types de dimensions :

Des dimensions de type « Rate » : ce sont les dimensions qui se terminent par le mot « Rate ». Ces dimensions sont alimentées une seule fois et contiennent les différents intervalles possibles pour les critères : Taux d'image, Taux de vidéo, Taux de texte littérale …

Des dimensions référentielles : ces dimensions stockent les différentes infrastructures et environnements d'exécutions utilisés pour chaque résultat renseigné et intégré dans cette base.

1.1.2.3 Traitements des connaissances

1.1.2.3.1 Construction des Exigences

Avant de passer à l'exploitation de la base de connaissance par la méthode BNL, nous devons en premier temps spécifier les exigences à considérer dans cette opération[78].

On distingue entre deux types d'exigences [78]:

- Des exigences Fixes : qui permettent de définir les caractéristiques fixes servant comme une condition dans la requête de sélection
- Des exigences à optimiser, ce sont les critères à utiliser dans les fonctions d'agrégation de cette même sélection.

1.1.2.3.1.1 Construction des exigences Fixes

Dans notre cas, les exigences fixes doivent être définies sur deux niveaux :

Exigences liées aux données source : il s'agit d'extraire les différents éléments qui caractérisent le message à envoyer. En fait, ça représente la même analyse effectuée précédemment sur les données sources sauf que dans cette partie, on souhaite plutôt construire les paramètres d'exécution de l'algorithme BNL.

Exigences liées aux différentes entités (sources, destinataires): Pour les entités qui souhaitent établir une communication confidentielle, il est aussi indispensable de connaitre les propriétés et les exigences de chacune d'eux, comme par exemple :

- L'infrastructure du réseau,

- Protocol de transport
- Disponibilité des licences
- Capacité des machines (CPU, RAM)
- Niveau de sécurité exigé

1.1.2.3.1.2 Construction des exigences à optimiser :

Dans le domaine décisionnel, ils s'agissent des mesures et des faits qu'on souhaite maximiser ou minimiser en fonctions de nos besoins.

Dans notre cas, nous pouvons considérer les contraintes suivantes :

- Le coût de chiffrement / déchiffrement
- La rapidité de chiffrement / déchiffrement
- La réputation de l'algorithme
- La qualité (taux de l'aléatoire /L'entropie)

1.1.2.3.2 Moteur d'inférence

Le moteur d'inférence dans notre cas, consiste à solliciter la base de connaissance et la table des Faits en utilisant l'algorithme BNL en se basant sur les différents critères préalablement spécifiés.

L'idée directrice de cet algorithme, extension de l'algorithme Nested-Loops, est de limiter le nombre d'E/S effectuées en chargeant en mémoire un ensemble de tuples candidats au Skyline appelé « fenêtre »[81]. De plus, un fichier temporaire stocké en mémoire secondaire dans lequel sont écrits tous les candidats non considérés faute de place. Ils seront traités lors d'une prochaine itération[87].

Trois cas de figure s'expliquent lors de la comparaison d'un tuplet avec ceux de la fenêtre:

Cas 1 : t est dominé par un tuple de la fenêtre

t est alors directement écarté et ne sera plus pris en compte pour le reste du calcul : il est dominé et ne fera donc jamais partie du Skyline.

Cas 2 : t domine un ou plusieurs tuples de la fenêtre

Les tuples de la fenêtre dominée par t sont écartés et ne seront plus pris en compte pour les itérations futures. t est inséré dans la fenêtre sans problème, puisqu'il y a au moins une place libre.

Cas 3 : t est incomparable avec l'ensemble des tuples de la fenêtre il doit être ajouté à la fenêtre mais il n'a éliminé aucun point sur son passage.

Dans ce qui suit, on considère :

f_i : critère fixe

d_i : critère à optimiser

t (f_i , d_i) : tuple de l'ensemble des critères à considérer

L : liste des solutions candidates (fenêtre)

S : liste des Skyline

1.1.2.3.2.1 Initialisation de la liste des candidats Skyline

Afin de gagner en efficacité et en rapidité, nous proposons de réduire l'espace des tuples à comparer par l'algorithme BNL. Comme dans [78] nous proposons alors d'introduire une étape préliminaire afin de filtrer sur la table Fact_ciphering_algo en utilisant les critères fixes comme suite :

```
L = Select * from Fact_ciphering_algo F
Inner join dim_image_rate dim1 on F.id_dim1 = dim1.id,
Inner join dim_video_rate dim2 on F.id_dim2 = dim2.id,
Inner join dim_struct_rate dim3 on F.id_dim3 = dim3.id,
…
 where  fi (>, < ou =)  dimi.interval_Rate
```

1.1.2.3.2.2 Application de l'algorithme BNL

L'application de l'algorithme BNL sera réalisée sur l'ensemble des Skyline définie dans la liste initialisée par les tuples ayant déjà vérifié les premiers critères fixes.

BNL algorithm
Start Input: L – the list of skyline candidate

```
        Skyline S = ∅
        IF L is empty
                Finish algorithm, returns S with no element
        P = first point in L
        FisrtSkylineInserted(p, S)
                FOR EACH point p in L
                        IF dominates (p, S)
                                Insert p into S
                                Remove all points in S that are dominated by p
                        ELSE IF !(dominates (p, S) || dominated(p, S))
                                Add p to S
                END FOR
// Return the skyline points
        RETURN S
End
```

Figure 37 : L'Algorithme BNL

1.2Expériences

Afin de pouvoir s'assurer du bon fonctionnement de cette stratégie, il faut en premier temps construire une base de connaissance prototype sur laquelle les algorithmes décisionnels vont se baser.

Il s'agit de lancer plusieurs expérimentations sur différents types de textes puis stocker les résultats de ces exécutions dans la base de données Stagging, pour finalement, les consolider et les intégrer dans notre Datawarehousse

1.2.1 Sta_analyse_date_source

Ci-dessous, un extrait des données renseignées dans la table Sta_analyse_date_source pour les dix premiers inputs considérés :

Id	taux d'image	taux de vidéo	taux de littérature	Autres	Input Size (Kb)
Input 1	0%	0%	80%	20%	30,217
Input 2	0%	0%	100%	0%	46,034
Input 3	0%	0%	82%	18%	56,002
Input 4	6%	0%	17%	77%	69,131
Input 5	10%	0%	27%	63%	78,245

Input 6	12%	0%	42%	46%	145,652
Input 7	0%	0%	68%	32%	168,339
Input 8	34%	0%	52%	6%	175,662
Input 9	0%	100%	0%	0%	294,851
Input 10	28%	0%	53%	19%	454,376
…	…	…	…		…

Tableau 3 : *extrait des données renseignées dans la table Sta_analyse_date_source pour les dix premier inputs*

1.2.2 Application des chiffrements

Pour chaque input, nous lançons le chiffrement avec les différents algorithmes qu'on souhaite intégrer dans notre système. Dans notre cas, nous avons opté pour une dizaine d'algorithmes les plus connus, à savoir :

- Triple DES
- IDEA
- AES
- Blowfish
- SEC et ses variantes (voir Annexe)
- DES

Dans cette table nous intégrons les résultats de chaque exécution comme suite (exemple des 10 premiers inputs de la base) :

Algorithm	Input_id	Run time (ms)	Memory used (KB)	Entropy
3DES	Input 1	61	18,53	2,9806
3DES	Input 2	112	18,3	2,9458
3DES	Input 3	147	18,88	2,8277
3DES	Input 4	191	18,9	2,9496
3DES	Input 5	232	18,01	2,9677
3DES	Input 6	450	18,32	2,9148
3DES	Input 7	532	18,63	3,0383
3DES	Input 8	558	19,84	2,9277
3DES	Input 9	644	19,65	2,8571
3DES	Input 10	788	20,04	2,8268
AES	Input 1	28	12,63	3,87314

AES	Input 2	63	12,4	3,83834
AES	Input 3	83	12,08	3,72024
AES	Input 4	114	12,91	3,84214
AES	Input 5	132	12,02	3,86024
AES	Input 6	274	12,33	3,80734
AES	Input 7	313	12,64	3,93084
AES	Input 8	344	13,85	3,82024
AES	Input 9	367	13,66	3,74964
AES	Input 10	449	14,05	3,71934
Blowfish	Input 1	8	7,21	3,97181
Blowfish	Input 2	24	6,89	3,93701
Blowfish	Input 3	35	7,38	3,81891
Blowfish	Input 4	47	8,21	3,94081
Blowfish	Input 5	56	7,32	3,95891
Blowfish	Input 6	125	7,63	3,90601
Blowfish	Input 7	147	7,94	4,02951
Blowfish	Input 8	151	9,15	3,91891
Blowfish	Input 9	165	8,96	3,84831
Blowfish	Input 10	208	9,35	3,81801
ASEC	Input 1	56	16,03	3,1287
ASEC	Input 2	107	15,8	3,0939
ASEC	Input 3	146	16,38	2,9758
ASEC	Input 4	187	17,3	3,0977
ASEC	Input 5	238	16,5	3,1158
ASEC	Input 6	451	16,9	3,0629
ASEC	Input 7	527	17,3	3,1864
ASEC	Input 8	556	18,6	3,0758
ASEC	Input 9	637	17,6	3,0052
ASEC	Input 10	783	17,99	2,9749

Tableau 4 : les résultats de chaque exécution

1.2.3 Application de l'algorithme BNL

Une fois la base de connaissance soit suffisamment rodée et alimenté avec les différents cas possibles, notre système sera capable de décider l'algorithme de chiffrement le plus efficace en fonction de l'analyse de la source de données et de nos exigences en termes de sécurité et de la confidentialité souhaitée.

Dans cette expérience, nous supposons que notre source de données capte des informations non littérales, dont le contenu est mixte entre 68% des images sensibles et 32% du texte non littérale.

En Revanche, nous souhaitons avoir un niveau de sécurité très efficace avec une rapidité dans la phase de chiffrement et de déchiffrement

Ce besoin peut être interprété comme suite :

Critère	Niveau
unstructured text	30%
Image	60%
Entropy	à maximiser
Decryption Run Time	à minimiser
Encryption Run Time	à minimiser

L'application de l'algorithme BNL a donné les résultats suivants :

	Runtime Ciphering (ms)	Runtime Deciphering (ms)	Memory used (KB)	Entropy
Blowfish	57	20	8,21	3,94081
AES	117	32	17,3	3,0977
ASEC	184	12	12,91	3,84214

Tableau 5 : Résultat de l'application de l'algorithme BNL

A cette étape on peut déjà décider l'algorithme de chiffrement le plus adapté pour notre input. Il s'agit bien de l'algorithme Blowfish puisque il domine dans 3 critères (case en vert) et atteint une bonne valeur dans le temps de déchiffrement aussi (case en orange).

1.3 Etude Comparative

En appliquant les trois systèmes élus sur notre input, on se retrouve avec les résultats suivants :

	Run time Ciphering (ms)		Run time Deciphering (ms)		Memory used (KB)		Entropy	
	Skyline	Valeur Réelle	Skyline	Valeur Réelle	Skyline	Valeur Réelle	Skyline	Valeur Réelle
Blowfish	57	61	20	28	8,21	10,03	3,94081	2,3402
AES	117	109	32	53	12,3	12,76	3,0977	2,3745
ASEC	184	194	12	15	17,91	19,6	3,84214	2,4142

Tableau 6 : Résultat de l'application des trois systèmes élus sur notre input

D'après cette comparaison entre les valeurs des indicateurs trouvés comme point skyline de notre problématique de chiffrement, et les valeurs réelles données par l'application des chiffrements en question, nous constatons que ces deux valeurs qui correspondent à l'un des quatre critères considérés dans cette expérience, sont très proches les uns des autres. Cette similitude peut être encore plus visible avec la représentation graphique présentée ci-dessous au niveau de chaque dimension d'analyse :

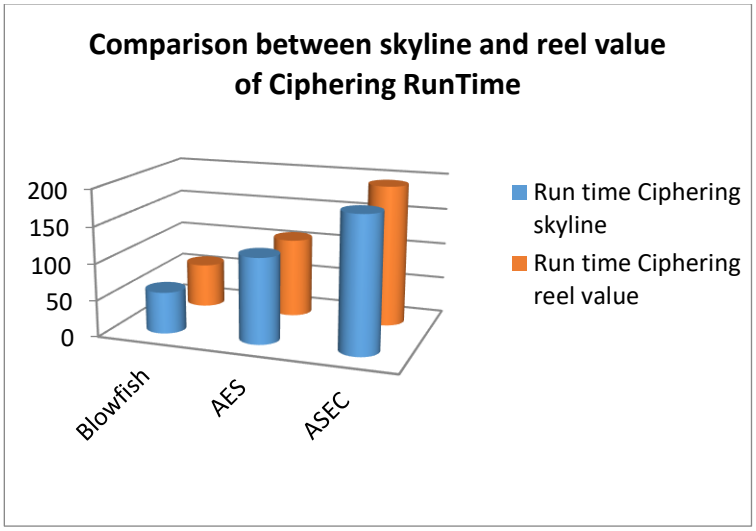


Figure 38 : Comparison between skyline and reel value of Ciphering RunTime

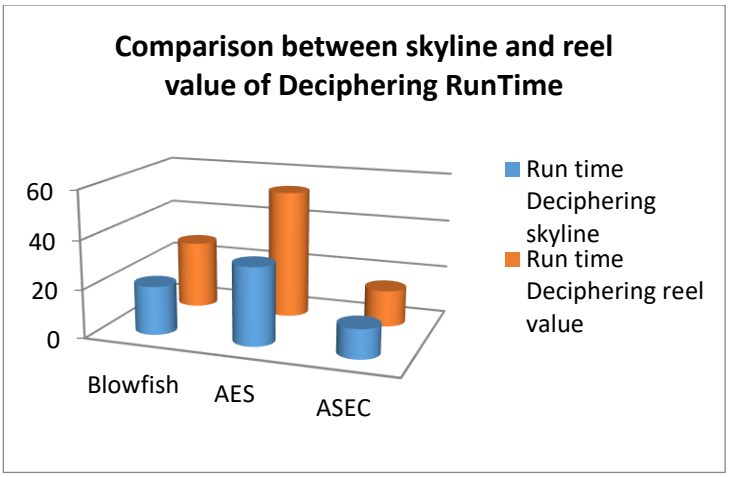


Figure 39 : Comparison between skyline and reel value of Deciphering RunTime

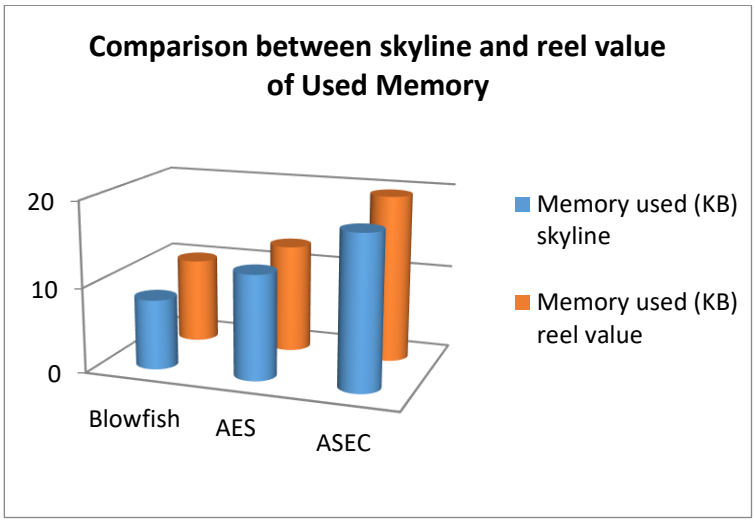


Figure 40 : Comparaison entre skyline et la valeur de la mémoire réel utilisée

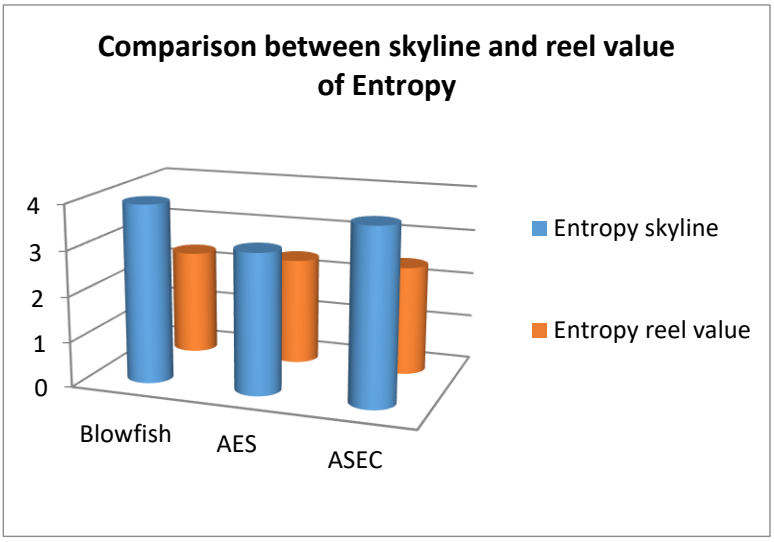


Figure 41 : Comparison between skyline and reel value of Entropy

1.4Discussion

Comme nous pouvons constater à travers les différentes comparaisons, les résultats de cette expérience ont confirmé le bon choix des algorithmes pour cette source de données. De plus, nous avons pu tirer un autre avantage très important, il s'agit de prédire une estimation préalable des valeurs des indicateurs pour chaque algorithme étant donné que les résultats trouvés sont très proches avec les résultats réels d'exécution.

VI- Conclusion

Grâce à ce travail, nous avons réalisé une nouvelle politique de chiffrement intelligent basée sur l'algorithme BNL. Cette approche a permis d'adapter le choix du système de chiffrement en tenant compte de l'ensemble du contexte dans lequel l'échange de ces données sera réalisé. Nous avons pu définir une stratégie basée sur six étapes principales depuis l'analyse des sources de données, la classification de ces données dans une architecture BI, les critères de sélection de construction, l'application de l'algorithme BNL, l'application du chiffrement élu, et enfin l'évaluation et l'enrichissement de la base de connaissances. Les résultats trouvés étaient généralement très proches des valeurs du Skyline élu, d'où l'importance de cette nouvelle méthode de prédiction d'indicateurs concrets, choses qui n'étaient pas possibles avant de se référer à une expertise humaine.

Chapitre 4 |

Étude des outils de la recherche Multidimensionnelle pour le Chiffrement Intelligent

Sommaire

I- Introduction

Dans la contribution décrite dans le chapitre 3 de ce mémoire, nous avons présenté une nouvelle approche de sécurité intelligente adaptable avec chaque échange d'information et permettant la communication avec plusieurs entités indépendamment de leur protocole de sécurité. Dans ce travail nous cherchons à augmenter encore plus le niveau de sécurité de ce système intelligent. L'apport original de ce travail réside dans l'ajout d'une nouvelle couche dans le système permettant la configuration de l'algorithme Skyline le plus approprié aux critères de recherches spécifiés. A travers ce chapitre, nous analysons, par une étude expérimentale, l'impact de chaque critère/dimension sur les performances de l'algorithme Skyline. Ensuite, nous déterminons l'ensemble des paramètres à considérer dans le choix du Skyline. Et finalement nous présenterons la nouvelle architecture de ce système.

II- Motivation

La solution, présentée dans le chapitre précédent, est un système décisionnel ayant pour but d'assurer la confidentialité des données de la manière la plus adéquate. En effet, avant de procéder à un chiffrement des données, dans cette solution nous avons proposé de considérer un ensemble de critères afin de décider le meilleur algorithme qui répond au maximum aux exigences de sécurité qu'on cherche à garantir.

Cette nouvelle méthode de chiffrement commence par une analyse des éléments qui impactent la sécurité de la communication au niveau de :

- l'environnement source,
- le canal de transmission,
- l'environnement de destination
- les types des données à transmettre,
- la génération des clés sinon la possession des clés.

Chacune de ces parties influence considérablement la sécurité des informations à échanger. Par conséquent, il faut les analyser et y extraire les critères les plus pertinents permettant

d'atteindre notre objectif. Une fois cette analyse soit faite, il faut passer à la classification et le stockage de toutes ces connaissances dans une architecture business intelligent. Il s'agit d'une base de connaissance qui regroupe un nombre important d'expériences couvrant les différents cas possible. Nous lançons ensuite via le moteur d'inférence, l'algorithme de la recherche Skyline BNL sur la base de connaissance préalablement rodée. Finalement, en fonction de l'algorithme de chiffrement choisi, nous procédons à l'application de ce chiffrement sur notre message et l'envoi de l'information chiffrée. Puis nous consolidons cette expérience en calculant les valeurs réelles de ce chiffrement en termes d'indicateurs de sécurité possible. Ce retour d'expérience que nous avons pu achever est stocké dans la base de connaissance.

Néanmoins, l'application réalisée dans le cadre de cette contribution a révélé quelques incohérences dans les points Skyline retournés ainsi que des lenteurs dans le cas de plusieurs critères de recherche. La diagnostique de cette problématique nous conduit à étudier d'autres algorithmes de la recherche Skyline afin d'optimiser les performances de notre application.

D'où vient l'idée de cette présente contribution, et dans laquelle nous proposons d'ajouter une couche de paramétrage supplémentaire permettant le choix de l'algorithme Skyline en fonction des critères en entrée. Plusieurs questions commencent à se poser comme: Quels sont les choix des algorithmes Skyline possibles ? Quels sont les attributs de ce paramétrage ? Comment et Quand ce paramétrage doit être appliqué ?

III- Méthodologie

3.1 Quels algorithmes Skyline ?

Dans la littérature, nous trouvons une grande variété d'algorithmes Skyline dont chacun répond à un besoin spécifique et performe mieux dans son contexte d'application [88] [89] [90]. Néanmoins, la plupart de ces algorithmes représentent une version améliorée de l'un des algorithmes de base présentés en détail dans le chapitre 2 [91]. Pour cette étude, nous optons en premier temps de rester sur cette catégorie basique d'algorithmes du fait qu'ils sont plus répandus, plus simples et bien maitrisés[91].

Pour cela, nous allons considérer dans la suite, les algorithmes Skyline suivants :

- Boucle imbriquée par blocs (BNL)
- Diviser & Conquérir (D&C)
- Bitmap
- Index
- NN

3.2 L'application SkylineApp

3.2.1 Description

"Skyline App" est une application Web qui renvoie les objets qui ne peuvent être dominés par aucun autre objet dans une base de données de n dimensions en utilisant 5 algorithmes (BNL, D&C, Index, Bitmap et NN) tout en choisissant les critères et les dimensions souhaités.

L'application des algorithmes Skyline a été réalisée par le langage de programmation Java via le framework Spring Boot pour le développement du Back End et Angular pour la partie Front office.

A travers cette application, les utilisateurs jouissent du droit de :

- Saisir le nombre de dimensions.
- Définir les dimensions.
- Choisir le type d'opération à appliquer sur chaque dimension.
- Choisir l'algorithme préféré pour le calcul des points Skyline.

3.2.2 Partie Backend

Le Backend de l'application Skyline a été développé sur la base des framework de développement Java les plus utilisé actuellement, et qui sont :

Le framework Spring: afin de bénéficier des fonctionnalités qu'apporte ce framework comme Spring Security, SpringMVC, Spring Batch, Spring Ioc, Spring Data, etc. Ces frameworks ont pour objectif de faciliter la tâche aux développeurs.

Le framework Spring Boot : il s'agit d'un sous projet de Spring qui vise à rendre l'utilisation du Spring plus facile en élimant plusieurs étapes de configuration. L'objectif de Spring Boot est de permettre aux développeurs de se concentrer sur des tâches techniques et non des tâches de configurations, de déploiements, etc. et par conséquent, il permet un gain en temps et en productivité.

JDBC (Java DataBase Connectivity) : il s'agit de l'API Java permettant de s'interfacer aux bases de données relationnelles et exécuter toutes les opérations SQL sur notre base de données.

JPA (Java Persistance API) : c'est une API Java pour manipuler des objets Java avec une grande simplicité. Cette API permet une abstraction des requêtes SQL pour réaliser des opérations de base. On peut utiliser des requêtes JPQL (Java Persistence Query Language) pour réaliser des requêtes SELECT plus avancées. Il est aussi possible d'utiliser les requêtes SQL natives en dernier recours mais JPA perd alors son intérêt.

Spring Data : c'est un concept Spring qui reprend JPA et l'intègre au framework Spring.

Le framework Spring MVC : Ce framework structure la réalisation de notre application dans une architecture de type MVC (Model-View-Controller). Il met à notre disposition un ensemble de composants performant pour le développement des applications Web d'une manière plus flexible et faiblement couplée. Le modèle MVC permet de séparer les différentes parties d'une application web à savoir la gestion de requêtes d'entrée envoyées par le client, la logique métier et la logique UI (affichage résultats en réponses aux requêtes) tout en assurant un couplage moins fort (Lazy) entre les différentes classes de l'application.

3.2.2.1 Architecture du Backend du SkylineApp

L'application SkylineApp a été développée en utilisant une conception MVC qu'on peut schématiser comme suite :

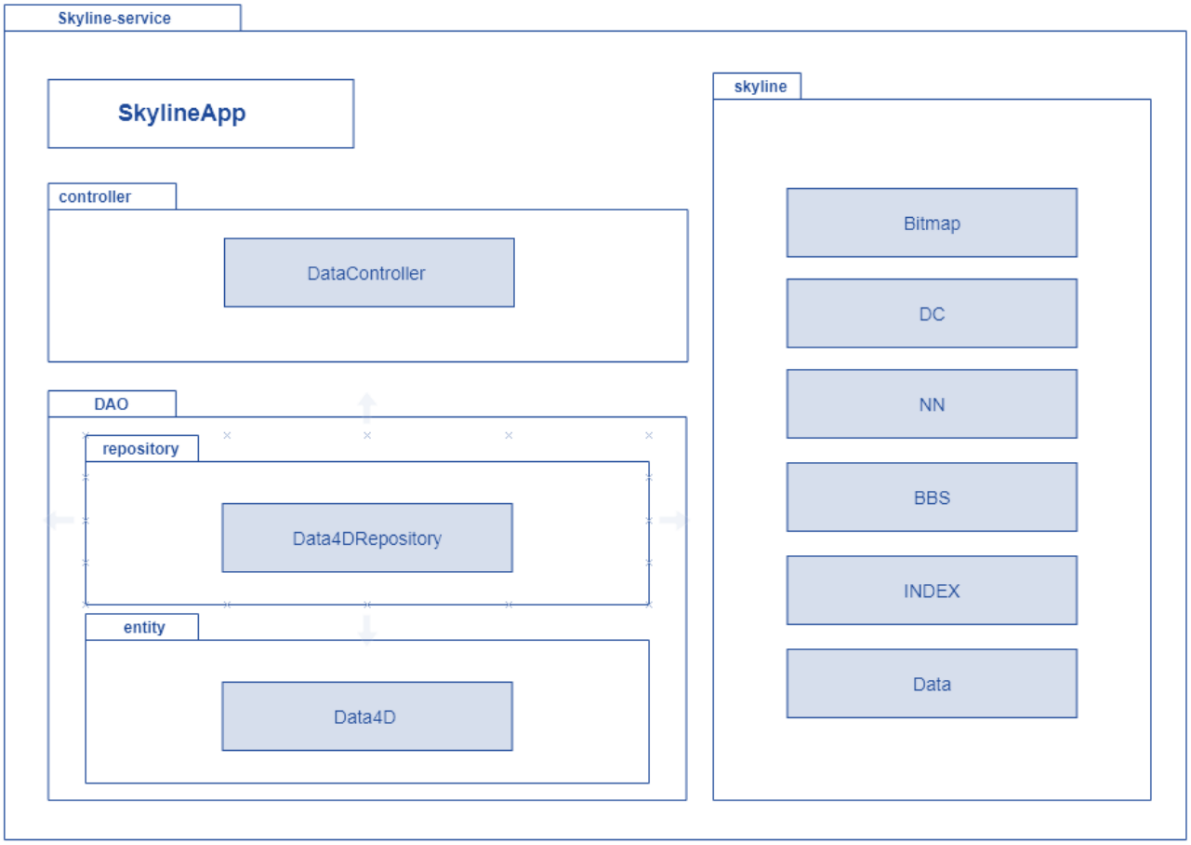


Figure 42 : Architecture technique de l'application SkylineApp

3.2.2.2 Liste des Packages Backend

Comme montré dans le schéma Fig.39, l'architecture adoptée dans notre application permet la réalisation de cinq packages qu'on peut récapituler dans le tableau suivant :

Tableau 7 : Description des packages de l'application SkylineApp

Package	Description
Controller	Package regroupant les différents contrôleurs du service.
DAO	Package regroupant les différentes définitions des classes relatives à l'accès à la base de données. Entre autres, on peut y retrouver les packages repository, entity.
Repository	Package regroupant les différentes définitions des classes permettant d'interagir avec la base de données via le contexte JPA.

Entity	Package regroupant les différentes définitions des classes représentant une instance de données au sein de la base de données.
Skyline	Package regroupant les différents algorithmes utilisés.

3.2.2.3 Liste des Classes Backend

L'application SkylineApp fait appel à cinq classes java principales. Le tableau ci-dessous regroupe les classes Java réalisées au niveau de chaque package ainsi que leurs descriptions :

Tableau 8 : Description des classes Backend du SkylineApp

Classe	**Description**
DataController.java	Classe permettant de contrôler les choix de l'utilisateur et de définir les endpoints (où le résultat de get request sera stocké sous forme JSON)
SkylineMain.java	Classe permettant d'exécuter le programme après la lecture de données stockées sur le fichier texte en les ajoutant à la base de données.
DataND.java	Classe qui représente l'entité "DataND" qui est définit avec un libellé (label) et N dimensions (dim1, dim2, dimN).
DataNDRepository.java	Interface permettant d'interagir avec la base de données via JPA.
CipheringtApplicationTests.java	Classe de test destinée à tester l'application sur la base de connaissance de notre problématique de chiffrement. Elle permet aussi de détecter la source du problème dans le cas d'une erreur.

3.2.2.4 Aperçu sur le Code source

➢ *Fichier « application.properties » :*

Toute la configuration de l'application est gérée dans un fichier de configuration nommé **application.properties**. Ce fichier se trouve dans le répertoire src/main/ressources, il permet de spécifier les valeurs des paramètres nécessaires pour l'exécution de l'application. Comme indiqué dans l'exemple ci-dessous dans lequel on trouve les valeurs des paramètres de la connexion à la base de données :

```
spring.datasource.url = jdbc:mysql://localhost:8889/skylineData?createDatabaseIfNotExist=true&serverTimezone=UTC
spring.datasource.username = root
spring.datasource.password = root
spring.jpa.hibernate.ddl-auto = create
spring.jpa.properties.hibernate.dialect = org.hibernate.dialect.MySQL5Dialect
server.port=9090
```

Par exemple, pour le port du serveur, nous changeons la propriété **server.port** pour activer l'application sur un port ouvert et non utilisé. Par défaut, le serveur intégré (Tomcat) démarre sur le port 8080, et pour le changer il suffit de choisir un port ouvert dans la machine (e.g. 9090).

➢ *SkylineMain.java :*

La classe main contient la méthode **SpringApplication.run(),** l'une des méthodes les plus importante pour l'exécution de notre application en utilisant Spring Boot.

La méthode public **void run(String... args)** a comme tâche de lire les données à partir d'un fichier en entrée et les enregistrer dans la base de données.

➢ *DataControler.java :*

- La méthode **public List<Data<String,Double>> allparams(@RequestParam Map<String,String> allParams)** :
  - ✓ prend comme paramètres une Map de type String,String et retourne le Skyline dans un Endpoint http://localhost:9090/api/data,
  - ✓ la collection Map <String,String> mappe les informations d'une HTTP Get request.
  - ✓ Par exemple :

http://localhost:9090/api/data?t=2&skyline=INDEX&choix1=dim1&cri1=Max&choix2=dim2&cri2=Min…&choixN=dimN&criN=Max

- les clés de la Map sont : choix {1..N}, Cri{1...N}, Dimension et Skyline.
- les valeurs sont dim{1...N}, {Max,Min...} , N et INDEX .

- La méthode public double **AlgoTemps()** retourne dans un Endpoint http://localhost:9090/temps le temps consommé par l'algorithme choisit pour retourner le résultat .

3.2.3 Partie Frontend

La partie Frontend de l'application SkylineApp a été réalisée sur la base des framework Java Angular et Bootstrap :

Le Framework Angular : est un framework JavaScript géré par Google basé sur le langage de programmation libre TypeScript. Il permet de compiler ce code en JavaScript pour qu'il soit compréhensible par les navigateurs.

Le Framework Bootstrap : est un framework open source développé par Twitter. Ce langage utilise les principaux langages de développement web (HTML, CSS & Javascript). Il s'agit d'un code qui raccourcit différentes fonctionnalités permettant au développeur de gagner du temps et de faciliter la réalisation des codes complexes (animation, carrousel, tableau, histogramme, …) tout en réduisant la quantité de caractères requis, et donc le poids du site web.

Le Frontend de l'application SkylineApp s'appuie sur trois composants essentiel pour l'exécution de cette application et qui sont :

Tableau 9 : Description des classes Frontend du SkylineApp

Classe	Description
skyline.component.ts	Code métier contrôlant le formulaire.
skyline.component.html	Vu contenant le formulaire et affichant le résultat dynamiquement.
skyline.service.ts	Service offrant des méthodes permettant la connexion entre la partie Backend et la partie Frontend en récupérant les données des endpoints.

3.2.4 Interface Graphique

La première page de l'application SkylineApp permet de définir le nombre de dimensions à prendre en considération dans chaque scénario de test. Une fois définie, d'autres champs de sélection se sont affichés automatiquement afin de pouvoir choisir les dimensions à tester. Et à la fin, un autre champ de sélection est affiché pour choisir l'algorithme Skyline à tester aussi.

Ci-dessous l'ensemble des interfaces graphiques montrant le mode d'utilisation de cette application ;

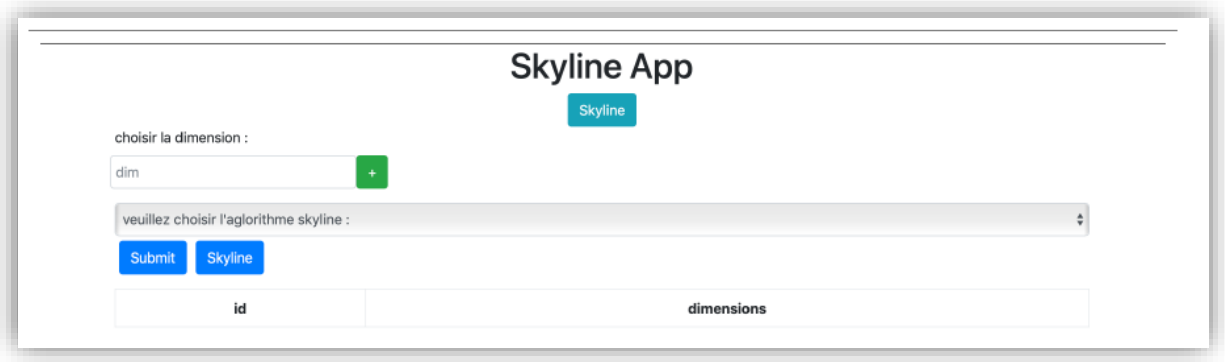


Figure 43: Capture de l'interface initiale de l'application SkylineApp

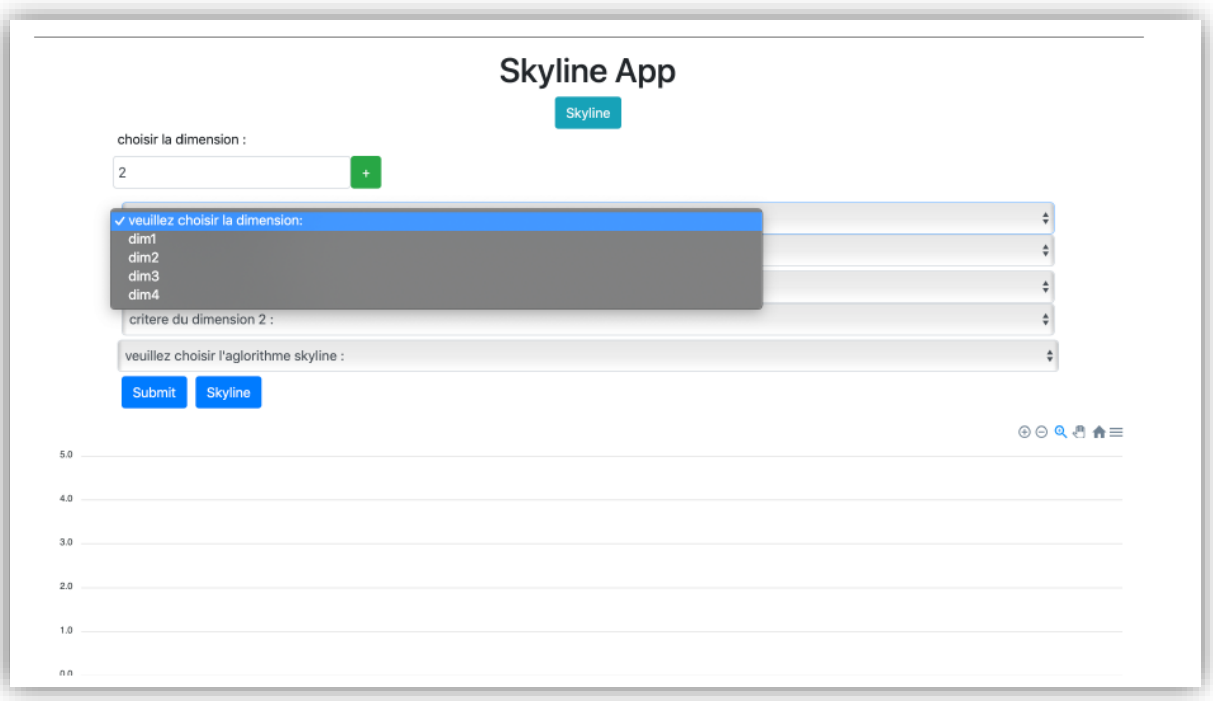


Figure 44: Choix des dimensions dans l'application SkylineApp

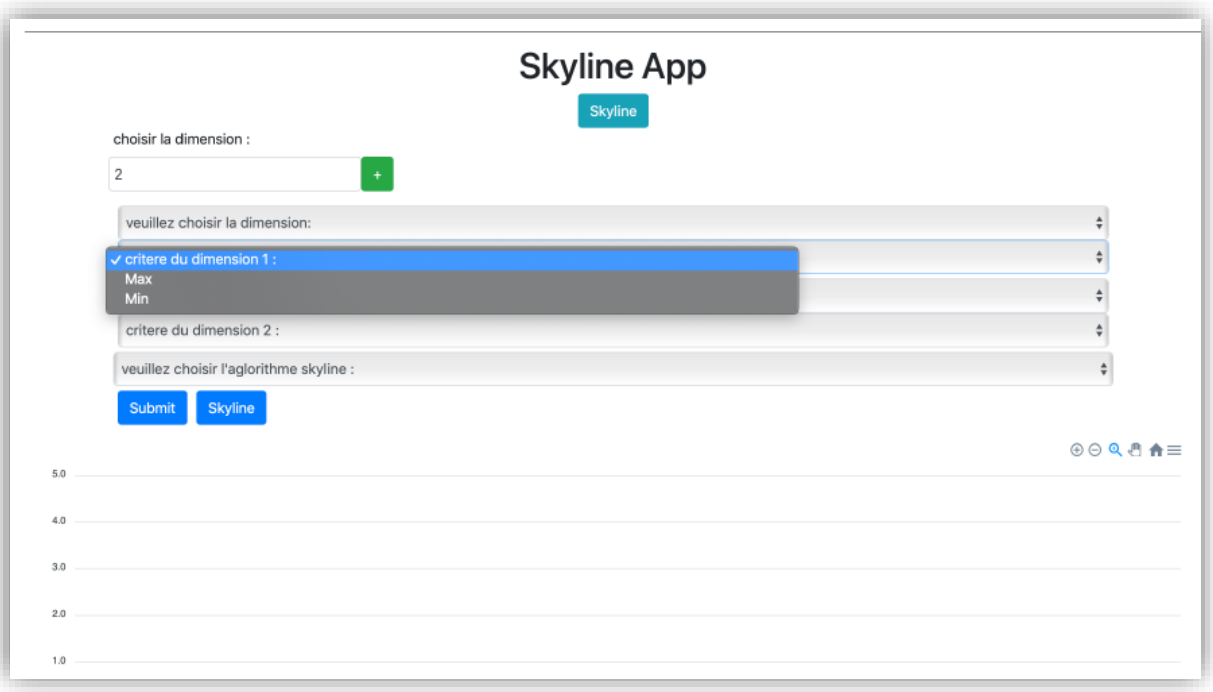


Figure 45 : Choix du critère d'agrégation pour chaque dimension

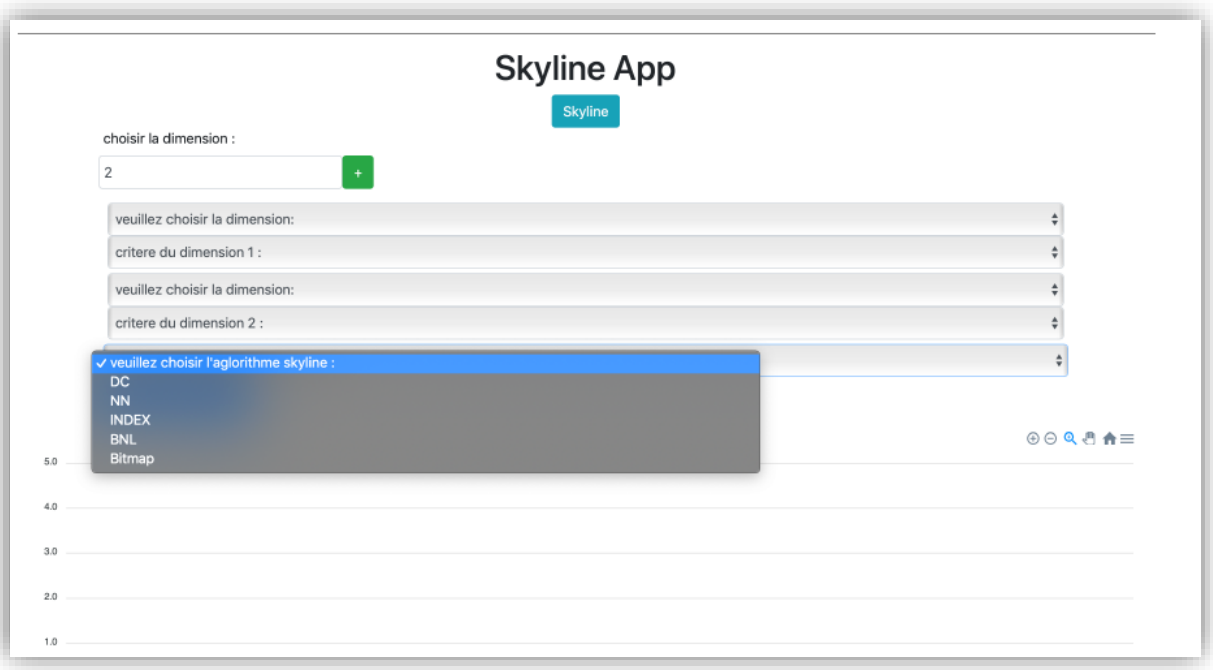
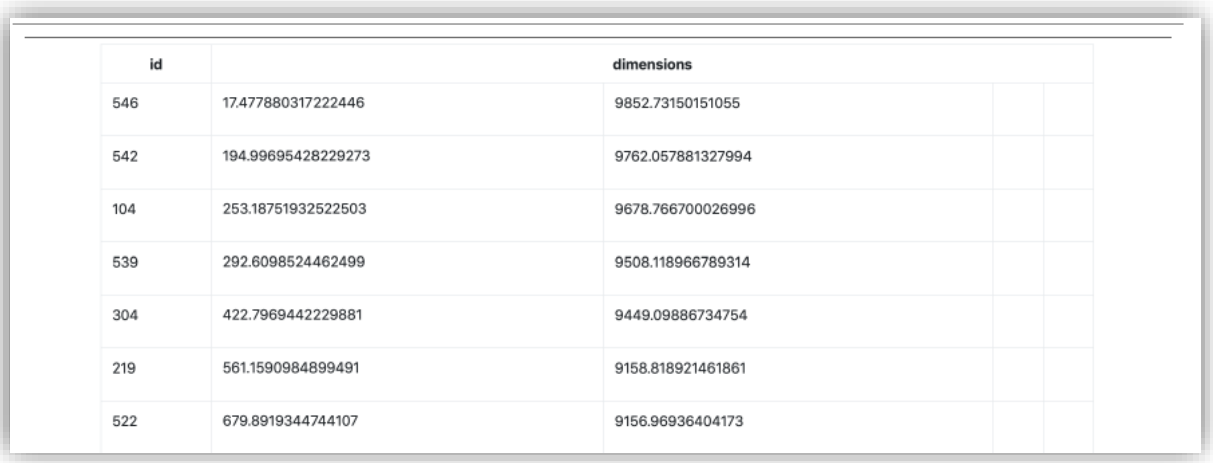


Figure 46: Choix de l'algorithme Skyline à appliquer

Les résultats de chaque expérience sont ensuite affichés dans la même page du formulaire préalablement rempli. Dans le cas de deux dimensions, nous disposons de deux types d'affichage : un affichage sous forme de tableau (fig. 44) et un autre affichage sous forme de graphe de points comme dans la figure (fig. 45) ci-dessous :



id		dimensions
546	17.477880317222446	9852.73150151055
542	194.99695428229273	9762.057881327994
104	253.18751932522503	9678.766700026996
539	292.6098524462499	9508.118966789314
304	422.7969442229881	9449.09886734754
219	561.1590984899491	9158.818921461861
522	679.8919344744107	9156.96936404173

Figure 47 : Affichage des points Skyline dans un tableau

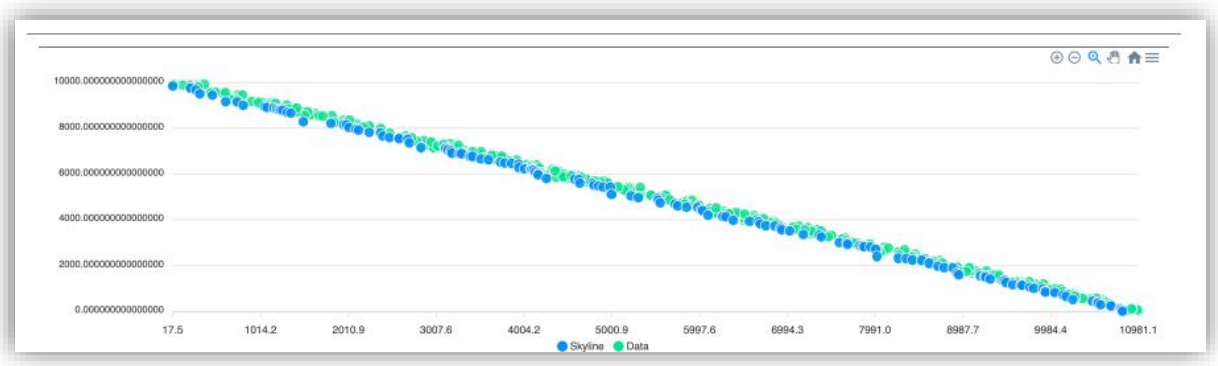


Figure 48 : Affichage des points Skyline dans un graphe

IV- Expériences

4.1Environnement de tests

Pour réaliser ces expériences, nous avons préparé une base de connaissance avec des données réelles. L'alimentation de cette base de données a été faite sur la base d'un échantillon de données en entrée choisies d'une manière arbitraire dont 100 textes de mêmes caractéristiques fixes. Ensuite, nous avons appliqué sur chaque texte les cinq algorithmes de chiffrements AES, Blowfish, DES, TripleDES et ASEC. Ce qui fait un total de plus de mille lignes de tests.

L'environnement de tests repose sur une machine Intel(R) Core(TM) i7-6700HQ , 2.60GHz, 16GB RAM, 64bit OS.

4.2 Spécification des exigences

En effet, dans une approche d'optimisation des traitements, cette méthode réalise deux requêtes pour attaquer la base de connaissances.

La première requête sert à minimiser la volumétrie des données à attaquer dans l'étape qui suit. Elle utilise des critères fixes dont les valeurs sont figées au début de chaque échange.

Les critères fixes représentent les propriétés de la communication courante comme par exemple :

- Type de la donnée à échanger

- caractéristiques du réseau source
- caractéristiques du réseau destination
- les licences disponibles.
- capacité de la machine de chiffrement
- capacité de la machine de déchiffrement
- robustesse du canal de transport entre les deux réseaux

La deuxième requête utilise l'algorithme Skyline et considère les critères à optimiser et à partir desquels on concrétise le souhait des utilisateurs en termes de niveau de sécurité à acquérir. D'où la distinction entre les deux types de critères.

4.3Les critères de dominance

- ✓ Le coût du chiffrement/déchiffrement
- ✓ La vitesse de chiffrement/déchiffrement
- ✓ La réputation de l'algorithme
- ✓ Qualité (taux d'aléatoire / entropie)

4.4Scénarios des tests

Les expériences réalisées dans le cadre de cette contribution, ont été définit sur un ensemble de scénarios de tests. Chaque scénario de test est définit en fonction du nombre de dimensions et de leurs types. On essaie alors de combiner au maximum les différents choix possibles dans notre contexte. L'application attaque la base de connaissances préalablement configurée et exécute l'algorithme Skyline choisi avec les critères saisis, puis retourne les points Skyline reflétant l'algorithme de chiffrement élu. Le but étant de démontrer l'efficacité de ces algorithmes Skyline dans chacun des cas spécifiés partant de deux dimensions jusqu'aux 10 dimensions. Nous avons ajouté d'autres dimensions dans nos scénarios de tests contenant des valeurs fictives afin de considérer un nombre important de dimensions que remplace les critères réseau/plateforme/canal de transport, et dont on souhaite les intégrer encore dans notre système très prochainement.

La fiche des différents scénarios de tests est capitalisée dans le tableau suivant :

Tableau 10 : La fiche des différents scénarios de tests

N°	Nbr de Dimensions	Nom des Dimensions	Critère de dominance
1	2	CipheringRuntime	Min
		Entropy	Max
2	2	DecipheringRuntime	Min
		DecipheringMemory	Min
3	4	CipheringRuntime	Min
		Entropy	Max
		DecipheringRuntime	Min
		DecipheringMemory	Min
4	4	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
5	6	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Entropy	Max
		Dim_fictive_1* (nbr de valeurs limité)	Min
6	6	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Dim_fictive_1* (nbr de valeurs limité)	Max
		Dim_fictive_2* (nbr de valeurs limité)	Min
7	10	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Entropy	Max
		Dim_fictive_1 (nbr de valeurs limité)	Min
		Dim_fictive_2 (nbr de valeurs limité)	Min
		Dim_fictive_3* (nbr de valeurs limité)	Min
		Dim_fictive_4* (nbr de valeurs limité)	Min
		Dim_fictive_5* (nbr de valeurs limité)	Min
8	10	CipheringRuntime	Min
		DecipheringRuntime	Min
		CipheringMemory	Min
		DecipheringMemory	Min
		Dim_fictive_1 (nbr de valeurs limité)	Max
		Dim_fictive_2 (nbr de valeurs limité)	Min

	Dim_fictive_3 (nbr de valeurs limité)	Min
	Dim_fictive_4 (nbr de valeurs limité)	Min
	Dim_fictive_5 (nbr de valeurs limité)	Min
	Dim_fictive_6* (nbr de valeurs limité)	Min

* : Les Dim_fictive_i représentent des dimensions fictives dont les valeurs ont été générées avec une règle approximative pour désigner les contraintes de l'infrastructure qu'on n'est pas en mesure de les calculer réellement et les intégrer dans la version actuelle de notre prototype. Ces dimensions ont été ajoutées à l'application afin

4.5Résultats et Discussions

4.5.1 Etude d'impact : 'Choix de la Dimension'

Pour réaliser cette étude nous allons considérer à chaque fois les deux scénarios ayant le même nombre de dimensions mais dont le choix des dimensions diffère.

Nous présenterons dans cette étude, le détail des résultats obtenus dans les deux premiers scénarios 1 et 2, puis nous confirmons notre constat sur les autres scénarios.

Résultat du Scénario 1 : (2 Dimensions)

N° du Scénario	Nbr de Dimensions	Nom des Dimensions	Critère de Dominance
1	2	CipheringRuntime	Min
		Entropy	Max

- L'Objectif de cette expérience est de :
  - Minimiser le temps de chiffrement
  - Maximiser l'entropie

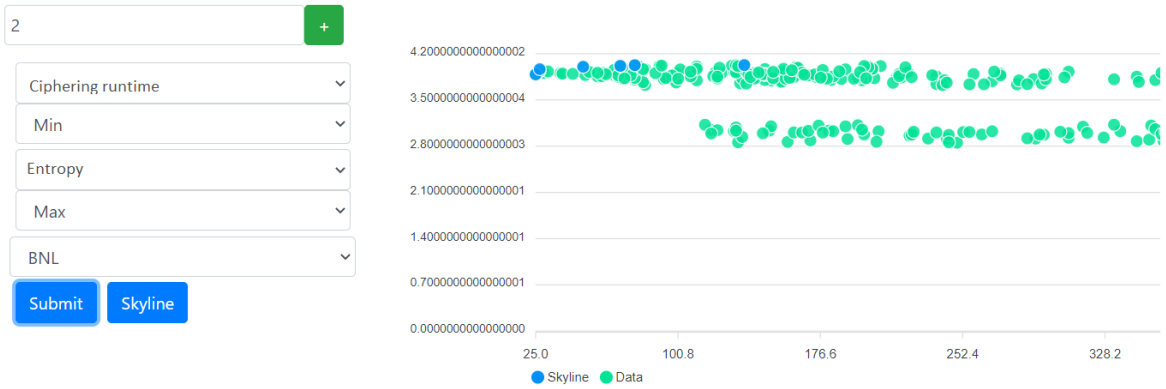


Figure 49 : Application de l'algorithme BNL avec 2 dimensions

Figure 50 : Application de l'algorithme DC avec 2 dimensions

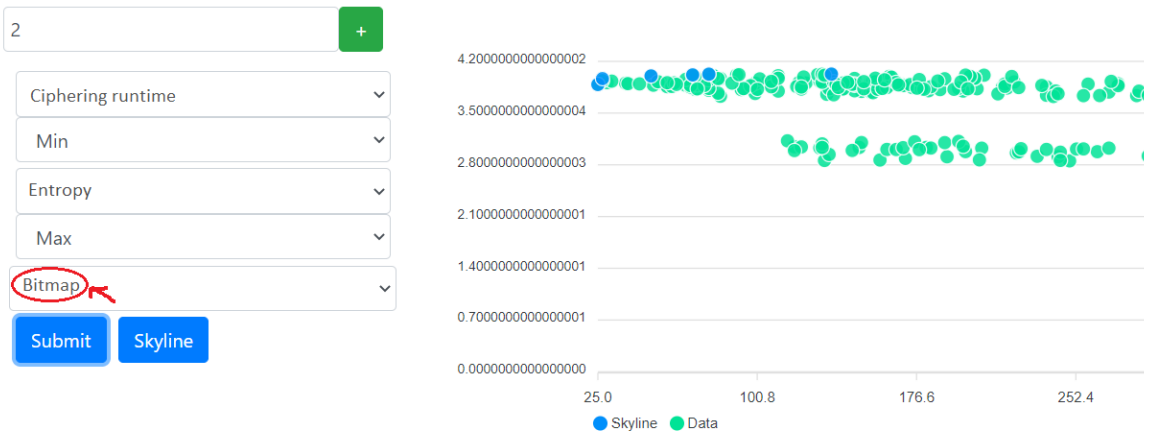


Figure 51 : Application de l'algorithme Bitmap avec 2 dimensions

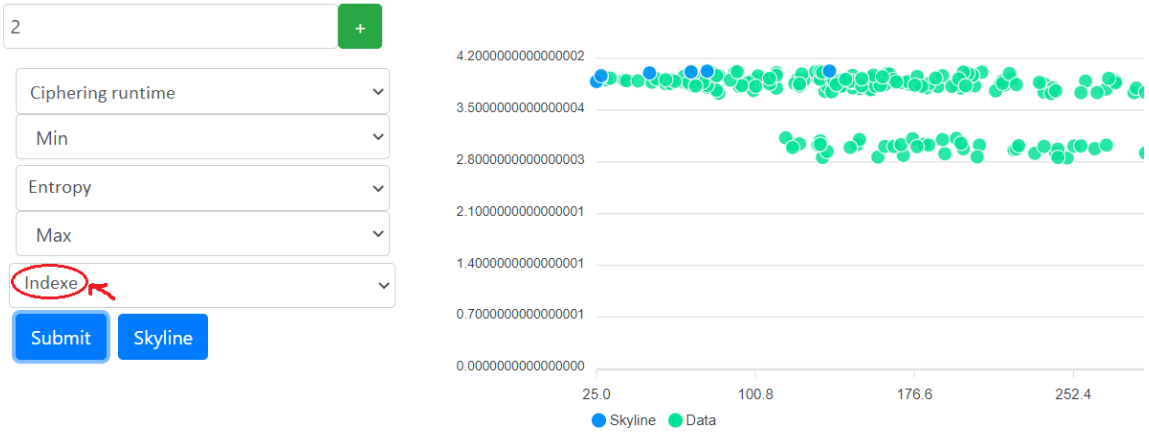


Figure 52 : Application de l'algorithme Indexe avec 2 dimensions

Figure 53: Application de l'algorithme NN avec 2 dimensions

Les exécutions sur les différents algorithmes ont été toutes pareilles et considèrent les 6 points en bleu comme solution dominante.

Le résultat de l'envoi de la requête Skyline a donné 6 points Skyline ayant les valeurs suivantes :

Algorithm	Ciphering Runtime (ms)	Entropy
BLOWFISH	250	3.8870066417181426
BLOWFISH	271	3.9658710817437597
BLOWFISH	487	4.00235367813974
BLOWFISH	661	4.01642913400371
AES	723	4.026504332171881

BLOWFISH	926	4.027404704120802

Tableau 11 : Le résultat de l'envoi de la requête Skyline

L'algorithme de chiffrement le plus performant est BLOWFISH vient après l'algorithme AES en deuxième position. En effet, dans ce cas de figure ces deux algorithmes sont très connus dans la communié de la sécurité informatique par leurs performances et leur niveau de sécurité. Maintenant, si on revient à la performance des algorithmes Skyline, Pour aboutir à ces résultats on constate une différence énorme dans le temps d'exécution de ces algorithmes.

Le diagramme ci-dessous présente le résultat du temps d'exécution de chaque algorithme Skyline.

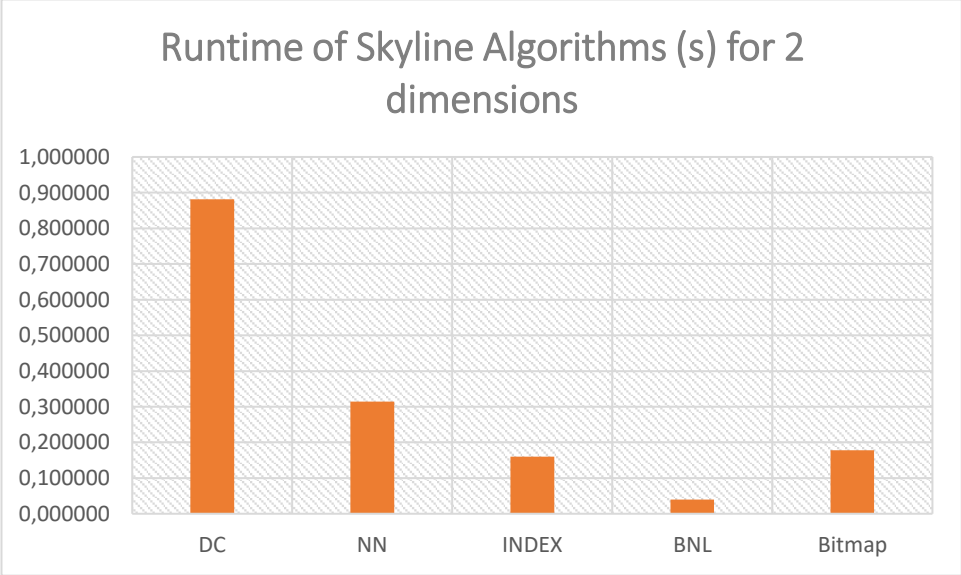


Figure 54 : Temps d'exécution des algorithmes Skyline (s) pour 2 dimensions

Donc, d'après le diagramme fig.51 on peut conclure que l'algorithme BNL est le plus simple et le meilleur dans le cas où on n'utilise les deux dimensions temps de chiffrement/Entropie.

Cependant, si on change les deux dimensions considérées dans la première expérience et on opte pour le scénario suivant :

Résultat du Scénario 2 : (2 Dimensions)

N°	Nbr de Dimensions	Nom des Dimensions	Critère de dominance
2	2	DecipheringRuntime	Min
		DecipheringMemory	Min

- L'Objectif de cette expérience est de :
  - Minimiser le temps de déchiffrement

- Maximiser la capacité mémoire nécessaire pour un déchiffrement.

Cette fois ci, les résultats en termes de points Skyline sont différents. Ci-dessous le tableau récapitulatif des résultats des différentes exécutions :

Algorithm	Deciphering Runtime(ms)	Deciphering Memory used (K)
SEC	98	1695
SEC	123	1841
ASEC	215	2503
AES	243	2547

Tableau 12 : tableau récapitulatif des résultats du scénario 2

L'algorithme de chiffrement SEC [10] est classé le premier, suivi de l'algorithme ASEC (version avancée de SEC), vient ensuite l'algorithme AES.

Le résultat donné par ce système apparait logique car SEC est un algorithme évolutionniste qui prend assez de temps pour effectuer le chiffrement et générer la clé, cependant, dans le processus de déchiffrement il applique une seule opération sur tout le texte d'un seul coup. Ce qui fait, ce type d'algorithme est le plus performent vis-à-vis au coût de déchiffrement.

Passant, maintenant au résultat du temps d'exécution des algorithmes Skyline. Le diagramme suivant illustre les résultats de ce scénario :

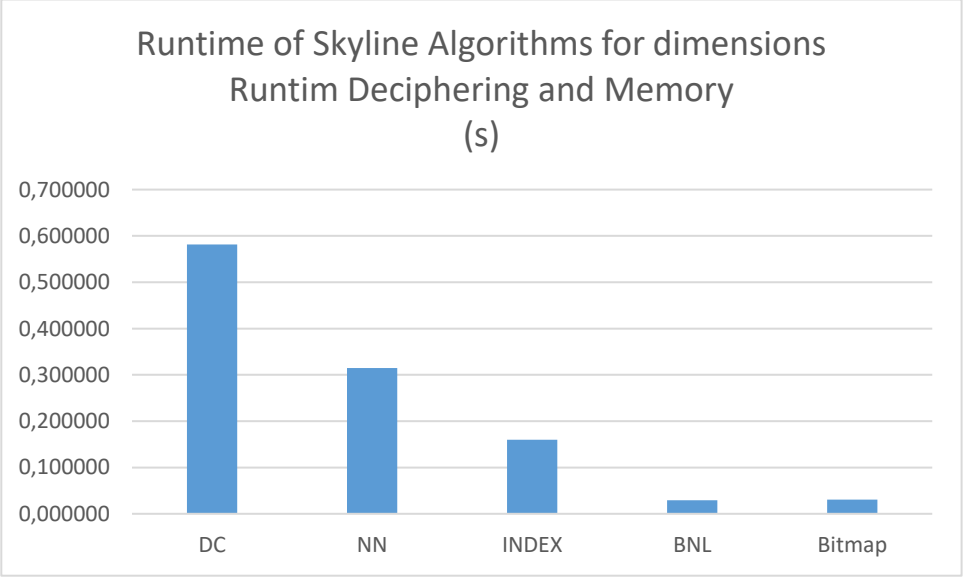


Figure 55 : Comparaison du temps d'exécution des algorithmes Skyline avec les dimensions - temps de déchiffrement et mémoire

D'après cette expérience, on observe que l'algorithme Bitmap a mieux performé dans ce scénario contrairement au 1er scénario.

Discussion 1 :

D'après les deux premiers scénarios dont le nombre de dimensions est '2', nous constatons qu'en changeant le type de dimensions les résultats changent, chose qui est normal du fait que la dominance est liée à l'ensemble des dimensions choisies. Cependant, ce qui nous a attiré plus c'est au niveau des performances des algorithmes Skyline qui ont aussi changé surtout par rapport à l'algorithme Bitmap. En effet, ce dernier a bien performé dans le 2ème scénario du fait que la représentation vectorielle Bitmap pour les deux dimensions choisies est moins complexe. On peut conclure alors que plus le nombre de valeurs d'une dimension sont maitrisées plus l'algorithme devient rapide et plus performant. Dans le scénario 1, la dimension de l'entropie avait des valeurs très variées et par conséquent la représentation vectorielle Bitmap sera d'une taille très considérable.

En revanche, l'algorithme BNL reste toujours performant même en modifiant le type de dimensions considérées.

Constat 1 :

- ✓ BNL performe bien dans le cas de 2 dimensions
- ✓ BNL n'a pas été impacté par les valeurs de la dimension choisie.
- ✓ Bitmap performe bien dans le cas de 2 dimensions
- ✓ Bitmap a été impacté par les valeurs de la dimension choisie.

Conclusion 1 :

- ↳ **Le paramétrage de l'application doit prendre en considération les dimensions dont l'écart entre leurs différentes valeurs est très petit.**

4.5.2 Etude d'impact : 'Nombre de Dimensions'

Après avoir confirmé à partir des premières expériences que les valeurs au niveau de chaque dimension impactent très considérablement les performances du système dans le cas de Bitmap, Dans la suite des expériences, nous allons se concentrer plus sur le nombre de dimensions afin de déterminer l'algorithme Skyline le plus performant dans ce cas de figure.

L'objectif étant de définir le paramétrage à prendre en compte en modifiant le nombre des dimensions, nous allons présenter dans cette rubrique les résultats des scénarios 3/5/8 qui s'appliquent respectivement sur 4/6/10 dimensions.

Les autres scénarios qui restent s'appliquent aussi sur ce même nombre de dimensions sauf qu'ils considèrent à chaque fois des types de dimensions différents dont ils ne font pas objet de cette étude d'impact.

Les diagrammes ci-dessous montrent le résultat du temps d'exécution des algorithmes Skyline dans chaque cas de figure.

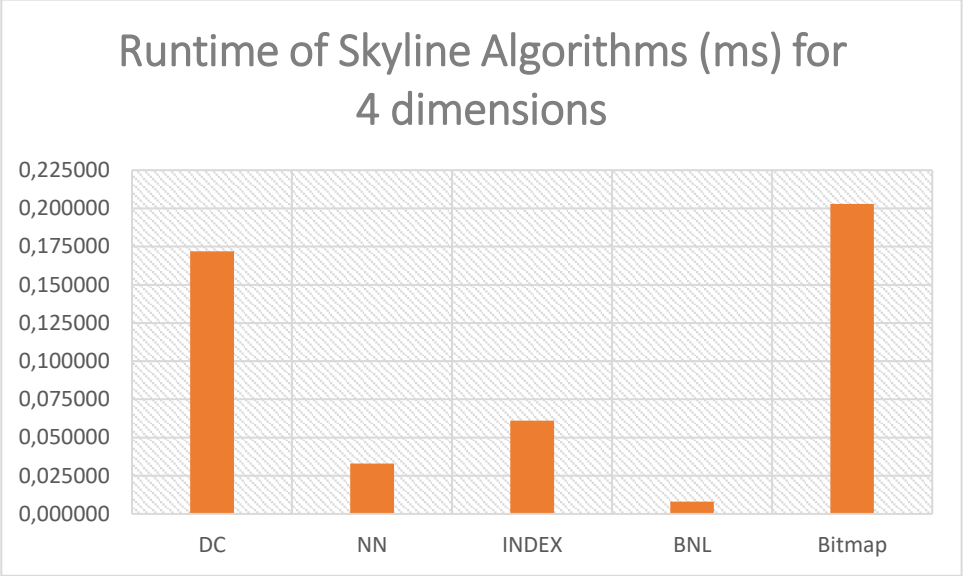


Figure 56 : Temps d'exécution des algorithmes Skyline avec 4 dimensions

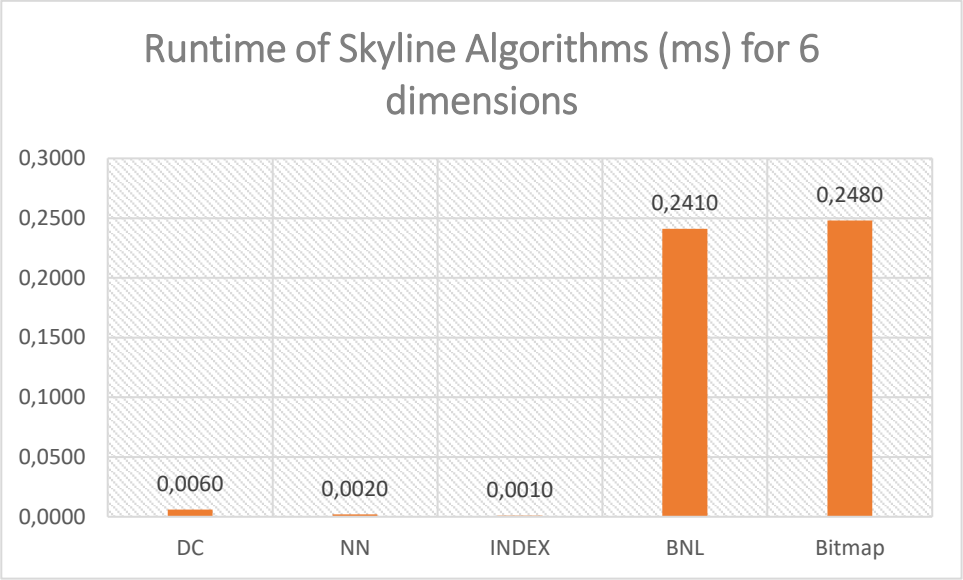


Figure 57 : Temps d'exécution des algorithmes Skyline avec 6 dimensions

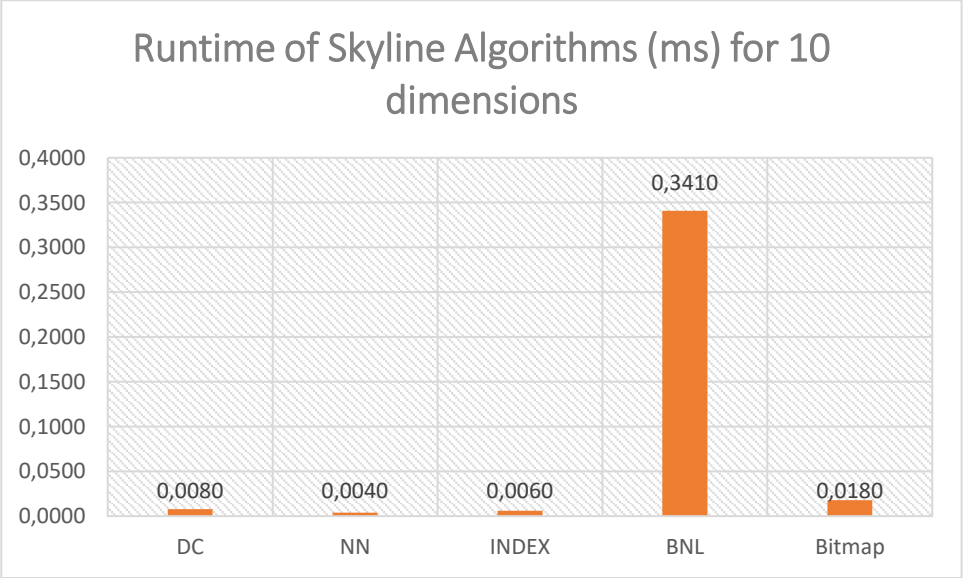


Figure 58 : Temps d'exécution des algorithmes Skyline avec 10 dimensions

Discussion 2 :

D'après les résultats de ces diagrammes, on constate que l'algorithme BNL devient de plus en plus lent et très complexe tandis que les autres algorithmes commencent à performer à partir de la 4ème dimension. Cette augmentation positive concerne plus les algorithmes DC, NN et index. L'algorithme Bitmap par contre n'est pas impacté par le nombre de dimensions car son exécution était plus rapide dans le cas de 10 dimensions alors qu'elle était très lente

dans le cas de 4 et 6 dimensions. Ce constat confirme alors le premier, car si on revient aux scénarios 3 et 5 de 4 et 6 dimensions, on va trouver la dimension 'entropie' dans les choix spécifiés. Cependant, le scénario 8 qui correspond à 10 dimensions on ne trouve plus la dimension de l'entropie.

Cette analyse peut être visualisée dans les deux graphes suivants :

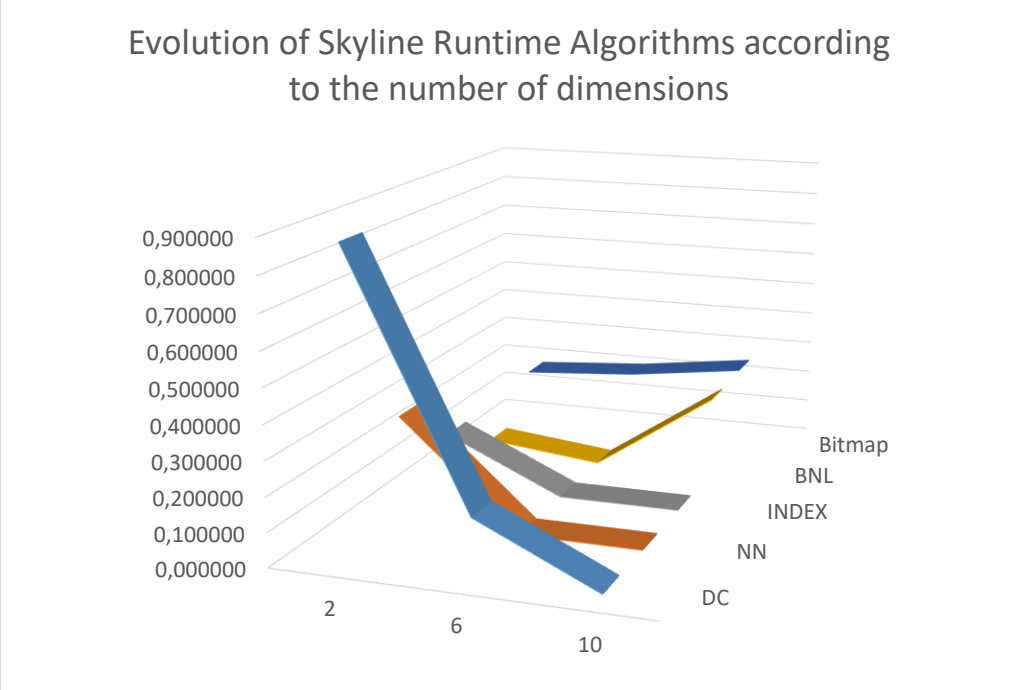


Figure 59 : Évolution du temps d'exécution des algorithmes Skyline en fonction du nombre de dimensions (scénarios 2, 4, 6, 7)

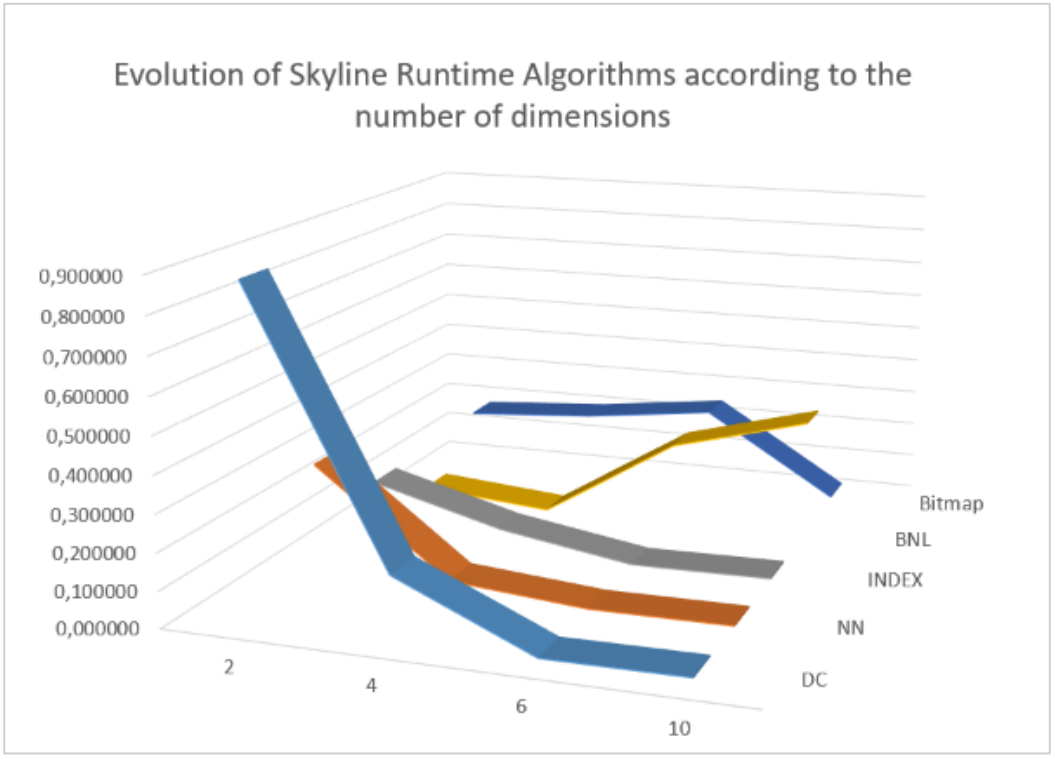


Figure 60 : Évolution du temps d'exécution des algorithmes Skyline en fonction du nombre de dimensions (scénarios 1, 3, 5, 8)

Constat 2 :

- ✓ DC, NN et indexe performe bien dans le cas de plusieurs dimensions
- ✓ DC, NN et indexe ne sont pas impacté par les valeurs de la dimension choisie.
- ✓ Bitmap performe bien dans le cas de 10 dimensions
- ✓ Bitmap a été impacté par les valeurs de la dimension choisie.

Conclusion 2 :

- ↳ **Le paramétrage de l'application doit prendre en considération le nombre de dimensions choisies.**

Au final, on peut conclure que pour augmenter les performances de notre système de chiffrement intelligent, on doit adapter l'algorithme Skyline en fonction des critères choisis par l'utilisateur. D'où la nécessité d'avoir une couche de paramétrage au niveau de moteur d'inférences afin de définir l'algorithme Skyline les plus approprié. Cette couche de paramétrage prend en compte les deux paramètres suivants :

- ✓ Le nombre des dimensions,
- ✓ Le choix de ces dimensions

Dans cette optique et en étudiant l'impact de chacun de ces paramètres nous avons pu distinguer entre 3 catégories :

- Catégorie 1 : Si le nombre de dimensions est ≤ 4, Alors performances abouties en utilisant BNL.
- Catégorie 2 : Si le nombre de dimensions est > 4, Alors performances abouties en utilisant DC ou NN.
- Catégorie 3 : Si toutes des dimensions choisies ont un nombre restreint de différentes valeurs, Alors performances abouties en utilisant Bitmap.

V- Conclusion

Durant ce travail, nous avons mis le point sur la nouvelle approche de chiffrement intelligent qui se base sur des concepts de l'intelligence artificielle dans le but d'améliorer ses performances. L'approche en question utilise un système à base de connaissance et applique les algorithmes Skyline afin de définir la politique à suivre pour assurer la confidentialité des échanges. Dans ce travail, nous avons appliqué différents algorithmes Skyline à travers l'application SkylineApp, Puis nous avons comparé les résultats des différentes exécutions en changeant dans chaque scénario le nombre et le type des dimensions à considérer. Après plusieurs scénarios de tests nous avons conclu que la performance des algorithmes Skyline sur notre problème de chiffrement dépend de deux paramètres essentiels, à savoir : le nombre des dimensions et le choix de ces dimensions. Par conséquent, rendre le choix de l'algorithme Skyline dynamique en fonction du nombre des dimensions augmente les performances de ce système en termes de temps d'exécution et de qualité de la décision de chiffrement retournée.

Partie II

Blockchain & Sécurité

“

Si vous n’y croyez pas ou si vous ne comprenez pas, je n’ai pas le temps de vous convaincre, désolé »

Satoshi Nakamoto,
Créateur anonyme de Bitcoin.

Chapitre 5 |

La Technologie Blockchain

Sommaire

I- Introduction

La technologie Blockchain est très en vogue durant cette dernière décennie[92]. Cette grande innovation, qui a vu le jour en 2008 (par Satoshi Nakamoto)[93], donnant naissance au Bitcoin a été au début, réservée uniquement à la crypto-monnaie. Mais au fil des années, ses champs d'application se sont élargis et elle est devenue omniprésente non seulement dans le secteur financier mais quasiment dans l'économie d'une manière générale[94].

Actuellement la Blockchain est au service de différents secteurs tels que les banques, assurances, santé et industrie pharmaceutique, chaine d'approvisionnement de nombreux secteurs (agroalimentaire, commerce international, distribution, aéronautique, automobile, etc), industrie musicale, énergie , immobilier, vote, etc[95]. D'après le concept de Klaus Schwab, (économiste allemand contemporain, fondateur et président du Forum économique mondial) ; elle est au cœur de la 4^{ième} révolution technologique numérique en parallèle avec l'intelligence artificielle, sachant que la troisième révolution industrielle était digitale (avènement de l'informatique en 1960) tandis que les 2 premières révolutions industrielles étaient techniques (Machine à vapeur entre 1760 et 1840, puis électricité et production de masse)[96].

En fait, le caractère décentralisé de la Blockchain, couplé avec sa sécurité et sa transparence, permet la collaboration entre les utilisateurs et le développement de l'intelligence collective afin d'assurer un vrai dynamisme dans l'économie[97]. La décentralisation de la Blockchain est due au fait qu'elle est constituée d'un réseau distribué fonctionnant en mode P2P (peer-to-peer)[93]. Ce réseau est un ensemble de nœuds obéissant tous à un même protocole informatique, et c'est à travers ces nœuds que se font les transmissions et stockages des informations structurées par des blocs chaînés et liés entre eux, en respectant des intervalles de temps réguliers. La validation des informations regroupées dans les blocs se fait par des utilisateurs - validateurs désignés par « mineurs » dans le cas de la crypto-monnaie[98]. Quant à la sécurité de la Blockchain, elle est principalement assurée par des outils de la Cryptographie[92] en particulier la cryptographie asymétrique tels que : Fonctions de hachage, signatures

numérique, etc. Elle est notamment fondée sur des problèmes réputés difficiles en mathématiques.

De son côté, Michel Bauwen, Fondateur de la P2P, a précisé que la Blockchain permet à des entités autonomes d'établir des contrats entre eux sans avoir besoin d'un collectif, de communauté ou quiconque tiers de confiance[94]. En effet, les smart-contrat en est une concrète application. Ces derniers sont des programmes autonomes qui exécutent automatiquement les conditions et les termes d'un contrat préalablement définis et déployé dans la Blockchain, sans avoir besoin d'une intervention humaine[92].

II- A propos de la Blockchain

2.1 Principe de Fonctionnement

Pour comprendre le fonctionnement de la Blockchain, il faut absolument passer par la compréhension du fonctionnement du Bitcoin puisqu'il représente la source d'inspiration et de la découverte de cette technologie.

Dans le contexte des systèmes Bitcoin, la Blockchain est utilisée comme archive publique sécurisée et de confiance pour toutes les transactions qui échangent des bitcoins sur ce réseau. Toutes les transactions sont enregistrées, organisées et stockées dans des blocs sécurisés en utilisant des outils cryptographiques très puissants[93]. Ces outils permettent de garantir l'enchaînement de ces blocs d'une manière fiable et persistante. La Blockchain est le garde-fou essentiel pour sécuriser les transactions bitcoin contre de nombreux problèmes connus et difficiles de sécurité, de confidentialité et de confiance, tels que les doubles dépenses, la divulgation non autorisée de transactions privées, le recours à une autorité centrale de confiance et le manque de fiabilité de l'informatique décentralisée[93]. La manière dont le bitcoin déployait la Blockchain a été l'inspiration pour de nombreuses autres applications.

Fonctionnellement, une Blockchain peut être considérée comme une base de données distribuée et sécurisée de journaux des transactions. Dans un réseau Bitcoin, si Alice souhaite envoyer des bitcoins à Bob, elle créera une transaction bitcoin et la diffuse à tous les nœuds du réseau afin de lancer le processus de minage. La transaction doit être

approuvée par les mineurs en premier temps. Ces derniers collecteront les transactions dans un bloc, vérifieront les transactions dans ce bloc et diffuseront le bloc et sa vérification à l'aide d'un protocole de consensus (Proof of Work) pour obtenir l'approbation du réseau. Lorsque les autres nœuds vérifient que toutes les transactions contenues dans le bloc sont valides, le bloc peut être ajouté à la Blockchain. Ce n'est que lorsque le « bloc » contenant la transaction est approuvé par les autres nœuds et ajouté à la Blockchain que ce transfert de bitcoin de Alice vers Bob deviendra finalisé et légitime[8]. La figure 27 fournit une illustration de ce processus.

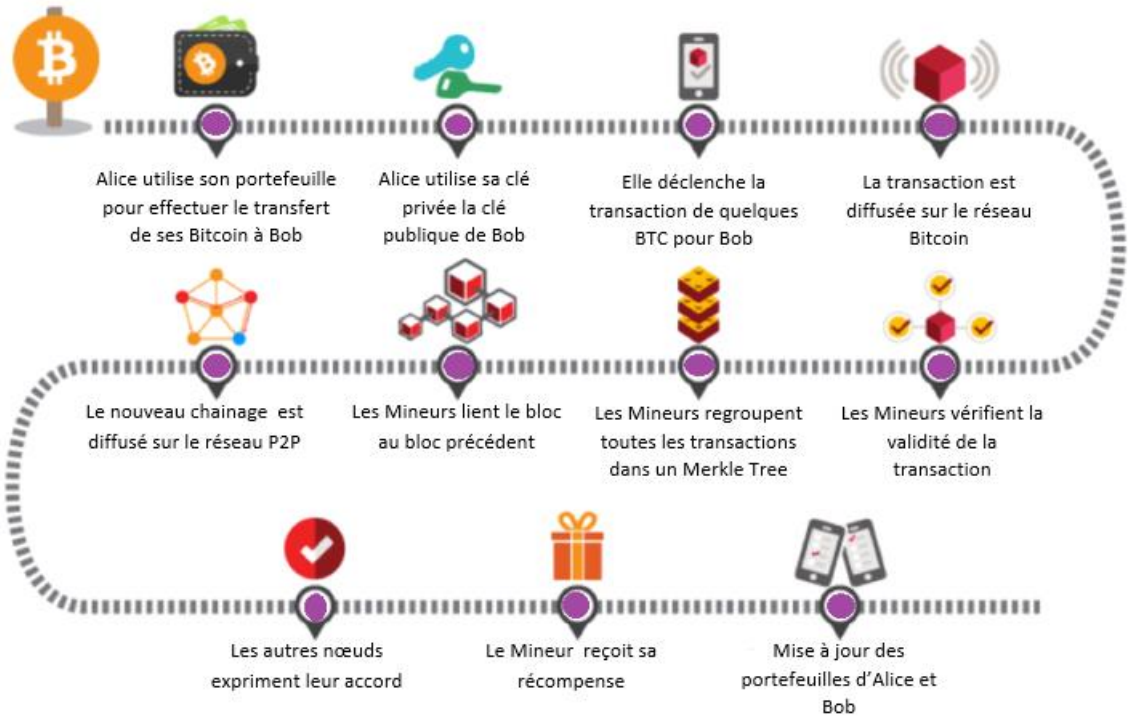


Figure 61 : Fonctionnement de la Blockchain Bitcoin

Globalement, on peut déduire trois principales fonctionnalités très importantes dans la mise en œuvre de la Blockchain dans Bitcoin [98]. Ces capacités sont :

- Le stockage chaîné par hachage,
- La signature numérique et
- Le consensus d'engagement pour l'ajout d'un nouveau bloc au registre chaîné.

Par une combinaison élégante d'une suite de techniques de sécurité les plus populaires, telles que la chaîne de hachage, l'arbre Merkle, la signature numérique, ainsi que les

mécanismes de consensus, la Blockchain Bitcoin peut à la fois éviter le problème de double dépense des bitcoins et arrêter la modification rétrospective de toutes les données de transaction dans un bloc après que le bloc soit validé avec succès au sein de la Blockchain.

2.2 Types de conception

Alors que la technologie Blockchain continue d'évoluer en ce qui concerne la façon dont elle est construite, consultée et vérifiée, elle est sont classée en trois grandes catégories[94] :

- **La Blockchain publique** : c'est une Blockchain ouverte à tout le monde pour lire, envoyer ou recevoir des transactions. Elle permet à tous les participants à se joindre à la procédure de consensus pour prendre la décision sur la validité des blocs, le contenu des transactions, puis l'opération de les ajouter à la Blockchain[94].
- **Consortium Blockchain :** qui a placé certaines contraintes sur les autorisations d'écriture telles que seul un ensemble présélectionné de participants dans le réseau peut influencer et contrôler le processus de consensus, même si la lecture reste ouverte à tout participant dans le réseau[96],
- **la Blockchain privée :** dont les autorisations d'écriture sont strictement limitées à un seul participant (ou organisation), même si ses autorisations de lecture sont ouvertes au publique ou limitées à un sous-ensemble de participants au réseau[99]. Bien que ce type de blockchains sont mieux de point de vue sécurité et performance, ils diffèrent par la vitesse de consensus (SoC) et de combien d'autorités de confiance (TA) sont nécessaires[8].

Comme résumé dans le tableau 13, ces trois catégories de Blockchain partagent certaines propriétés[98] communes :

(1) Ils utilisent tous des réseaux décentralisés Peer to Peer pour les opérations ;
(2) Ils exigent tous que chaque transaction soit signée numériquement et ne soit annexée qu'au Blockchain, et chaque nœud pair maintient une réplique d'un

grand livre mondial distribué de transactions ;

(3) Ils reposent tous sur un consensus pour synchroniser les répliques à travers le réseau.

Types	Description	TA	Vitesse de consensus	Scenarios
Publique	Tout le monde peut participer et est accessible à l'échelle mondiale	0	Lent	Globalement décentralisé
Consortium	Contrôlée par des nœuds présélectionnés par consortium	1	Rapide et léger	Un Business sélectionnées parmi autres organisations
Privée	Les droits d'écriture sont contrôlés par une organisation	1	Rapide	Partage et gestion des informations au sein d'une organisation

Tableau 13 : les 3 Types de la Blockchain

2.3 Concept du Consensus

Dans le contexte décentralisé d'une Blockchain, lorsqu'un nouveau bloc est envoyé par diffusion au réseau, chaque nœud a la possibilité d'ajouter ce bloc à sa copie du grand livre global ou de l'ignorer. Le consensus est utilisé pour chercher à ce que la majorité du réseau se mette d'accord sur une mise à jour unique de l'état afin de sécuriser l'expansion du grand livre global et empêcher les tentatives malhonnêtes ou les attaques malveillantes[92].

Concrètement, étant donné que la Blockchain est un immense registre mondial partagé, n'importe qui peut le mettre à jour, une infraction contradictoire pourrait se produire lorsqu'un nœud décide de falsifier l'état de sa copie du registre, ou lorsque plusieurs nœuds tentent de manière collusoire une telle falsification[92]. Par exemple, si Alice envoyait 10 bitcoins à Bob depuis son portefeuille, elle aimerait être sûre que personne dans le réseau ne peut falsifier le contenu de la transaction et changer 10 bitcoins en 100 bitcoins. Et donc, pour permettre à la Blockchain de fonctionner à l'échelle mondiale avec une garantie de sécurité et d'exactitude, le registre partagé a besoin d'un algorithme de

consensus efficace et sécurisé, qui doit être défectueux tolérant, et assurer les propriétés suivantes :

(i) Tous les nœuds maintiennent simultanément une chaîne de blocs identique

(ii) Pas d'autorité centrale pour empêcher des adversaires malveillants de perturber la coordination et le processus d'obtention d'un consensus.

Bref, chaque message transmis entre les nœuds doit être approuvé par la majorité des participants du réseau au travers d'un accord consensuel. Aussi, le réseau dans son ensemble doit être résilient aux défaillances partielles et aux « attaques », comme dans le cas d'un groupe des nœuds malveillants ou lorsqu'un message en transit est corrompu. Un bon mécanisme de consensus utilisé dans la mise en œuvre de la Blockchain garantit également un solide registre des transactions avec deux propriétés importantes : la persistance et la vivacité.

Garantie de persistance : la réponse cohérente du système concernant l'état d'une transaction. Par exemple, si l'un des nœuds sur le réseau indique qu'une transaction est dans l'état « stable », les autres nœuds du réseau doivent également le signaler comme stable[95].

Garantie de vivacité : déclare que tous les nœuds ou les processus finissent par s'entendre sur une décision ou une valeur. Éventuellement, il peut prendre un temps significatif pour parvenir à un accord. En combinant persévérance et vivacité, il garantit qu'un registre des transactions est robuste de sorte que seules les transactions authentiques sont approuvées et deviennent permanentes[94].

En résumé, le rôle de la Blockchain dans le système Bitcoin est de remplacer la base de données centralisée avec un contrôle d'accès faisant autorité. Une fois que certaines données ont été enregistrées dans le bloc du grand livre global chaîne, il devrait être « impossible » de changer la Blockchain, et en faisant respecter l'accord majoritaire de validité de mise à jour par consensus, il assure l'état de cohérence et empêche le double problème de dépenses.

III- Blockchain & Sécurité

Il est important de souligner que cette thèse s'articule plus sur le volet sécurité de la Blockchain et sur aucuns autres enjeux tels que ceux liés aux domaines métiers de ses applications comme la logistique, la finance, la santé, etc. Néanmoins, même au cours des études primaires dans ces domaines, les chercheurs se confrontent toujours aux problèmes de sécurité à part entière et les classent dans une priorité bien avancée. Dans cet esprit, il convient à noter que ce focus sur la sécurité et la blockchain est transverse et très attendu par tout le monde[100].

En générale, on distingue entre deux orientations dans les travaux de recherches réalisés dans ce domaine. Des travaux qui se situent globalement sur l'une des deux branches suivantes :

- Sécurité pour la blockchain
- Blockchain pour la sécurité

La première consiste à renforcer la sécurité au sein de la blockchain et la deuxième cherche à solutionner la problématique de la sécurité en utilisant la technologie blockchain comme outil de basse.

3.1 La Blockchain pour la Sécurité

La blockchain et les technologies environnantes n'offrent pas de solution miracle pour les problèmes de cyber-sécurité. Au contraire, ils renforcent simplement les efforts existants pour sécuriser les réseaux, les communications et les données[100]. Le chiffrement et le hachage des utilitaires Blockchain pour stocker des enregistrements immuables et de nombreuses solutions de cyber-sécurité existantes utilisent également une technologie très similaire[101]. La majorité des mesures de sécurité existantes reposent sur une seule autorité de confiance pour vérifier les informations ou stocker des données chiffrées. Cela laisse le système vulnérable aux attaques ; de nombreux acteurs malveillants pourraient concentrer leurs efforts sur une cible unique pour commettre des attaques par déni de service, injecter des informations malveillantes et extorquer des données par le biais du vol ou du chantage[101]. Les Blockchain ont le cap sur les contraintes de sécurité actuelles dans la mesure où les véritables Blockchain sont décentralisées et ne nécessitent

l'autorité ou la confiance d'aucun membre du groupe ou du réseau ; le système ne nécessite pas de confiance car chaque nœud, ou membre, dispose d'une copie complète de toutes les informations historiques disponibles et ce n'est qu'en parvenant à un consensus de la majorité que davantage de données seront ajoutées à la chaîne d'informations précédentes[102]. Comme indiqué dans d'autres parties de ce document, cela est réalisé de différentes manières, mais l'essentiel est le suivant : de nombreux membres d'un groupe qui ont tous accès aux mêmes informations seront en mesure de sécuriser ce groupe bien mieux qu'un groupe composé d'un leader et une multitude de membres qui s'appuient sur le leader pour leurs informations, en particulier lorsque les mauvais acteurs pourraient venir sous la forme de membres du groupe ou du leader lui-même. Sur la base des applications Blockchain, nous abordons la manière dont la blockchain a été appliquée pour améliorer la cyber-sécurité dans l'IoT, le stockage et le partage de données, la sécurité du réseau, les données des utilisateurs privés, la navigation et l'utilité du World Wide Web [103]:

La recherche dans ce domaine met l'accent sur les chantiers techniques suivants :

- la sécurité de l'internet des objets (IoT) [104]étant donné que presque la moitié des publications sur le cyber sécurité des Blockchain sont autour de l'IoT. Cela peut être dû à la prolifération des solutions IoT dans notre vie quotidienne et la demande croissante de cette technologie surtout en médecine et dans le domaine militaire. les opportunités d'améliorer la sécurité de l'Internet des objets sont surtout abordées pour répondre aux besoins suivants :
    3 Authentification des appareils sur le réseau et authentification des utilisateurs finaux sur les appareils eux-mêmes.
    4 Détection des menaces et prévention des logiciels malveillants
    5 Déploiement sécurisé du frimware via la propagation Peer-to-Peer des mises à jour.
- La sécurité des échanges de données et de leur stockage est aussi un axe de recherche très actif et permet de mettre en place les moyens techniques pour assurer l'intégrité des données et empêcher toute altération non autorisées ou destruction volontaire ou accidentelle de l'expédition à la réception au stockage[105]. Cet axe concerne :

6 La sécurité des traitements des données sur le cloud
7 La sécurité de la transmission des données sur une architecture P2P
8 La sécurité des registres de stockage distribués

- La sécurité du réseau : une grande partie du réseau Blockchain se fonde sur une utilisation massive des machines virtuelles et le principe de la containerisation pour le déploiement de l'application Blockchain. Ce réseau virtuel est géré par des logiciels qui ne sont pas sécurisés à 100%. Dans le contexte de la Blockchain, les études de la sécurité du réseau[101] vise plus les sujets liés aux :

  9 Vulnérabilités au niveau des systèmes d'exploitation des VM et la sécurité d'isolement matériel implémenté dans les VM sur lesquels la Blockchain est déployée et stocke ses données sensibles.
  10 Le risque de fuite au niveau des conteneurs : leur code source, leur Bibliothèques, leurs permissions excessives（Fonction des Privilèges CAP_SYS_ADMIN）Même le moteur de conteneur lui-même peut avoir une vulnérabilité de sécurité.
  11 Hybridation des deux concepts VM/conteneur pour offrir une couche d'isolement au conteneur entre l'application et le noyau hôte.

- Données d'utilisateur privées [106] : y compris les paramètres de l'utilisateur final pour les appareils Bluetooth portables et la protection des informations personnelles identifiables échangées avec d'autres parties.

- Navigation et utilité du World Wide Web : assurer la validité du point d'accès Internet sans fil connecté à naviguer vers la bonne page Web via des enregistrements DNS précis utilisé en toute sécurité par les applications Web et communiqué avec les autres via des méthodes sécurisées et chiffrées[94].

3.2 La Sécurité pour la Blockchain

Nous discutons d'abord des exigences de sécurité des transactions en ligne, chacune de ces exigences est ciblée sur un type de vulnérabilités connues. Ensuite, nous décrivons les propriétés de sécurité de base (et inhérentes) de la Blockchain sur la base de sa première implémentation dans Bitcoin, et présentons l'ensemble des propriétés de sécurité et de confidentialité supplémentaires importantes de la Blockchain, qui sont soit présentes dans certains systèmes de Blockchain existantes, soit souhaitées par de nombreuses applications de Blockchain.

3.2.1 Cohérence

Le concept de cohérence dans le contexte de la Blockchain en tant que grand livre mondial distribué fait référence à la propriété selon laquelle tous les nœuds ont le même grand livre en même temps[107]. La propriété de cohérence a soulevé un débat controversé. Certains soutiennent que les systèmes Bitcoin ne fournissent qu'une cohérence éventuelle, ce qui est une cohérence faible. D'autres prétendent que Bitcoin garantit une cohérence forte, et non une cohérence éventuelle[108]. La cohérence à terme est un modèle de cohérence proposé pour les systèmes informatiques distribués en cherchant un compromis entre disponibilité et cohérence. Formellement, cela garantit que toutes les mises à jour des répliques sont propagées de manière paresseuse et que tous les accès en lecture à un élément de données obtiendront finalement la dernière valeur mise à jour si l'élément ne reçoit aucune nouvelle mise à jour[108]. En d'autres termes, la cohérence éventuelle garantit que les données de chaque entrée à chaque nœud du système deviennent cohérentes à terme, et permet ainsi d'obtenir une haute disponibilité et une faible latence au risque de renvoyer des données obsolètes. Avec une cohérence éventuelle, le temps mis par les nœuds du système pour devenir cohérents peut ne pas être défini. Ainsi, la cohérence des données signifie finalement que :

- ✓ il faudra du temps pour que les mises à jour soient propagées vers d'autres répliques ; et
- ✓ si quelqu'un lit à partir d'une réplique qui n'est pas encore mise à jour (puisque les répliques sont éventuellement mises à jour), il existe alors un risque de renvoyer des données obsolètes.

Au sein d'un système de réseau blockchain, le modèle de cohérence forte signifie que tous les nœuds ont le même grand livre en même temps, et pendant le temps où le grand livre distribué est mis à jour avec de nouvelles données, toutes les demandes de lecture/écriture ultérieures devront attendre jusqu'à ce que le commit de cette mise à jour soit fini[100]. En revanche, le modèle de cohérence éventuel signifie que la blockchain à chaque nœud du système devient cohérente à terme, même si certaines demandes de lecture/écriture à la blockchain peuvent renvoyer des données obsolètes. Le principal défi pour une cohérence forte est que le coût des performances (par rapport à la latence/la disponibilité) est trop élevé pour être abordable dans tous les cas. Le principal défi pour une cohérence éventuelle est de savoir comment supprimer l'incohérence qui peut être causée par des données obsolètes[108].

La blockchain de Bitcoin adopte un modèle de cohérence qui cherche un meilleur compromis entre une cohérence forte et une cohérence éventuelle pour atteindre la tolérance de partition (P) et la cohérence (C) avec une disponibilité différée. Dans Bitcoin, les transactions sont regroupées en blocs. Lorsqu'un nœud expéditeur envoie une transaction au réseau blockchain, les nœuds mineurs l'exploitent en l'ajoutant à un bloc avec d'autres transactions non vérifiées et en effectuant un jeu de défi de preuve de travail. Une fois son défi de preuve de travail terminé, un mineur envoie son bloc et sa preuve au réseau pour solliciter les acceptations d'autres nœuds, qui vérifieront toutes les transactions dans le bloc. Les autres nœuds acceptent le bloc en travaillant à la génération du bloc suivant en utilisant le hachage du bloc accepté comme hachage précédent. Le mineur dont le bloc est contenu dans la chaîne la plus longue et qui est le premier à obtenir μ confirmations (les blocs μ sont ajoutés en haut du bloc, et $\mu = 6$ par défaut dans le protocole de consensus Bitcoin) est le gagnant pour l'enchaînement de cette transaction dans le grand livre mondial distribué. Nous pouvons voir le μ paramètre en tant que mécanisme pour fournir une cohérence forte configurable ou paramétrée dans la blockchain.

En résumé, la blockchain est une approche élégante pour résoudre le problème du CAP pour le stockage d'un grand livre distribué dans un système décentralisé. Pour Bitcoin, la blockchain implémente la tolérance de partition (P) tout en prenant en charge la cohérence (C) et la disponibilité (A) sur la blockchain écrêtée avec la plus récente μ blocs ignorés.

En bref, le protocole de consensus accepte une mise à jour de la blockchain (le grand livre mondial distribué) uniquement lorsque le nombre de confirmations reçues par un mineur sur sa solution de défi est égal ou supérieur à µ, ainsi, la disponibilité de la mise à jour est retardée jusqu'au µ les confirmations sont obtenues auprès du réseau. Le protocole read ne lit que la blockchain avec le dernier µ blocs sur la chaîne clippés pour assurer la cohérence forte et la disponibilité en lecture sur la blockchain µ-clippé. Ainsi, certains ont fait valoir que la blockchain dans Bitcoin garantit une cohérence bien plus forte qu'une éventuelle cohérence. Il offre une sérialisation avec une probabilité qui diminue de façon exponentielle avec la latence. D'un autre côté, certaines applications blockchain sont moins averses au risque et peuvent bénéficier d'une garantie de cohérence plus faible pour plus de commodité et de performance. Par exemple, quand µ = 0, cela signifie qu'une confirmation zéro est requise à la fois pour le protocole de consensus et le protocole de lecture. Cela peut être un choix pratique pour ces applications distribuées sans risque. Le blog d'Emin Gün Sirer est un excellent point de départ pour plus de lectures sur ce sujet. De plus, le temps nécessaire pour confirmer une transaction Bitcoin avec le µ La contrainte de cohérence forte peut être d'une lenteur prohibitive pour certaines applications, par exemple 10 minutes en moyenne pour générer un bloc en Bitcoin, et cette latence élevée est aggravée lorsque µ est configuré avec une valeur plus élevée. Récemment, certains efforts de recherche tentent de créer des systèmes de blockchain à débit beaucoup plus rapide et beaucoup plus élevé qui offrent de meilleures garanties que les transactions à 0 confirmation de Bitcoin. PeerCensus étend la blockchain Bitcoin pour prendre en charge une cohérence forte et découpler la création de blocs et la confirmation de transaction[100].

3.2.2 La résistance à la falsification

La résistance à la falsification fait référence à la résistance à tout type de falsification intentionnelle d'une entité par les utilisateurs ou les adversaires ayant accès à l'entité, qu'il s'agisse d'un système, d'un produit ou d'un autre objet logique/physique[109]. La résistance à la falsification de la blockchain signifie que toute information de transaction stockée dans la blockchain ne peut pas être falsifiée pendant et après le processus de

génération de blocs. Plus précisément, dans un système Bitcoin, de nouveaux blocs sont générés par les nœuds de minage. Il existe deux manières possibles de falsifier les informations de la transaction :

(1) les mineurs peuvent tenter de falsifier les informations de la transaction reçue ;

(2) L'adversaire peut tenter de falsifier les informations stockées sur la blockchain. Nous analysons pourquoi de telles tentatives de falsification sont élégamment empêchées par les protocoles de la blockchain dans Bitcoin.

Pour le premier type de falsification, un mineur peut tenter de changer l'adresse du bénéficiaire de la transaction pour lui-même. Cependant, une telle tentative ne peut aboutir, puisque chaque transaction est compressée par une fonction de Hash sécurisée, telle que SHA-256, puis signée par le payeur à l'aide d'un algorithme de signature sécurisé, tel que ECDSA, dans un réseau Bitcoin, et enfin, la transaction 0est envoyé à l'ensemble du réseau pour vérification et approbation via le minage. Ainsi, plusieurs mineurs peuvent recevoir et récupérer la transaction à miner, ce qui se fait de manière non déterministe. Si un mineur modifie des informations de la transaction, il sera détecté par d'autres lorsqu'ils vérifieront la signature avec la clé publique du payeur, car le mineur ne peut pas générer une signature valide sur les informations modifiées sans la clé privée du payeur. Ceci est garanti par la non-falsification de l'algorithme de signature sécurisée[109].

Pour le deuxième type de falsification, un adversaire échouera dans ses tentatives de modifier les données historiques stockées sur la blockchain. Cela est dû aux deux techniques de protection utilisées dans le stockage distribué de la blockchain dans Bitcoin : le pointeur de hachage, et la prise en charge à l'échelle du réseau pour le stockage et la vérification de la blockchain. Plus précisément, si un adversaire veut falsifier les données sur un certain bloc (disons k), la première difficulté rencontrée par l'adversaire est le problème de non-concordance, à savoir, le bloc falsifié k a une valeur de hachage incohérente par rapport au hachage du précédent. Bloc k maintenu dans le bloc k+1. En effet, en utilisant une fonction de hachage avec résistance aux collisions, les sorties de la fonction de hachage résistant aux collisions avec deux entrées différentes seront complètement incohérentes avec une probabilité écrasante, et une telle incohérence peut

être facilement détectée par d'autres sur le réseau. Même si l'adversaire tente de déguiser cette falsification en cassant le hachage du bloc précédent et ainsi de suite le long de la chaîne, cette tentative échouera finalement lorsque la tête de la liste (alias bloc de genèse) est atteinte. De plus, dans la blockchain du réseau Bitcoin, tout le monde a une copie de la blockchain. Il est très difficile pour un adversaire de modifier toutes les copies dans l'ensemble du réseau. En bref, comme chaque transaction en Bitcoin est signée et distribuée sur tous les nœuds du réseau via la blockchain, il est pratiquement impossible de falsifier les données de transaction sans que le réseau le sache, ce qui montre le pouvoir de la foule pour stocker et distribuer la blockchain. Cette propriété est attrayante pour de nombreuses applications. Par exemple, dans le domaine de la santé, la blockchain pourrait aider à créer des pistes d'audit immuables, à maintenir la fiabilité des essais de santé et à préserver l'intégrité des données des patients[100].

3.2.3 Résistance aux attaques DDoS

Une attaque par déni de service est appelée attaque DoS sur un hôte. C'est le type de cyberattaques qui perturbent les services Internet hébergés en rendant la machine hôte ou la ressource réseau sur l'hôte indisponible pour ses utilisateurs prévus. Les attaques DoS tentent de surcharger le système hôte ou les ressources du réseau hôte en inondant de demandes superflues, ce qui retarde par conséquent l'exécution des services légitimes. L'attaque DDoS fait référence à une attaque DoS « distribuée »[101], c'est-à-dire que l'attaque par inondation de trafic entrant vers une victime provient de nombreuses sources disparates réparties sur Internet. Un attaquant DDoS peut compromettre et utiliser l'ordinateur d'une personne pour attaquer un autre ordinateur en tirant parti de la sécurité. Vulnérabilités ou faiblesses. En exploitant un ensemble d'ordinateurs compromis, un attaquant DDoS peut envoyer d'énormes quantités de données à un site Web d'hébergement ou envoyer du spam à des adresses e-mail particulières. Cela rend effectivement très difficile d'empêcher l'attaque par simple brouillage. Sources individuelles une par une. La course à l'armement dépend du taux de réparation de ces nœuds compromis par rapport au taux de réussite des nœuds informatiques compromis dans le réseau.

La principale préoccupation dans une attaque DDoS concerne la disponibilité de la blockchain et est liée à la question de savoir si un attaquant DDoS peut rendre la blockchain indisponible en éliminant un réseau partiel ou entier. La réponse à cette question est non, grâce à la construction et à la maintenance entièrement décentralisées du système blockchain et Bitcoin et au protocole de consensus pour la génération de nouveaux blocs et l'ajout à la blockchain, qui garantit que le traitement des transactions blockchain peut continuer même si plusieurs blockchain les nœuds se déconnectent. Pour qu'un cyber-attaquant réussisse à mettre la blockchain hors ligne, l'attaquant devrait collecter des ressources de calcul suffisantes qui peuvent compromettre une très grande partie des nœuds de la blockchain dans l'ensemble du Bitcoin. Plus le réseau Bitcoin devient grand, plus il est difficile de réussir une attaque DDoS à si grande échelle[101].

3.2.4 Résistance aux attaques à double dépense

L'attaque à double dépense dans le contexte de la blockchain Bitcoin fait référence à un problème spécifique propre aux transactions en monnaie numérique. Notez que l'attaque par double dépense peut être considérée comme un problème de sécurité général en raison du fait que les informations numériques peuvent être reproduites relativement facilement. Plus précisément, avec les transactions d'échange de jetons numériques, telles que la monnaie électronique, il existe un risque que le détenteur puisse dupliquer le jeton numérique et envoyer plusieurs jetons identiques à plusieurs destinataires. Si une incohérence peut survenir en raison des transactions de jetons numériques en double (par exemple, double dépensé le même jeton bitcoin), alors le problème de double dépense devient une menace sérieuse pour la sécurité. Pour éviter les doubles dépenses, Bitcoin évalue et vérifie l'authenticité de chaque transaction à l'aide des journaux de transactions de sa blockchain avec un protocole de consensus. En veillant à ce que toutes les transactions soient incluses dans la blockchain, dans laquelle le protocole de consensus permet à chacun de vérifier publiquement les transactions dans un bloc avant de valider le bloc dans la blockchain globale, en veillant à ce que l'expéditeur de chaque transaction ne dépense que les bitcoins qu'il possède légitimement.[109] De plus, chaque transaction est signée par son expéditeur à l'aide d'un algorithme de signature numérique sécurisé.

Cela garantit que si quelqu'un falsifie la transaction, le vérificateur peut facilement la détecter. La combinaison de transactions signées avec des signatures numériques et de vérification publique des transactions avec un consensus majoritaire garantit que la blockchain Bitcoin peut être résistante à l'attaque de double dépense[93].

3.2.5 Résistance à l'attaque par consensus de la majorité (51%)

Cette attaque fait référence aux risques de tricheries dans le protocole de consensus majoritaire. L'un de ces risques est souvent appelé l'attaque des 51 %, en particulier dans le contexte de la double dépense[102]. Par exemple, l'attaque à 51% peut se produire en présence de mineurs malveillants. Par exemple, si un mineur (utilisateur de vérification) contrôle plus de 50% de la puissance de calcul pour maintenir la Blockchain, le grand livre distribué de toutes les transactions de trading d'une crypto-monnaie. Un autre exemple de l'attaque à 51% peut se produire lorsqu'un groupe de mineurs s'entend pour réaliser un complot, par exemple, en ce qui concerne le décompte des votes des mineurs pour vérification. Si un utilisateur puissant ou un groupe d'utilisateurs complices contrôle la Blockchain, diverses attaques de sécurité et de confidentialité peuvent être lancées, telles que le transfert illégal de bitcoins vers certains portefeuilles cibles, l'annulation de transactions authentiques comme si elles n'avaient jamais eu lieu, etc[100].

3.2.6 Pseudononymat

Le pseudonyme fait référence à un état d'identité déguisée. Dans Bitcoin, les adresses dans la Blockchain sont des hachages de clés publiques d'un nœud (utilisateur) du réseau. Les utilisateurs peuvent interagir avec le système en utilisant leur hachage de clé publique comme pseudo-identité sans révéler leur vrai nom[109]. Ainsi, l'adresse qu'un utilisateur utilise peut être considérée comme une pseudo-identité. Nous pouvons considérer le pseudonyme d'un système comme une propriété de confidentialité pour protéger le vrai nom de l'utilisateur. De plus, les utilisateurs peuvent générer autant de paires de clés (adresses multiples) qu'ils le souhaitent, de la même manière qu'une personne peut créer plusieurs comptes bancaires à sa guise. Bien que le pseudonyme puisse atteindre une

forme faible d'anonymat au moyen des clés publiques, il existe toujours des risques de révéler les informations d'identité des utilisateurs[105].

3.2.7 Dissociabilité

La dissociation fait référence à l'incapacité d'établir la relation entre deux observations ou deux entités observées du système avec un niveau de confiance élevé[110]. L'anonymat fait référence à l'état d'être anonyme et non identifié. Bien que la blockchain dans Bitcoin assure le pseudononymat en offrant la pseudo-identité comme support de l'anonymat de l'identité d'un utilisateur, elle ne fournit pas aux utilisateurs la protection de la dissociation pour leurs transactions. Intuitivement, l'anonymat complet d'un utilisateur ne peut être protégé qu'en garantissant à la fois le pseudonyme et la dissociation si l'utilisateur utilise toujours sa pseudo-identité pour interagir avec le système. En effet, la dissociation rend difficile le lancement d'attaques d'inférence de dés-anonymisation, qui relient les transactions d'un utilisateur ensemble pour découvrir la véritable identité de l'utilisateur en présence de connaissances de base[110]. Concrètement, dans les systèmes de type Bitcoin, un utilisateur peut avoir plusieurs adresses pseudonymes[103]. Cependant, cela ne garantit pas un anonymat parfait pour les utilisateurs de blockchain, car chaque transaction est enregistrée sur le grand livre avec les adresses de l'expéditeur et du destinataire, et est traçable librement par toute personne utilisant les adresses associées de son expéditeur et de son destinataire. Ainsi, n'importe qui peut relier la transaction d'un utilisateur à d'autres transactions impliquant ses comptes par une simple analyse statistique des adresses utilisées dans les transactions Bitcoin. Par exemple, par analyse sur le compte d'un expéditeur, on peut facilement connaître le nombre et le montant total des bitcoins sortant ou entrant sur ce compte[93]. Alternativement, on peut lier plusieurs comptes qui envoient/reçoivent des transactions à partir d'une adresse IP. Plus sérieusement, un utilisateur peut perdre son anonymat et donc sa confidentialité pour toutes les transactions associées à son adresse Bitcoin si le lien entre son adresse Bitcoin et l'identité réelle de l'utilisateur est exposé. De plus, étant donné la nature ouverte de la blockchain publique, n'importe qui peut tenter d'effectuer ce type d'attaque de dés-anonymisation silencieusement et secrètement sans que l'utilisateur cible se rende même

compte qu'elle est attaquée ou que sa véritable identité a été compromise. Par conséquent, la mise en œuvre de la blockchain dans Bitcoin n'obtient que le pseudonyme, mais pas la dissociabilité et donc pas l'anonymat complet défini par le pseudonyme avec dissociabilité[100].

IV- Conclusion

Ainsi la Blockchain a le potentiel de transformer d'une manière radicale nos façons d'échanger l'information, de produire des biens et des services, d'investir et de financer nos entreprises. Son appréhension ainsi que son application est un défi pour moderniser notre économie.

L'écosystème de la Blockchain se développe rapidement avec l'augmentation des investissements et des intérêts de l'industrie, du gouvernement et des universités.

Néanmoins, l'adoption de cette technologie ne manque pas de difficultés et d'obstacles qui empêchent sa mise en pratique et provoque encore des inquiétudes chez les entrepreneurs et les investisseurs. Des problématiques sérieuses mais qu'on ne peut pas éluder du fait que cette innovation offre des potentialités et des atouts vivement demandées. On peut projeter ces enjeux sur trois échelles essentielles, d'un côté la sûreté des outils techniques et leur capacité à garantir les différentes propriétés que promet cette technologie, d'un autre côté, le volet fonctionnel lié aux domaines métiers de son application y compris les enjeux financiers et économiques. Puis, et finalement la portée juridique reflétant la fiabilité revendiquée par la technologie pour protéger l'ordre publique, contrôler le consommateur ainsi éliminer les usages frauduleux.

Chapitre 6 |

Protocole pour la Blockchain Préservant une Partielle Confidentialité & Transparence

Sommaire

I- Introduction

Dans le cadre de cette contribution, nous nous focalisons sur l'aspect technique de la technologie Blockchain et plus précisément le volet de la confidentialité des transactions qui fait partie des défis majeurs pour l'adoption de ce protocole[111]. Notre solution consiste à chercher un compromis entre la Transparence et la Confidentialité des données qui est un dilemme de l'adoption de la Blockchain. Elle permet d'introduire une nouvelle notion de confidentialité que nous avons nommée « confidentialité partielle ». Par la suite, elle l'applique sur les transactions échangées tout en assurant le processus de leur validation par les différents nœuds de la Blockchain. Grâce à l'utilisation de fonctions de hachage et de signature numérique, ce protocole garantit également l'intégrité et l'authentification au sein de son processus de validation.

Pour présenter ce travail, nous aborderons d'abord la motivation derrière ce travail, puis nous présentons l'état de l'art à propos des différentes approches actuelles qui abordent la même problématique. Ensuite, nous expliquerons en détail les différentes étapes de ce processus, ses avantages et les points à améliorer dans le cadre de nos perspectives.

II- Motivation

La divulgation publique des informations dans une Blockchain, que ça soit lors du processus du consensus ou au niveau des registres distribués, présente un risque important pour la sécurité des entreprises. Pour remédier à cette problématique, la première solution adoptée par certaines entreprises est le déploiement de cette technologie dans un réseau fermé et restreint, ce que l'on désigne par Blockchain privée[92]. Il s'agit, en fait, d'un protocole Blockchain installé au sein d'un environnement bien maitrisé et qui permet d'échanger les transactions uniquement entre des parties autorisées. Cependant, ce type d'architecture nécessite une autorité centrale pour maintenir la gestion des accès et les autorisations entre les participants[97]. Chose qui peut nous induire à l'ancienne attitude de centralisations afin de gérer ces autorisations, au lieu de la décentralisation qui représente l'un des potentialités majeures de la Blockchain[112]. Manipuler une Blockchain Privée pour des fins personnelles est très probable et dépend de l'honnêteté

de l'autorité centrale.

Privée ou publique, la sécurité dans la Blockchain reste toujours une question ouverte pour laquelle il faut faire intervenir plusieurs paramètres pour pouvoir englober au même temps les différents enjeux de cette technologie[113], à savoir :

2.1 Enjeu de la Confidentialité

La confidentialité des données de la Blockchain fait référence à la propriété selon laquelle la Blockchain peut assurer la confidentialité de toutes les données ou de certaines données sensibles qui y sont stockées[111]. Bien que la Blockchain ait été conçue à l'origine comme un grand livre mondial distribué pour le système de monnaie numérique Bitcoin, son champ d'application potentiel est beaucoup plus large que les monnaies virtuelles. Par exemple, la Blockchain peut être utilisée pour gérer les contrats intelligents[114], les œuvres protégées par le droit d'auteur, la numérisation des registres commerciaux ou organisationnels. Sans surprise, une propriété de sécurité souhaitable commune à toutes les applications Blockchain est la confidentialité des informations de transaction, telles que le contenu de la transaction (par exemple, les montants des transactions en Bitcoin) et les adresses. Malheureusement, cette propriété de sécurité n'est pas prise en charge dans les systèmes Bitcoin. Dans Bitcoin, le contenu et les adresses de la transaction sont visibles publiquement, même si le pseudonyme est utilisé comme adresse de l'expéditeur et du destinataire d'une transaction au lieu de l'identité réelle[92]. Nous supposons que la capacité de garder le contenu de la transaction privé aidera à réduire le risque de lien entre le pseudonyme et l'identité réelle de l'utilisateur. Ceci est essentiel pour promouvoir le partage basé sur le besoin de savoir au lieu d'être visible publiquement de l'ensemble de la Blockchain. De plus, les systèmes de Blockchain, qui utilisent des contrats intelligents pour mettre en œuvre des transactions complexes, comme Ethereum, nécessitent

(1) que les données de chaque contrat et le code qu'il exécute sur les données soient publiques
(2) que chaque mineur émule l'exécution de chaque contrat.

Cela conduira à la fuite d'informations sur les utilisateurs[101]. Par exemple, un utilisateur configure un contrat intelligent pour transférer une certaine quantité d'ETH à un autre utilisateur à un certain moment. Si un adversaire a des informations de base sur l'une des deux parties, cet adversaire peut exposer et lier cette partie à sa véritable identité. Par conséquent, il est essentiel de concevoir et de mettre en œuvre des mécanismes de protection plus solides pour les contrats intelligents préservant la confidentialité. En résumé, la recherche sur la confidentialité des données au cours des dernières décennies a montré les risques de fuite de la vie privée dus à diverses attaques par inférence, qui lient des données de transaction sensibles et/ou un pseudonyme à la véritable identité des utilisateurs réels, même si seul un pseudonyme est utilisé [113]. Une telle fuite de confidentialité peut conduire à une violation de la confidentialité des informations de transaction. Ainsi, la confidentialité et la vie privée constituent un défi majeur pour la blockchain et ses applications qui impliquent des transactions sensibles et des données privées[100]. Nous consacrerons la section qui suit pour discuter de techniques principales qui peuvent aider à améliorer la confidentialité des données et la confidentialité des transactions sur la blockchain.

2.2 Enjeu de l'évolution : Calcul hors chaîne

La technologie Blockchain actuelle atteint un consensus sur la chaîne de telle sorte qu'un contrat intelligent s'exécute sur chaque nœud de cette Blockchain. Il n'y a pas d'autorité finale dans un réseau décentralisé, comme Bitcoin, de sorte que chaque mineur doit valider chaque transaction avant de l'accepter et de l'enregistrer sur la Blockchain. En d'autres termes, il n'y a aucun moyen de vérifier le résultat sans l'exécuter sur un réseau sans confiance. Le problème le plus important est le montant des frais de transaction pour chaque transaction. Plus le réseau se développe, la quantité de calculs dans le réseau combiné peut facilement dépasser la limite de gaz provoquant plus de fourches et soulevant plus de problèmes de sécurité[99].

Le calcul hors chaîne résout ce problème en déplaçant le travail de calcul hors du réseau publique et en vérifiant le résultat en publique, uniquement en cas de litige[100]. La clé du calcul hors chaîne est d'avoir une propriété vérifiable dans la tâche de calcul

externalisée. Cette propriété résout efficacement le problème d'évolutivité sur les réseaux Blockchain.

2.3 Enjeu des Dataset Centralisés

Les grandes sociétés Internet collectent des données sur les activités en ligne des utilisateurs pour former des systèmes de recommandation qui prédisent les intérêts et les actions futurs des clients. Les données de santé de différents hôpitaux et organisations gouvernementales peuvent être utilisées pour produire de nouveaux modèles de diagnostic, tandis que les sociétés financières et les réseaux de paiement peuvent combiner l'historique des transactions, les données des commerçants et les informations sur les titulaires de compte pour former des moteurs de détection de fraude plus précis.

Toutes ces données doivent être agrégées à un niveau individuel, mais les bénéfices réalisés à partir des données n'ont jamais été partagés avec les contributeurs de données. Le règlement général sur la protection des données (RGPD) a clairement indiqué que les détenteurs de données ont la responsabilité de faciliter la portabilité des données. Pour vraiment appliquer cette réglementation, nos données ne doivent pas être stockées sur le disque dur de l'entreprise car elles font face au risque inévitable d'être violées.

Malheureusement, les entreprises ont trouvé difficile de protéger leurs données critiques. La convention avec la conservation des données est mal formée :

1. Données à risque centralisées

2. Distribution de valeur injuste - les utilisateurs individuels ne sont jamais payés pour leurs données

3. La domination établie des joueurs crée un fossé hostile qui dissuade les nouveaux challengers innovants de concourir.

4. Les individus talentueux ne peuvent pas construire un bon modèle sans rejoindre l'entreprise

Nous avons besoin d'un système de partage de données décentralisé qui permet à chacun de construire des modèles dessus, de récompenser ceux qui y contribuent et en même

temps de sécuriser les données[115].

2.4 Enjeu de la Sécurité des échanges

Il y a une tension qui surgit lorsque des individus, des entreprises ou des gouvernements traitent des données sensibles. D'une part, la science des données est une composante fondamentale de l'ère de l'information. On entend souvent dire que les données sont le nouveau pétrole : il y a une immense valeur noncière et sociale à acquérir des données brutes. Un article récent de Google confirme un fait bien connu que la taille des données est positivement corrélée aux performances du modèle[116], quelle que soit la qualité du modèle. Il explique la motivation des entreprises à collecter une grande quantité de données utilisateur. D'un autre côté, plus il y a de valeur à gagner en rassemblant des données, plus tout le monde est devenu prudent quant au partage de données, car les violations de données peuvent causer des dommages financiers, juridiques et politiques. Cela semble être un compromis central : nous pouvons partager des données afin d'acquérir de nouvelles connaissances qui profitent à la société dans son ensemble, ou nous pouvons isoler les données dans des silos protégés qui protègent notre vie privée[117].

2.5 Enjeu de la Transparence

L'un des principaux arguments de vente de la technologie Blockchain est son potentiel à apporter une plus grande transparence aux marchés financiers[8]. À la base, la Blockchain est un moyen d'effectuer des transactions sécurisées, vérifiables et enregistrables en ligne sans une partie centralisée. Comme un consensus est atteint en publique et que le résultat est auditable, la Blockchain est censée être transparente et accessible au publique par conception[8]. Cependant, une transparence totale peut être problématique dans le monde réel[118]. Imaginez que vous ayez ouvert un compte dans une banque et que vous découvriez bientôt que ses registres sont publiques, où n'importe qui peut accéder à l'historique des transactions de votre compte (et d'autres). Ou, si vous avez effectué une transaction dans un magasin en utilisant Bitcoin et que le caissier serait en mesure de

savoir combien d'argent se trouve sur votre compte. Malheureusement, la demande croissante des technologies Blockchain pose le défi de protéger les utilisateurs contre le vol de propriété intellectuelle et d'autres attaques

En conclusion, la sécurité, la vie privée et la confidentialité sont les principaux problèmes qui freinent l'adoption de la Blockchain. Peu de solutions ont été proposées jusqu'à ce jours permettant d'apporter une réponse sur l'un de ces enjeux sans impacter les autres[106]. Nous essayons à travers la solution qu'on propose dans cette contribution de trouver un compromis entre ces différents enjeux et ouvrir notre blockchain sur un environnement partiellement publique!

III- Etat de l'art

Une fonction de confidentialité qui permet de couvrir tous les enjeux précédemment discutés n'est pas encore fournie avec la technologie Blockchain d'aujourd'hui. La plupart des travaux qui visent à résoudre le problème de la vie privée dans la Blockchain, sont davantage axés sur ce que l'on appelle par « le calcul sécurisé »[100]. En effet, ce type de solutions repose sur des techniques permettant la réalisation de calculs sur des données sans dévoiler la partie secrète de chaque transaction. Pour ce faire, ces travaux se basent essentiellement sur une ou deux des approches de chiffrement suivantes :

- L'approche ABE (Attribute-Based Encryption)
- L'approche Secure Multi Part Calculation (SMPC)
- L'approche de Chiffrement Homomorphe (HE)
- La preuve non interactive à connaissance nulle (NIZK)

Dans cette section, nous essaierons d'aborder ces approches en décrivant certains travaux effectués dans le cadre de chaque approche.

3.1 Chiffrement homomorphe (HE)

Le chiffrement homomorphe (HE) est une cryptographie puissante[107]. Il peut effectuer certains types de calculs directement sur le texte chiffré et garantir que les opérations

effectuées sur les données chiffrées, lors du déchiffrement des résultats calculés, généreront des résultats identiques à ceux effectués par les mêmes opérations sur le texte en clair. Il existe plusieurs crypto-systèmes partiellement homomorphes ainsi que des systèmes totalement homomorphes[119]. On peut utiliser des techniques de chiffrement homomorphes pour stocker des données sur la Blockchain sans modification significative des propriétés de la Blockchain. Cela garantit que les données sur la Blockchain seront chiffrées, répondant aux problèmes de confidentialité associés aux Blockchain publiques. L'utilisation d'une technique de chiffrement homomorphe offre une protection de la vie privée et permet un accès facile aux données chiffrées via la Blockchain publique à des fins d'audit et à d'autres fins, telles que la gestion des dépenses des employés[120]. Les contrats intelligents Ethereum fournissent un chiffrement homomorphe sur les données stockées dans Blockchain pour plus de contrôle et de confidentialité.

3.2 Chiffrement basé sur les attributs (ABE)

Le chiffrement basé sur les attributs (ABE) est une méthode cryptographique dans laquelle les attributs sont les facteurs de définition et de régulation du texte chiffré à l'aide de la clé secrète d'un utilisateur[107]. On peut déchiffrer les données chiffrées à l'aide de la clé secrète de l'utilisateur si ses attributs sont en accord avec les attributs du texte chiffré[121]. La résistance à la collusion est une propriété de sécurité importante de l'ABE. Il garantit que lorsqu'un utilisateur malveillant s'entend avec d'autres utilisateurs, il ne peut accéder à d'autres données que celles qu'il peut déchiffrer avec sa clé privée.

Le concept de chiffrement basé sur les attributs a été proposé en 2005 avec une seule autorité. Depuis lors, un certain nombre d'extensions ont été proposées à l'ABE de base, y compris ABE avec plusieurs autorités pour générer conjointement les clés privées des utilisateurs[121], les schémas ABE qui prennent en charge les prédicats arbitraires.

Le chiffrement basé sur les attributs est très puissant, mais peu d'applications le déploient à ce jour en raison du manque de compréhension des concepts de base et d'une mise en œuvre efficace. ABE n'a pas encore été déployé sous quelque forme que ce soit sur une Blockchain pour un fonctionnement en temps réel à ce jour. En 2011, un schéma ABE décentralisé a été proposé pour utiliser ABE sur une Blockchain[122]. Par exemple, sur

une Blockchain, les autorisations pourraient être représentées par la propriété de jetons d'accès. Tous les nœuds du réseau, auxquels un certain jeton leur a été attribué, auront accès aux droits et privilèges spéciaux associés au jeton. Le jeton fournit un moyen de suivre qui possède certains attributs et un tel suivi doit être effectué de manière algorithmique et cohérente par l'entité d'autorité qui distribue le jeton. Les jetons peuvent être considérés comme des badges qui représentent des attributs ou des qualifications, et doivent être utilisés comme quantificateurs non transférables de réputation ou d'attributs[123].

Dans [124] , il est montré qu'il n'y a pas besoin d'une autorité fixe pour effectuer un chiffrement basé sur les attributs. Il est possible d'avoir plusieurs autorités dans un réseau décentralisé et d'accomplir le même accomplissement. Par exemple, s'appuyer sur des témoins pour le rôle de ces autorités peut être possible dans une Blockchain, avec des technologies, récemment rendues possibles, telles que Steemit[125], Storj, IPFS [126], SAFE Network. Cependant une mise en œuvre du chiffrement basé sur les attributs utilisant une approche Blockchain reste un défi ouvert.

3.3 Calcul multi-parties sécurisé

Le modèle de calcul multipartite (MPC) définit un protocole multipartite pour leur permettre d'effectuer certains calculs conjointement sur leurs entrées de données privées sans violer leur confidentialité d'entrée, de sorte qu'un adversaire n'apprend rien sur l'entrée d'une partie authentique mais le sortie du calcul conjoint[127]. Andrew Yao a formellement défini le calcul sécurisé à deux parties en 1982 et l'a généralisé en 1986[128] pour le problème des millionnaires. Goldreich et al. ont proposé une généralisation du calcul bipartite au calcul multipartite, en supposant que toutes les entrées du calcul et des preuves à connaissance nulle font partie du partage secret[129]. Cette généralisation a servi de base à de nombreux protocoles MPC ultérieurs et de plus en plus efficaces. Le succès de l'utilisation de MPC dans le vote distribué, les enchères privées et la récupération d'informations privées en a fait une solution populaire à de nombreux problèmes du monde réel. Le premier déploiement à grande échelle de MPC a eu lieu en 2008 pour un problème d'enchères réel au Danemark.

Ces dernières années, MPC a été utilisé dans les systèmes de blockchain pour protéger la vie privée des utilisateurs. Andrychowicz et al. Ont conçu et mis en œuvre des protocoles de calcul multipartites sécurisés sur le système Bitcoin en 2014[130]. Ils ont construit des protocoles pour des loteries multipartites sécurisées sans aucune autorité de confiance. Leurs protocoles sont en mesure de garantir l'équité pour les utilisateurs honnêtes, quel que soit le comportement des malhonnêtes. Si un utilisateur viole ou interfère avec le protocole, il devient perdant et ses bitcoins sont transportés vers les utilisateurs honnêtes.

Une plate-forme de calcul SMP décentralisée, appelée Enigma, est proposée en 2015 par Zyskind et al. [131]. En utilisant une version avancée du calcul SMP, Enigma utilise un schéma de partage de secret vérifiable pour garantir la confidentialité de son modèle de calcul. En outre, Enigma code les données secrètes partagées à l'aide d'une table de hachage distribuée modifiée pour un stockage efficace. De plus, il exploite une blockchain externe en tant qu'enregistrement d'événements résistant à la corruption et régulateur du réseau Peer to Peer pour la gestion des identités et le contrôle d'accès. Semblable au système Bitcoin, Enigma fournit un contrôle et une protection autonomes des données personnelles tout en éliminant la nécessité et la dépendance d'un tiers de confiance.

3.4 Preuve non interactive à connaissance nulle (NIZK)

Une autre technologie cryptographique qui possède de puissantes propriétés de préservation de la vie privée est les preuves à connaissance nulle, proposées au début des années 1980 [132]. L'idée de base est qu'une preuve formelle peut être formulée pour vérifier qu'un programme exécuté avec une entrée connue de manière privée par l'utilisateur peut produire une sortie publiquement ouverte sans divulgation d'aucune autre information. En d'autres termes, un certificateur peut prouver à un vérificateur qu'une affirmation est exacte sans fournir aucune information utile au vérificateur. En tant que variante des preuves à connaissance nulle, avec la variante non interactive des preuves à connaissance nulle, appelée NIZK, on peut atteindre la connaissance informatique zéro sans exiger que le certificateur et le vérificateur interagissent du tout, à condition que le certificateur et le vérificateur partagent une chaîne de référence commune [133]. Dans

une application blockchain, tous les soldes des comptes sont chiffrés et stockés dans la chaîne. Lorsqu'un utilisateur transfère de l'argent à un autre utilisateur, il peut facilement prouver qu'il dispose d'un solde suffisant pour le transfert avec des preuves à connaissance nulle, sans révéler le solde du compte. Une autre variante est la preuve zéro-connaissance Succinct Non-interactive ARgument of Knowledge (zk-SNARK), introduite en 2012 par Bitansky et ses coauteurs [134] et sert de colonne vertébrale au protocole Zcash . Ce dernier utilise zk-SNARK[135] pour vérifier les transactions tout en protégeant la confidentialité des utilisateurs. Récemment, le groupe Zcash a amélioré le langage contractuel Ethereum pour fournir efficacement la vérification des preuves zk-SNARK. Plus précisément, ils ont adopté une précompilation snark-verify (comme un opcode) vers un fork de "Parity" qui utilise lib-snark pour vérifier les preuves génériques. Ils ont également utilisé le nouveau vérificateur zk-SNARK[136] pour appliquer un contrat de mélange de pièces original, qui adopte une version simplifiée de Zerocash, un protocole académique dont l'implémentation est utilisée pour construire Zcash. Ainsi, il s'appelle « baby » ZoE, acronyme de Zerocash over Ethereum. Le contrat permet à un utilisateur de stocker des montants discrets (unités d'ETH) en ajoutant un « numéro de série » en tant qu'engagement dans un arbre Merkle, qui est maintenu par le contrat.

Certes, toutes les approches qu'on vient de citer proposent des solutions très innovantes pour assurer la confidentialité des données sensibles, cependant, comme nous pouvons le constater à travers la description de chacune, ces approches restent néanmoins limitées à des cas spécifiques et ne peuvent pas être applicable sur tout type de consensus[137]. Autrement dit, elles s'appliquent plus dans les cas où le processus de validation d'une transaction nécessite le calcul d'une ou plusieurs opérations arithmétiques[100]. Donc, dans le cas d'un consensus basé sur un smart contrat procédural et purement fonctionnel, l'application de ce type de solutions ne sera pas possible sinon elle coutera chère.

D'autres inconvénients peuvent entraver l'utilisation ce type d'algorithme comme leur lenteur et leur coût en termes de consommation des ressources physiques [124]. Ceci est dû au fait que les calculs/vérifications qu'elles utilisent, ne se font pas directement sur la donnée brute. D'autre part, ces algorithmes prennent en compte une quantité restreinte de données et ne peuvent pas supporter une grande volumétrie en entrée[111] . Et finalement,

ces approches souffrent toujours de leur manque de maturité et de leurs complexités liées principalement à la difficulté de leur implémentation et leur mise en place[107].

L'approche de chiffrement qu'on propose dans ce travail est générique et applicable sur n'importe quel type de consensus. A l'envers des solutions présentées ci-dessus, notre modèle, propose plus de fluidité et permet d'exploiter les algorithmes du chiffrement symétrique qui ont largement prouvé leurs robustesses et leurs performances dans le domaine. De plus, elle intègre d'autres outils de sécurité comme les fonctions de hachage et les signatures numériques. La flexibilité de ce protocole ne concerne pas seulement les outils de sécurité misent en place, mais aussi au niveau de la distribution des rôles à chaque transaction, chose qui prévient la vulnérabilité du système.

IV- Contribution

4.1 Définition : Confidentialité partielle :

Ce travail présente un nouveau concept de confidentialité qui est la confidentialité partielle. Comme son nom l'indique ce concept vise à assurer la propriété de la confidentialité juste sur une partie des données jugées être sensibles. De ce faire, l'algorithme de la confidentialité partielle requiert une indentification préalable des informations sensibles, ensuite, il procède aux processus de chiffrement ordinaire pour chiffrer ces données et les réintégrer à la fin de l'algorithme dans leur contexte de début.

Pour mieux assimiler cet algorithme, ci-dessous le détail des opérations effectuées au niveau de chaque étape :

- ✓ **Etape** 1 – Identification : La première étape de ce système consiste à identifier les informations sensibles dans le texte à partager et ceci en utilisant un mécanisme de séparation avant et après chaque partie.

- ✓ **Etape** 2 - Extraction : La deuxième étape consiste à extraire les informations sensibles du texte à partager en se référant au séparateur utilisé pour ce protocole. Ensuite, on les regroupe dans un bloc appart qu'on appelle black-blocks (BBs) et qui fera objet du chiffrement dans l'étape suivante.

- ✓ **Etape** 3 – Chiffrement : Le chiffrement est effectué sur le black-blocks (BBs) contenant les données sensibles en utilisant un algorithme de chiffrement symétrique dont la clé secrète est celle des entités autorisées à valider la transaction ou la partie du smart-contrat

- ✓ **Etape** 4 – Reconstruction : L'étape de la reconstruction consiste à intégrer les différents octets des BBs chiffrés dans leurs positions initiales du texte clair en se référant une autres fois au séparateur mis en place lors de l'étape de l'identification.

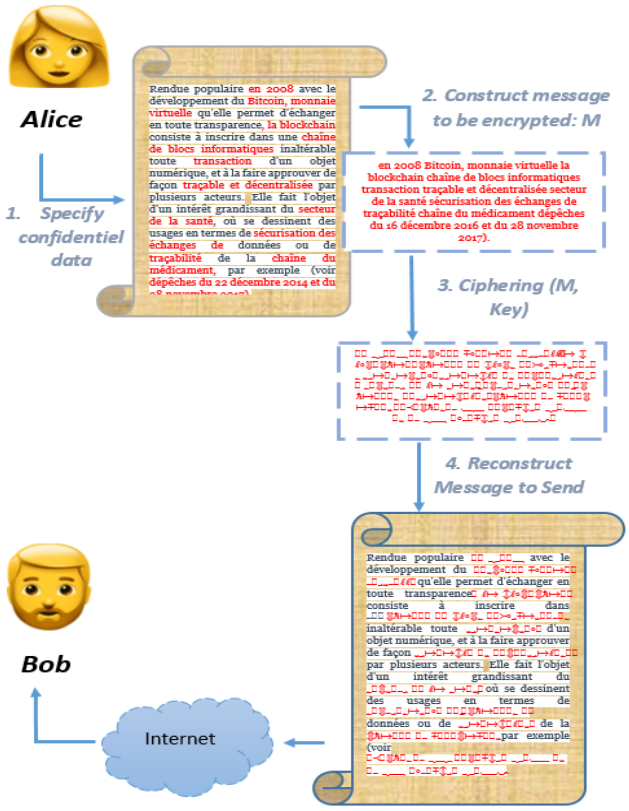


Figure 62: Principe de la Confidentialité Partielle

Le nouveau protocole PPCT propose de garder une partie des informations confidentielle et laisser les autres informations non sensibles transparentes et lisibles par tout le monde. Il vise à utiliser le principe de la confidentialité partielle afin de pouvoir partager les

transactions d'une manière publique tout en gardant les données sensibles secrètes sauf aux yeux des nœuds autorisés. Cette solution favorise la parallélisassions de l'activité de vérification des transactions pour garantir la sécurité du système en séparant les taches entre les différents participants du réseau distribué.

4.2 Groupe de Confiance de la Transaction : TTG

Chaque participant/organisme identifie en avance son groupe de confiance avec lesquels il doit assurer la validité des informations échangées en toute transparence. Ce groupe de confiance peut être considéré comme un sous réseau virtuel et privé de notre réseau distribué dont l'échange des clés cryptographique est préalablement effectué à l'extérieur de la Blockchain.

4.3 Socle de transaction

Dans une Blockchain plusieurs types de transactions peuvent circuler et échanger entre les différents nœuds du réseau, chaque transaction reflète une fonctionnalité bien précise dans le processus à digitaliser. Il peut présenter un achat, un virement, une vérification de contrat, le résultat d'un diagnostic, ou autres. Peu importe le type de la transaction, dans notre système, nous exigeons de préciser la structure de toutes les transactions possibles et leurs donner un modèle bien défini qu'on nomme 'Socle de Transaction'. Le Socle de chaque transaction doit distinguer entre les informations publiques nécessaires pour la validation publique et celles qui peuvent rester secrètes.

4.4 Application de la Confidentialité partielle

Comme nous l'avons présentée auparavant, la confidentialité partielle est une technique qu'on vient de définir afin d'assurer le caractère privé d'un ensemble d'informations incluses dans un modèle de données destiné au publique. Pour se faire, il faut passer par la séparation de ces deux catégories de données, chiffrer la partie sensible en se basant

sur un outil cryptographique de chiffrement et finalement réintégrer cette partie dans le modèle en question.

4.5 Validation privée

Afin de valider la transaction en cours, chaque nœud du groupe de confiance procède au déchiffrement partiel du modèle, calcule le haché du contenu en clair puis passe à la réalisation de leurs taches nécessaires pour leur validation.

Les validateurs envoient leurs réponses signées et chiffrées en utilisant le haché du texte clair comme clé de chiffrement, on le nomme également « clé de validation publique Kv »

Il est à noter que chaque nœud validateur possède un socle aussi pour sa réponse et utilise sa propre signature pour signer sa décision. Le tout est chiffré en utilisant le haché du clair. Cette opération de chiffrement a un rôle très important dans ce protocole de sécurité. D'un côté, elle permet d'assurer la crédibilité de la décision du validateur du fait qu'il a eu le bon haché et par conséquent a pu se baser sur les bonnes informations contenues dans le modèle en clair. D'autre part, à travers ce chiffrement on arrive à protéger le groupe contre l'infiltration des nœuds malveillants afin de ne pas impacter la décision finale de la transaction, et par conséquent, on assure la neutralité de cette décision. Cette opération est considérée comme une enveloppe fermée destinée à être diffusée sur le réseau.

4.6 Validation publique

Une fois recevoir les enveloppes des différents validateurs du groupe de confiance, une notification est envoyée à l'initiateur de la transaction afin d'envoyer le haché du clair dans la Blockchain. La transaction définitive est construite en se basant sur le modèle, le haché et les enveloppes du groupe de confiance.

La transaction est ajoutée au bloc de la Blockchain et envoyée au réseau publique pour validation définitive. Le réseau publique ouvre les enveloppes de chaque validateur en

effectuant l'opération du déchiffrement avec la clé Kv qui représente le haché du clair, Puis compare les résultats et réalise ses vérifications globales pour valider l'opération.

Si les validateurs publiques arrivent à déchiffrer le message avec le haché du clair cela signifie que le validateur privé avait bien le bon contenu du modèle que l'initiateur de la transaction et donc sur la base des différentes réponses le publique valide la transaction finale.

Une fois réussir le consensus de la validation définitive, le bloc est stocké et ajouté au registre via les mécanismes ordinaires de la Blockchain.

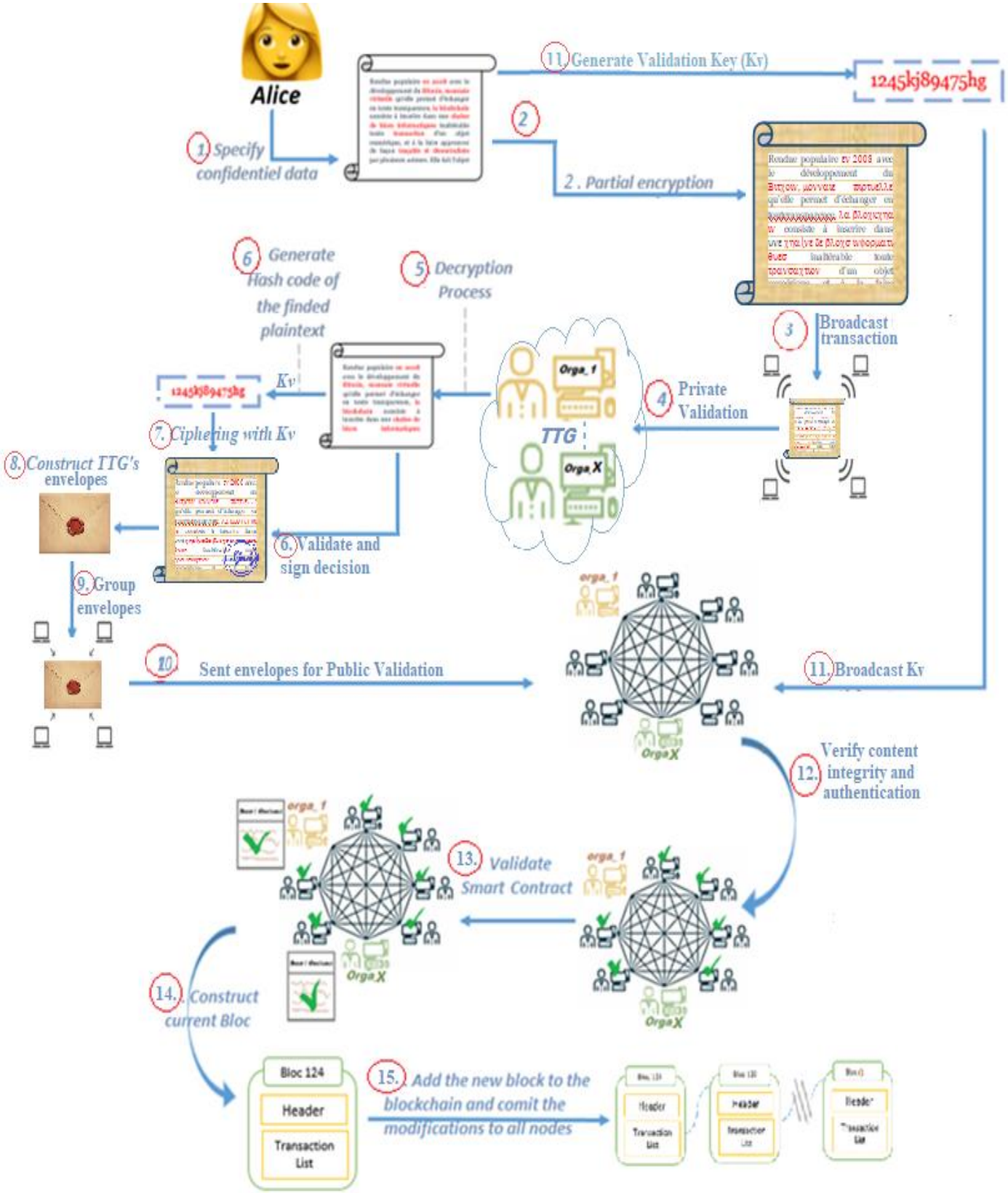


Figure 63 : Protocole Préservant la Confidentialité et la Transparence (PPCT)

V- Cas d'utilisation : Remboursement des soins

Dans le domaine de la santé ou la confiance tient une place dominante et décisive, l'application de la Blockchain peut ouvrir des horizons et perspectives très prometteuses. le désir d'une traçabilité parfaite dans un environnement de partage et de stockage des informations d'une manière assez transparente, tout en assurant un degré de sécurité raisonnable pour une vraie maîtrise des données qui y circulent. Ce champ d'application qui exige la coexistence de ces deux qualités, en apparence antagonistes, est très favorable pour la mise en production de notre solution. Appréhender ces processus de la santé dans une Blockchain est d'une grande utilité. En fait, elle permet de réduire leur complexité en facilitant la gestion des relations entre différents systèmes d'information hétérogènes tout en garantissant un échange d'informations sécurisé et transparent. Cette Blockchain regroupe l'ensemble du corps médical, les patients, les assurances, les centres de radiologies, les laboratoires, les pharmacies, les centres de Kinésithérapie, etc. Bref, tous les professionnels dans le domaine de la santé doivent se réunir autour de cette Blockchain, chacun de son angle, pour veiller au bon déroulement de différentes normes misent en place au service de la santé. Ces apports en termes de : confiance, sécurité, simplification, et parallélisassions des procédures de vérifications se traduisent immédiatement et non seulement en un gain économique en temps et en argent mais aussi à une amélioration énorme de la qualité de vie des populations.

On peut imaginer un cas d'utilisation simple et très récurrent dans notre vie quotidienne et il s'agit de la procédure de remboursement des soins par les assurances maladies. L'automatisation de remboursement des soins est une tâche assez complexe car elle dépend de la crédibilité et l'engagement de toutes les entités dans le domaine de la santé. Le déroulement normal de ce processus de remboursement exige avoir un dossier médical complet contenant toutes les justificatives sur les consultations, les diagnostics, les soins et les traitements réalisés. Ces preuves doivent être signés et validés par les différents interlocuteurs du domaine. L'ensemble de ces papiers est ensuite déposé chez l'assureur qui à son tour réalise ses vérifications et valide le dossier pour un remboursement en fonction du contrat d'assurance maladie de l'assuré. La communication entre les différents interlocuteurs d'un dossier médical est réalisée par le patient. Nous partons sur

ce même principe et nous proposons le scénario modélisé dans le diagramme de séquence suivant :

Patient

- identifie ses interlocuteurs dans la blockchain
- remplit son dossier médical
- Demande de la validation de son dossier

Professionnel de la santé

- Validation du dossier médical par les interlocuteurs

Assurance :

- Validation du dossier médical par l'assurance
- Remboursement du dossier
- Validation de la transaction de remboursement par le réseau
- Stockage de la transaction

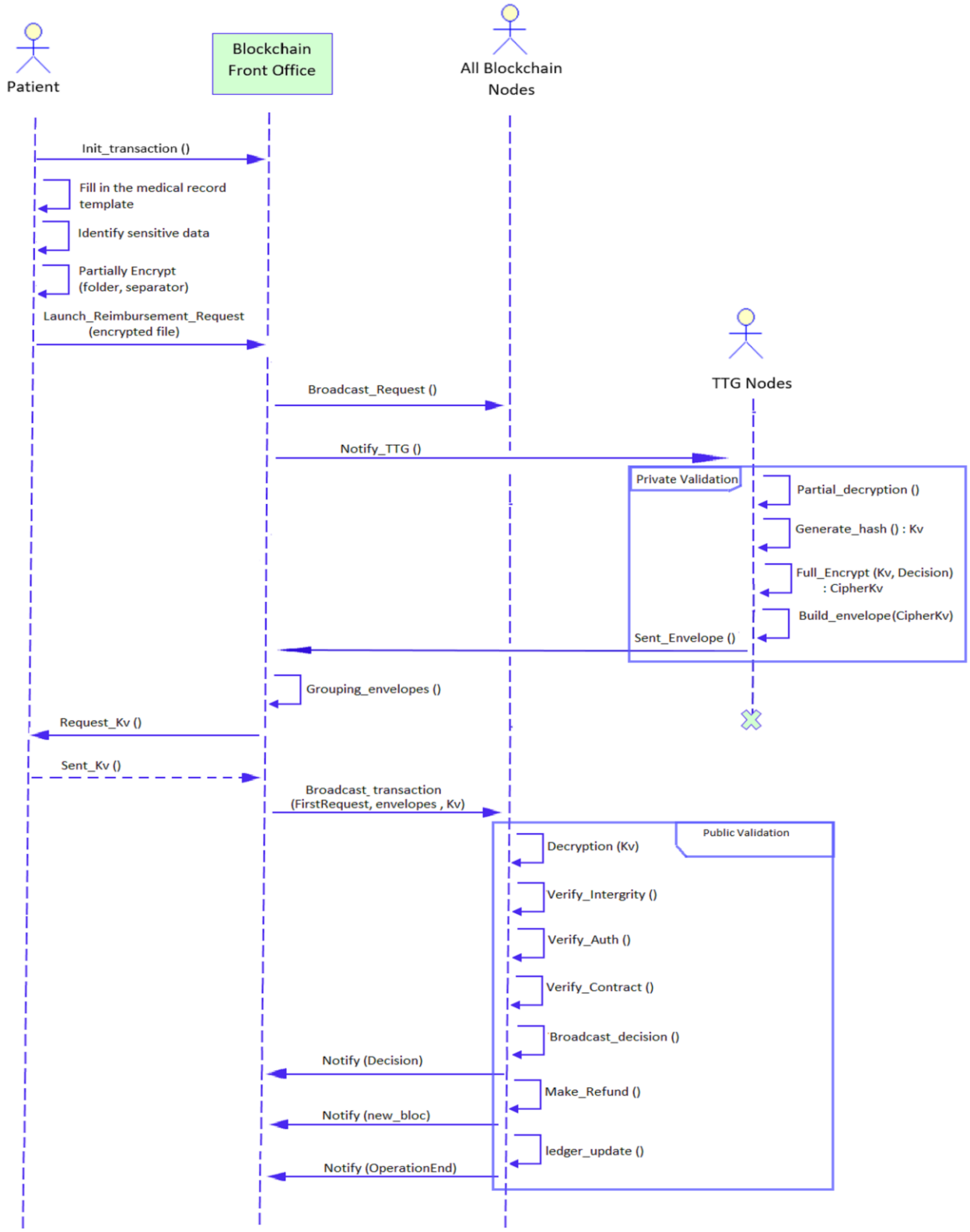


Figure 64 : le diagramme de séquence pour le remboursement des soins

VI- Discussion et Comparaison

5.1 Discussion

Comme la technologie Blockchain est fondée sur un réseau distribué, la notion d'une autorité centrale n'existe plus. De ce fait, tous les membres de ce réseau sont invités à exécuter le consensus en question afin de valider les transactions de ce réseau. Néanmoins, avec l'évolution immense de la taille des Blockchain actuelles, l'intervention collective devienne plus couteuse que ça soit en termes de quantité de calculs ou de frais de transaction. Ainsi, il provoque de vrais problèmes de latence et de performance sur le réseau[108]. En outre, l'aspect de la sécurité est aussi impacté et devient de plus en plus difficile à assurer. Notre protocole répond à cette problématique et offre des avantages permettant de garder au maximum les potentialités de la blockchain tout en favorisant son évolutivité et sa sécurité. Ces avantages peuvent être discutés dans les points suivants :

- Confidentialité

La confidentialité est assurée en se basant sur le processus de la confidentialité partielle. Ce processus permet de chiffrer les données sensibles non nécessaires pour la validation publique de la transaction et les réintégrer par la suite dans le socle global de la transaction contenant d'autres clauses publiques. De cette façon, les clauses publiques ainsi que le socle de ce type de transaction seront validé en publique, En revanche le groupe de confiance de la transaction en question garantissent la validation des éléments privées dans leur ensemble en effectuant le déchiffrement partiel et vérifiant ainsi le contenu global de la transaction en cours.

- Intégrité

De plus de la confidentialité des données sensibles, Le protocole PPCT permet aussi d'assurer l'intégrité de ces données en se basant le hachage[35] de la partie en claire. Ce dernier est désigné dans notre protocole par la clé de validation Kv. Cette clé est la même pour tous les participant du groupe TTG. Elle est utilisée par chaque membre de ce groupe afin de chiffrer l'ensemble des éléments de son enveloppe et le diffuser au sein du réseau distribué. Une fois tous les enveloppes sont récupérés, les nœuds du réseau publique demande à l'initiateur de de dévoiler la clé de validation Kv en la diffusant dans le réseau, puis ils procèdent à la vérification de l'intégrité de la transaction validée. Ils réalisent le déchiffrement de différentes enveloppes du groupe TTG. Si le déchiffrement de chaque enveloppe passe bien cela signifie que le haché du clair était le bon pour ce participant et sa validation peut être considérée dans la validation publique.

- Authentification

Assurer l'authentification fait partie aussi des avantages de cette solution. En effet, ajouter la signature du validateur du groupe TTG permet de vérifier son identité[36] et de

s'assurer que la décision envoyée dans le réseau concerne bien la bonne personne désignée dans le groupe TTG.

- Traitements hors chaîne sur un réseau réduit, privé et dynamique (cyclique)

Le PPCT permet de déplacer une partie du travail hors le réseau public. Il permet de créer une propriété vérifiable dans la tâche de calcul externalisée[138]. Cette propriété est assurée par la Clé de validation Kv qui se réalise en fonction des données sensibles de chaque transaction. La Clé Kv est un moyen efficace permettant au publique de confirmer leur validation sans avoir recours aux données sensibles de cette transaction. Le réseau délègue cette tache au TTG tout en gardant le contrôle sur le travail de ce petit réseau privé. Cette solution est d'une très grande utilité dans la mesure où elle applique le principe du calcul hors chaîne [107] afin de pouvoir résoudre efficacement ce problème. Ainsi, il :

- ✓ Ne dépend pas de la taille de la blockchain
- ✓ Favorise la scalabilité de la blockchain
- ✓ Réduit les frais des calculs
- ✓ Réduit le temps de validation globale de la transaction

- Traitement Indépendant de la confiance du réseau publique

Le PPCT ne gère pas l'accès à la Blockchain d'une manière stricte et permanente, ce n'est qu'au moment de la création de la transaction que l'initiateur de la transaction désigne son groupe de confiance. Ceci augmente la sécurité du protocole pour deux raisons, le groupe TTG représente des entités concrètement de confiance dans la réalité et non définit d'une manière procédurale. La deuxième liée au dynamisme de ce groupe donc même si l'un des nœuds malveillants arrive à intégrer le groupe TTG de la transaction courante, il ne pourra pas l'être forcément l'autre fois. Et donc l'attaque ne peut concerner qu'une seule transaction et pour laquelle on peut bien l'identifier dans l'étape de la validation publique.

- ✓ Les autorisations ne sont pas figée et change d'une transaction à une autre en favorisant le choix de l'initiateur de la transaction
- ✓ L'initiateur de la transaction désigne son groupe de confiance

- Capacité de traitement efficace et sécurisée :

Pour le protocole PPCT, le temps de calcul et de validation passe plus rapidement car il se réalise après le déchiffrement des informations sensibles par le groupe de confiance. Grâce au déchiffrement, les calculs de validation ne représentent aucune complexité et ne demande aucune exigence en termes de puissance de calcul.

- Indépendance de type et de la taille de donnée.

Numérique ou littérale, fichier Log ou donnée brute, le PPCT peut être appliqué en toute aisance sur les différents type de transaction contrairement aux autres protocoles de confidentialité qui sont restreints aux données numériques pour réaliser de calculs ou à quelques données privées pour représenter l'identité [REF]. Dans notre cas, Il suffit de choisir les bons algorithmes de Chiffrement Symétriques qui performent plus sur la volumétrie et les types de données à échanger pour améliorer les performances du protocole.

- Flexibilité des outils cryptographique

Les outils cryptographiques peuvent être appropriés en fonction des contraintes des entités qui entrent en jeu dans cette Blockchain

5.2Comparaison

L'étude comparative de ce travail consiste à analyser les différents concepts de confidentialité suite aux dimensions suivantes :

5.2.1 Dépendance de confiance

La confiance en ces approches dépend de deux facteurs principaux, ou bien en se basant sur la fiabilité de la méthode utilisée dans ces approches, ou bien sur la fiabilité de la conception matérielle de l'infrastructure mise en place. On distingue alors entre deux directions de développement : le calcul sécurisé et le calcul de confiance.

- Le **calcul sécurisé** se concentre principalement sur la sécurité de la cryptographie apportée par des problèmes difficiles qui sont minutieusement conçus à l'aide de mathématiques. La sécurité est purement dérivée de la théorie des nombres, ce qui signifie que même avec une puissance de calcul inabordable, un adversaire ne peut pas casser le chiffrement.
- Le **calcul de confiance** part d'un point différent, qui essaie de forcer l'implication de l'informatique à se comporter de manière cohérente de la manière attendue. L'application du comportement est obtenue par l'authentification matérielle, le chiffrement de la mémoire, la conception d'instructions spécifiques, etc. Ces technologies sont toujours liées à la plate-forme matérielle telle que le stockage en mémoire ou les éléments de traitement.

De toute évidence, le chiffrement homomorphe, l'ABE, le MPC ainsi que notre protocole PPCT appartiennent tous à la catégorie qui sa confiance de la sécurité des méthodes déployées dans sa couche logicielle. En effet, le calcul sécurisé est plus adapté pour être utilisé dans un réseau sans autorisation, tandis que le calcul de confiance repose davantage sur des entités physiques et peut être utilisé dans un réseau avec autorisation tel que Hyperledger.

5.2.2 Évolutivité et flexibilité

Compte tenu du fondement des méthodes de calcul mentionnées, on peut observer que l'on peut évaluer un calcul sécurisé sur n'importe quelle puissance de calcul, qu'il s'agisse d'un centre de données ou d'un périphérique tel qu'une automobile et un téléphone. Compte tenu de l'indépendance sous-jacente des périphériques, MPC ou HE peuvent tirer parti des chaînes d'outils mutuelles des fabricants de matériel et ont un potentiel futur pour les accélérations matérielles. En ce qui concerne les différents scénarios de calcul, MPC peut se transformer en un modèle approprié pour faire un compromis entre la complexité du calcul et la sécurité. La nature de MPC peut également satisfaire diverses conditions d'entrée de confidentialité des données. L'HE peut également prendre en charge l'entrée multipartite en utilisant le chiffrement et le déchiffrement distribué, mais la surcharge de l'amorçage limite les fonctions sécurisées évaluées sur l'HE. Maintenant, pour le PPCT, l'indépendance est bien claire entre l'évolutivité de l'infrastructure et son fonctionnement, les validations complexes passent sur un réseau restreint qui ne peut pas dépasser une certaine limite, il reste aussi flexible et ouvert sur tout type de chiffrement adéquat aux calculs demandés.

5.2.3 Efficacité pratique

Pour la mise en œuvre S/FHE et MPC, la multiplication sur texte chiffré est l'opération la plus fréquemment appelée et la plus fondamentale. En pratique, on peut observer que MPC est plus rapide que SHE de 1 à 2 ordres de grandeur. Et il devrait y avoir un potentiel d'amélioration du débit de 10 à 100 fois pour MPC si on considère qu'à l'avenir, une

accélération matérielle dédiée des blocs de construction MPC sera développée. Par conséquent, la grandeur d'écart de performance entre HE et MPC est très remarquable. L'opération la plus coûteuse de FHE est l'amorçage qui permet une profondeur de multiplication illimitée. Cela peut entraîner un autre ralentissement du calcul FHE. Pour le protocole PPCT, le temps de calcul et de validation passe plus rapidement car il se réalise après le déchiffrement des informations sensibles par le groupe de confiance. Grâce au déchiffrement ASEC qui est très rapide dans son processus de déchiffrement, les calculs de validations ne représentent aucune complexité et ne demande aucune exigence en termes de puissance de calcul.

Bref, le tableau suivant récapitule la comparaison entre les différentes approches de confidentialité y compris notre approche PPCT :

Tableau 14 : Comparaison entre le PPCT et les différentes approches de confidentialités

Technique	**Avantage**	**Inconvénient**	**Application**
HE	Il peut réaliser des calculs préservant la confidentialité en effectuant des calculs directement sur le texte chiffré.	Seuls certains types d'opérations, telles que l'addition et la multiplication, peuvent être mises en œuvre efficacement. L'efficacité de calcul des fonctions complexes est très faible.	Etherium
MPC	Il permet à plusieurs parties d'effectuer des calculs conjointement sur leurs entrées de données privées sans violer leur confidentialité d'entrée.	Seules certaines fonctions simples peuvent être prises en charge, et les fonctions complexes sont moins efficaces.	Enigma
ABE	Il peut simultanément assurer la confidentialité des données et un contrôle d'accès précis.	La délivrance et la révocation du certificat d'attribut dans un environnement distribué doivent encore être résolues.	Pas encore

NIZK	L'utilisateur peut facilement prouver qu'il dispose d'un solde suffisant pour le transfert avec NIZK, sans révéler le solde du compte.	Moins efficace	Zcash
PPCT	Assure la confidentialité partielle tout en gardant de la transparence sur la transaction, l'intégrité et l'authentification	Temps de latence à améliorer	Pas encore

VII- Conclusion

Le travail présenté dans ce chapitre a ouvert une nouvelle piste de compromis entre la transparence et la confidentialité au sein de la blockchain. Tout d'abord, à travers l'introduction et après avoir donné un aperçu général sur la technologie blockchain, nous avons exposé la problématique étudiée au cours de ce travail ainsi que l'objectif principal de notre proposition qu'est : la confidentialité des données sensibles tout en préservant la transparence au sein de l'infrastructure Blockchain. Nous avons présenté, ensuite, les différents travaux connexes dans ce contexte qui sont généralement conçus à la base des algorithmes de calculs sécurisés à savoir : les ABE, le SMPC, le chiffrement Homomorphe et les algorithmes NIZK. Puis, nous avons donné la description détaillée de notre solution intitulée «Protocol for Preserving Confidentiality & Transparency (PPCT)», en discutant, à la fin de ce chapitre, ses avantages vis à vis aux autres systèmes équivalents. À travers ce travail, on peut conclure que le protocole PPCT proposé ici, a pu résoudre le problème de la confidentialité en exploitant les outils conventionnels de la cryptographie et les adaptant aux concepts de la Blockchain dans une nouvelle optique. Nous avons défini une nouvelle notion de la confidentialité qu'est la confidentialité partielle puis nous avons intégré les fonctions de hachage et les algorithmes de signatures au cœur du processus de la validation. L'intégration de ces derniers assure l'augmentation du niveau de sécurité du système en garantissant les deux autres propriétés piliers de la sécurité, à savoir : l'intégrité et l'authentification. Ainsi, le

protocole PPCT permet d'assurer la confidentialité des données sensibles de chaque transaction via le chiffrement partiel de celle-ci, puis l'intégrité de ces données en utilisant leur haché, puis l'authentification des premiers validateurs décideurs de cette transaction via leurs signatures. Comme présenté dans la section de la discussion, les performances de ce protocole sont bien au rendez-vous vis à vis aux autres solutions de confidentialité qui s'appuient sur des outils de calculs sécurisés. Dans notre prochain travail, nous souhaitons se focaliser plus sur le temps de latence entre la validation initiale et la validation finale de la transaction tout en étudiant les pistes d'amélioration de ce critère en ajoutant l'aspect du parallélisme dans ce processus. Nous souhaitons aussi réaliser des expériences comparatives entre les différentes combinaisons possibles en proposant un schéma d'utilisation plus claire sur chaque type de besoin.

Apports de la thèse :

La présente thèse met le focus sur un enjeu très complexe et manque de nouvelles stratégies plus récentes adaptées aux exigences technologiques actuelles. Il s'agit d'assurer la confidentialité des données sensibles et protéger la vie privée des utilisateurs. À travers les travaux de recherches réalisés dans le cadre de cette thèse, nous avons réussi à établir un rapprochement de ce volet dans les technologies dites disruptives, à savoir l'intelligence artificielle et la Blockchain.

Nous avons présenté dans la 1ère contribution une nouvelle approche intelligente pour le chiffrement des données. Cette approche utilise un système à base de connaissance afin de décider le système de chiffrement le plus adéquat aux échanges réalisés. Il se base sur une base de connaissance dont ils sont stockés un ensemble très varié de cas de chiffrements grâce à un entrainement préalable du système. La base de connaissance est ensuite exploitée via un moteur de référence utilisant l'algorithme de recherche du Skyline BNL pour retourner l'algorithme qui répond mieux aux contraintes exigées par la communication en cours. Les résultats de ce travail étaient très concluants par rapport aux décisions de chiffrements proposées.

À travers les différentes comparaisons, les résultats des expériences réalisés dans le cadre de cette contribution ont confirmé le bon choix des algorithmes pour chaque source de données. De plus, nous avons pu tirer un autre avantage très important, il s'agit de prédire une estimation préalable des valeurs des indicateurs pour chaque algorithme, étant donné que les résultats trouvés sont très proches des résultats réels d'exécution.

Au cours de la deuxième contribution de cette thèse, nous avons proposé une nouvelle variante du premier système proposé. Une nouvelle version dans laquelle nous avons introduit une couche supplémentaire permettant de brancher le moteur d'inférence avec l'algorithme Skyline le plus adéquat aux contraintes réellement exploitées lors de chaque communication. En effet, à travers ce travail, nous avons étudié les différents algorithmes Skyline dans notre contexte pour en sortir avec une nouvelle brique de paramétrage

permettant de sélectionner les meilleurs algorithmes de chiffrement sans impacter les performances du système.

À travers ce travail, et après plusieurs scénarios réalisés dans ce contexte, nous avons conclu que ce système doit absolument rendre le choix de l'algorithme Skyline dynamique en fonction de nombre des dimensions. Nous avons pu conclure que la performance des algorithmes Skyline sur notre problème de chiffrement dépend de deux paramètres, à savoir : le nombre des dimensions et le choix des dimensions ayant un écart très fin entre ses valeurs. Et donc, on a intégré les 3 catégories dans la brique de paramétrage dédiée aux algorithmes Skyline comme suite :

- Catégorie 1 : BNL pour un petit nombre de dimensions (moins de 5)
- Catégorie 2 : DC ou NN pour un nombre entre 5 et 10 dimensions, peu importe les valeurs
- Catégorie 3 : Bitmap au-delà de 5 dimensions avec restriction sur le nombre de valeurs.

Finalement, dans la dernière contribution, nous avons pu concevoir un système assurant la confidentialité des données sensibles tout en préservant la propriété de la Transparence au sein de la Blockchain. A travers ce travail, nous avons trouvé un compromis entre les deux exigences Confidentialité & Transparence au sein d'un même protocole de validation qu'on a conventionné de le nommer PPCT pour dire « Protocol for Preserving Confidentiality & Transparency », ce nouveau protocole consiste à appliquer une confidentialité partielle sur les transactions échangées tout en assurant le processus de leurs validations par les différents nœuds de la Blockchain. En effet, le PPCT repose sur deux étapes de validation. La première validation s'effectue sur un réseau restreint de nœuds qu'on nomme Groupe de Confiance de la Transaction (Transaction Trust Group TTG) qui décide et signe leurs retours via une clé de validation générée à la base des données confidentielle. La deuxième validation s'effectue d'une manière publique en divulguant la clé de validation sur l'ensemble des nœuds pour confirmer que les premières décisions de validation effectuées par le TTG sont bien correctes et reposent bien sur les bonnes informations sans pour autant accéder aux informations sensibles sujettes de cette confidentialité. Cette première version de ce protocole s'appuie sur différents concepts et outils cryptographiques ce qui augmente sa résistance et son efficacité pour répondre largement aux objectifs définis et tracés tout au long de cette thèse qui est la confidentialité et le respect de la vie privée pour les technologies disruptives.

Perspectives :

Au cours de la réalisation de ces travaux, différentes pistes de recherche nous ont inspiré mais que nous avons dû les temporiser afin de pouvoir atteindre nos premiers objectifs tracés pour cette thèse. Des perspectives tantôt pour améliorer les travaux déjà misent en place et tantôt pour affranchir d'autres techniques dont on prévoit des bons résultats.

Sur les deux axes de recherche que nous avons abordés, nous prévoyons explorer les idées suivantes :

Sur l'axe de l'Intelligence Artificielle :

- ✓ Mettre en place une Nouvelle variante de l'algorithme Bitmap pour le système de chiffrement Intelligent, une variante qui prend en compte les valeurs des dimensions comme l'entropie, l'avalanche et bien d'autres.
- ✓ Appliquer les méthodes des Machine Learning dans la partie de l'acquisition et la classification des données en entrée au sein de notre stratégie de chiffrement Intelligent.
- ✓ Dans le même principe notre stratégie de chiffrement Intelligent, utiliser une autre technique de l'IA au lieu des Systèmes à Base de Connaissances.

Sur l'axe de la technologie Blockchain :

- ✓ Effectuer une étude expérimentale d'une Blockchain basée sur le protocole PPCT en le comparant avec d'autres types de Blockchain.
- ✓ Intégrer l'une des techniques de l'IA dans le protocole PPCT au niveau de la confidentialité partielle.
- ✓ Mettre en place un nouveau consensus intelligent pour la Blockchain.

Liste des Publications

1) S. Trichni, F. Omary, and M. Bougrine, "New Blockchain Protocol for Partial Confidentiality and Transparency (PPCT)," International Journal of Advanced Computer Science and Applications, vol. 13, no. 2, pp. 617–617, 2022, doi: 10.14569/IJACSA.2022.0130273.

2) S. Trichni, F. Omary, and M. Bougrine, "New Smart Encryption Approach based on Multidimensional Analysis Tools," International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 666–675, 2021, doi: 10.14569/IJACSA.2021.0120579.

3) S. Trichni, F. Omary, A. Idrissi, M. Bougrine, and M. Abourezq, "New intelligent strategy for encryption decisional support system," International Journal of High Performance Systems Architecture, vol. 9, no. 4, pp. 173–181, 2020, doi: 10.1504/IJHPSA.2020.113678.

4) M. Bougrine, F. Omary, and S. Trichni, "Security of a new hybrid ciphering system," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, pp. 694–699, 2020.

5) M. Bougrine, S. Trichni, and F. Omary, "Improving performance of the symmetrical evolutionary ciphering system SEC," International Journal of High Performance Systems Architecture, vol. 10, no. 1, pp. 12–19, 2021, doi: 10.1504/IJHPSA.2021.115502.

6) S. Trichni, F. Omary, B. Boulahiat, and M. Bougrine, "A new approach of mutation operator applied to the ciphering system SEC," 2011, pp. 680–685.

7) M. Bougrine, F. Omaiy, S. Trichni, and B. Boulahiat, "New evolutionary tools for a new ciphering system SEC version," 2012, pp. 140–146. doi: 10.1109/CCST.2012.6393549.

[1] A. Kerckhoffs, “LA CRYPTOGRAPHIE MILITAIRE.pdf.” Journal des Sciences Militaires, Jan. 1983.

[2] J. M. Alfred, C. V. O. Paul, and A. V. Scott, “Handbook of applied Cryptography,” *CRC press*, p. 815, 1997.

[3] B. Aurélien, “La théorie de l’information,” *ISBN : 9782070138098 - 243527*.

[4] P. Thomas, “Implantation et optimisation des primitives cryptographiques.” École Normale Supérieure,Université Paris 7, Oct. 25, 2001.

[5] L. HUGO, “1965-2020 : La loi de Moore est morte - Science & Vie.” https://www.science-et-vie.com/science-et-culture/1965-2020-la-loi-de-moore-est-morte-53613

[6] H. Zbinden, “QKD: from the concept to a commercial application,” p. 49.

[7] R. T. Thew, N. Gisin, and G. Ribordy, “Quantum Technologies for IT security and privacy protection online,” p. 3.

[8] B. Singhal, G. Dhameja, and P. S. Panda, *Beginning Blockchain*. Berkeley, CA: Apress, 2018. doi: 10.1007/978-1-4842-3444-0.

[9] M. Lefevre, “CM8 : Système à Base de Connaissances,” p. 61, 2020.

[10] B. Martin, *Codage, cryptologie et applications*. PPUR presses polytechniques, 2004.

[11] C. Bidan, “Cryptographie et Cryptanalyse,” p. 81.

[12] B. Gérard, “Cryptanalyses statistiques des algorithmes de chiffrement à clef secrète.,” p. 225.

[13] D.-J. Mercier, “Cryptographie classique et cryptographie publique à clé révélée,” p. 17.

[14] “Histoire du chiffrement et de ses méthodes.”. [Online]. Available: https://www.thawte.fr/assets/documents/guides/history-cryptography.pdf

[15] ANSSI-PG-083. “GUIDE DES MÉCANISMES CRYPTOGRAPHIQUES.” 01/01/2020, [Online]. Available:https://www.ssi.gouv.fr/uploads/2021/03/anssi-guide-mecanismes_crypto-2.04.pdf

[16] “Introduction à la cryptographie et au chiffrement par substitution mono-alphabétique.” [Online]. Available: https://repository.root-me.org/Cryptographie/Sym%C3%A9trique/FR%20-%20Introduction%20%C3%A0%20la%20cryptographie%20et%20au%20chiffrement%20par%20substitution%20mono-alphab%C3%A9tique.pdf

[17] S. Even, “A construction of a cipher from a single pseudorandom permutation,” p. 11.

[18] European Network and Information Security Agency ., *Algorithms, key size and parameters: report – 2014.* LU: Publications Office, 2013. [Online]. Available: https://data.europa.eu/doi/10.2824/36822

[19] C. Berrou and V. Gripon, *Petite mathématique du cerveau: Une théorie de l'information mentale*. Odile Jacob, 2012.

[20] H. Lehning, "Strasbourg, témoin de l'évolution de la cryptologie du 16e au 17e siècle," *rbnu*, no. 13, pp. 46–57, May 2016, doi: 10.4000/rbnu.1501.

[21] Univ. Lille 1, "Principes et Algorithmes de Cryptographie- Chiffrement Poly Alphabétique" [Online]. Available: https://www.fil.univ-lille.fr/~wegrzyno/portail/PAC/Doc/Cours/ChiffrementPolyAlphabetique/chiffrementPolyAlphabetique.pdf

[22] R. Masram, V. Shahare, J. Abraham, and R. Moona, "Analysis and Comparison of Symmetric Key Cryptographic Algorithms Based on Various File Features," *IJNSA*, vol. 6, no. 4, pp. 43–52, Jul. 2014, doi: 10.5121/ijnsa.2014.6404.

[23] "ECRYPT_2012_Recomandation.pdf."

[24] H. Lehning, "Les révolutions cryptographiques autour du XXe siècle," p. 7.

[25] D. Lamas, "La cryptographie," Haute école de gestion de Genève, 2015. [Online]. Available: https://doc.rero.ch/record/258630

[26] S. Bruce, "Essays: Why Cryptography Is Harder Than It Looks - Schneier on Security," *Information Security Bulletin 1997*. https://www.schneier.com/essays/archives/1997/01/why_cryptography_is.html

[27] M. Coles and R. Landrum, "Introduction to Encryption," in Expert SQL Server 2008 Encryption, M. Coles and R. Landrum, Eds. Berkeley, CA: Apress, 2009, pp. 1–20. doi: 10.1007/978-1-4302-3365-7_1.

[28] Ö. Ismet and S. Ibrahim, "Analysis and Comparison of Image Encryption Algorithms.pdf." International Journal of Information Technology, Volume 1 Number 2.

[29] A. K. Lenstra and E. R. Verheul, "Selecting Cryptographic Key Sizes," *J. Cryptology*, vol. 14, no. 4, pp. 255–293, Sep. 2001, doi: 10.1007/s00145-001-0009-4.

[30] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 26, no. 1, pp. 96–99, Jan. 1983, doi: 10.1145/357980.358017.

[31] G. J. Simmons, "Symmetric and Asymmetric Encryption," in *Secure Communications and Asymmetric Cryptosystems*, Routledge, 1982.

[32] C. Hanin, "Nouvelles Approches pour la Sécurité Informatique basées sur les Automates Cellulaires.". Thesis call Number 3055, Mohammed V University of Rabat, p. 120, Décembre 2017

[33] D. Liu *et al.*, Eds., "Chapter 3 - An Introduction To Cryptography," in *Next Generation SSH2 Implementation*, Burlington: Syngress, 2009, pp. 41–64. doi: 10.1016/B978-1-59749-283-6.00003-9.

[34] S. R. Ellis, "Chapter 63 - Fundamentals of Cryptography," in *Computer and Information Security Handbook (Second Edition)*, J. R. Vacca, Ed. Boston: Morgan Kaufmann, 2013, pp. 1031–1038. doi: 10.1016/B978-0-12-394397-2.00063-5.

[35] N. Judy, "Understanding Hashing in Cryptography," *Engineering Education (EngEd) Program | Section*. https://www.section.io/engineering-education/understand-hashing-in-cryptography/

[36] IEvangelist, "Cryptographic Signatures." https://docs.microsoft.com/en-us/dotnet/standard/security/cryptographic-signatures

[37] N. Svetlin, "Practical Cryptography for Developers - Digital Signatures." https://cryptobook.nakov.com/

[38] L. Ben, "What is a Digital Signature?," *SearchSecurity*. https://www.techtarget.com/searchsecurity/definition/digital-signature

[39] F. Fürst, "Histoire de l'Intelligence Artificielle," p. 13.

[40] S. J. Russell, P. Norvig, "Artificial Intelligence - A Modern Approach." ISBN 0-13-103805-2, Library of Congress Cataloging-in-Publication Data Q335.R86 1995

[41] G. Frege and C. Besson, *Idéographie*. Vrin, 1999.

[42] D. A. Réda, "Intelligence Artificielle," p. 26. https://www.researchgate.net/publication/309238376

[43] P. De Loor, " INTELLIGENCE ARTIFICIELLE ET SIMULATION" Cours– ENIB – LabSTICC- CNRS 2017-2018

[44] Y. LeCun, " Qu'est-ce que l'intelligence artificielle ? ", UPL4485925235409209505.

[45] L. Derome, " Le cerveau selon Aristote ", Université de Montréal Département de philosophie Faculté des arts et des sciences, [Online]. Available: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18784/Derome_L%C3%A9a_2016_memoire.pdf?sequence=6&isAllowed=y

[46] "Neuropsychologie." https://courspsycho.blog4ever.com/neuropsychologie

[47] A. Tixier-Vidal, "De la théorie cellulaire à la théorie neuronale," *Biologie Aujourd'hui*, vol. 204, no. 4, pp. 253–266, 2010, doi: 10.1051/jbio/2010015.

[48] G. Bugnicourt, "Adhésion, croissance et polarisation de neurones sur substrats micro-et nano-structurés," p. 215.

[49] "The Man Who Tried to Redeem the World with Logic," *Nautilus | Science Connected*, Jan. 29, 2015. https://nautil.us/the-man-who-tried-to-redeem-the-world-with-logic-2885/

[50] O. Ezratty, “Les usages de l’intelligence artificielle,” p. 522.

[51] Inria, “ Intelligence Artificielle LIVRE BLANC N°01 - Les défis actuels et l’action d’Inria”

[52] L. Gugerty, “Newell and Simon’s Logic Theorist: Historical Background and Impact on Cognitive Modeling,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 880–884, Oct. 2006, doi: 10.1177/154193120605000904.

[53] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.

[54] H. L. Dreyfus, S. E. Drey-fus, and L. A. Zadeh, “Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer,” *IEEE Expert*, vol. 2, no. 2, pp. 110–111, Jun. 1987, doi: 10.1109/MEX.1987.4307079.

[55] M. C. F. LAOUBI,“ Contrôle d’un pendule inversé par un réseau de neurones artificiels.”, Université Akli Mohand Oulhadj, p. 15, 2018.

[56] C2RP Carif-Oref Hauts-de-France, “C2dossier : L'Intelligence Artificielle.”, 2021

[57] “Introduction aux réseaux neuronaux,” *De Boeck Supérieur*. https://www.deboecksuperieur.com/ouvrage/9782804137960-introduction-aux-reseaux-neuronaux

[58] B. Victorri, “Chapitre 7. Le connexionnisme,” in *Traité de neuropsychologie clinique*, Louvain-la-Neuve: De Boeck Supérieur, 2008, pp. 53–64. doi:10.3917/dbu.eusta.2008.01.0053.

[59] C. Jouve, “Représentation des connaissances pour les problèmes de conception. Application à un système à base de connaissances pour la conception de réseaux informatiques: NEST.,” p. 249.

[60] H. Thierry, “ Modélisation et Représentation des Connaissances -Introduction”. Institut Galilée - Université Paris 13, p. 122, Janvier 2019.

[61] J.-L. Ermine, “Les systèmes de connaissances,”. Hermes Science Publication, pp.144, 2000. hal-00856172.

[62] T. M. Mitchell, “*Machine learning*, Nachdr,” Nachdr New York: McGraw-Hill, ISBN: 0070428077, p. 432. [Online]. Available: https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf

[63] G. Bonaccorso, *Machine Learning Algorithms*, Packt Publishing Ltd, ISBN: 978-1-78588-451-1, p. 352, Juillet 2017

[64] Md. M. Ahamad and Md. I. Abdullah, “Comparison of Encryption Algorithms for Multimedia,” *Rajshahi Univ. j. sci. eng.*, vol. 44, pp. 131–139, Nov. 2016, doi: 10.3329/rujse.v44i0.30398.

[65] F. Omary, A. Mouloudi, A. Tragha, and A. Bellaachia, “A new ciphering method associated with evolutionary algorithm,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3984 LNCS, pp. 346–354, 2006, doi: 10.1007/11751649_38.

[66] S. Trichni, F. Omary, A. Idrissi, M. Bougrine, and M. Abourezq, “New intelligent strategy for encryption decisional support system,” *International Journal of High Performance Systems Architecture*, vol. 9, no. 4, pp. 173–181, 2020, doi: 10.1504/IJHPSA.2020.113678.

[67] M. Bougrine, F. Omary, and S. Trichni, “Security of a new hybrid ciphering system,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 694–699, 2020.

[68] M. N. A. Wahid, A. Ali, B. Esparham, and M. Marwan, “A Comparison of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish for Guessing Attacks Prevention,” p. 7, 2018.

[69] A. Devi, A. Sharma, and A. Rangra, “Performance analysis of Symmetric Key Algorithms: DES, AES and Blowfish for Image encryption and decryption,” vol. 4, no. 6, p. 6, 2015.

[70] M. Oulehla and D. Malanik, “Comparison of cryptographic methods Triple DES, AES and a method based on the arithmetic of elliptic curves (ECC) on the Android mobile platform. - extended version,” *International Journal of Computers and Communications*, vol. 9, p. 62, Jan. 2015.

[71] S. Douzi, Thesis:“Vers un Système Deep Learning de Detection des Intrusions.”, Mohammed V University of Rabat, p. 109, Juillet 2019.

[72] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, and K. Li, “AI^2: Training a Big Data Machine to Defend,” in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, New York, NY, USA, Apr. 2016, pp. 49–54. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.79.

[73] P. Vähäkainu and M. Lehto, “Artificial intelligence in the cyber security environment Artificial intelligence in the cyber security environment,” Dec. 2019.

[74] M. Landauer, F. Skopik, M. Wurzenberger, and A. Rauber, “System Log Clustering Approaches for Cyber Security Applications: A Survey,” *Computers & Security*, vol. 92, p. 101739, May 2020, doi: 10.1016/j.cose.2020.101739.

[75] M. Abadi and D. G. Andersen, “Learning to Protect Communications with Adversarial Neural Cryptography,” *arXiv:1610.06918 [cs]*, Oct. 2016. [Online]. Available: http://arxiv.org/abs/1610.06918

[76] S. Trichni, F. Omary, B. Boulahiat, and M. Bougrine, “A new approach of mutation operator applied to the ciphering system SEC,” 2011, pp. 680–685.

[77] M. Bougrine, S. Trichni, and F. Omary, “Improving performance of the symmetrical evolutionary ciphering system SEC,” *International Journal of High Performance Systems Architecture*, vol. 10, no. 1, pp. 12–19, 2021, doi: 10.1504/IJHPSA.2021.115502.

[78] M. ABOUREZQ, Thesis: “Cloud Service Selection using the Skyline and Multi Criteria Decision Aiding.”, Mohammed V University of Rabat, p. 204, 21 Octobre 2017.

[79] S. Börzsönyi, D. Kossmann, and K. Stocker, “The Skyline operator,” *Proceedings 17th International Conference on Data Engineering*, 2001, doi: 10.1109/ICDE.2001.914855.

[80] P. K. Wanko, “Optimisation des requêtes skyline multidimensionnelles,” p. 155.

[81] E. Tiakas, A. Papadopoulos, and Y. Manolopoulos, “Skyline queries: An introduction,” *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2015, doi: 10.1109/IISA.2015.7388053.

[82] K.-L. Tan, P.-K. Eng, and B. Ooi, “Efficient Progressive Skyline Computation.,” Jan. 2001, pp. 301–310.

[83] S. Berchtold, C. Böhm, D. Keim, and H. Kriegel, “A cost model for nearest neighbor search in high-dimensional data space,” 1997. doi: 10.1145/263661.263671.

[84] G. Cormier, “ Systèmes Intelligents - Chapitre 2: Systèmes à base de règles,” GIND5439, Université de Moncton, p. 45.

[85] F. L. Ber, J. Lieber, and A. Napoli, “Les systèmes à base de connaissances,” p. 14.

[86] N. Sendrier, “Introduction à la théorie de l’information,” p. 70.

[87] S. Nadouri, Z. Sahnoun, and A. Hadjali, “Using G-skyline to improve Decision-Making,” p. 4, 2018.

[88] D. Sacharidis, P. Bouros, and T. Sellis, “Caching Dynamic Skyline Queries,” in *Scientific and Statistical Database Management*, vol. 5069, B. Ludäscher and N. Mamoulis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 455–472. doi: 10.1007/978-3-540-69497-7_29.

[89] D. Kossmann, F. Ramsak, and S. Rost, “Shooting Stars in the Sky: An Online Algorithm for Skyline Queries,” p. 12.

[90] T. Bouadi, M.-O. Cordier, and R. Quiniou, “Requêtes skyline hiérarchiques,” p. 18.

[91] T. Kian-Lee, E. Pin-Kwang, and O. Beng Chin, “Efficient Progressive Skyline Computation.” http://www.vldb.org/conf/2001/P301.pdf

[92] I. Bashir, *Mastering Blockchain*. Packt Publishing Ltd, 2017.

[93] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” p. 9.

[94] X. Xu, I. Weber, and M. Staples, *Architecture for Blockchain Applications*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-03035-3.

[95] V. Dhillon, D. Metcalf, and M. Hooper, *Blockchain Enabled Applications*. Berkeley, CA: Apress, 2017. doi: 10.1007/978-1-4842-3081-7.

[96] V. Morabito, *Business Innovation Through Blockchain*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-48478-5.

[97] M. Swan, *Blockchain: blueprint for a new economy*, First edition. Beijing : Sebastopol, CA: O'Reilly, 2015.

[98] J. M. Graglia and C. Mellon, "Blockchain and Property in 2018: At the End of the Beginning," *Innovations: Technology, Governance, Globalization*, vol. 12, no. 1–2, pp. 90–116, Jul. 2018, doi: 10.1162/inov_a_00270.

[99] W. K. Chan, J.-J. Chin, and V. T. Goh, "Simple and scalable blockchain with privacy," *Journal of Information Security and Applications*, vol. 58, p. 102700, May 2021, doi: 10.1016/j.jisa.2020.102700.

[100] R. Zhang, R. Xue, and L. Liu, "Security and Privacy on Blockchain," *arXiv:1903.07602 [cs]*, Aug. 2019. [Online]. Available: http://arxiv.org/abs/1903.07602

[101] P. J. Taylor, T. Dargahi, A. Dehghantanha, R. M. Parizi, and K.-K. R. Choo, "A systematic literature review of blockchain cyber security," *Digital Communications and Networks*, vol. 6, no. 2, pp. 147–156, May 2020, doi: 10.1016/j.dcan.2019.01.005.

[102] Y. I. Alzoubi, A. Al-Ahmad, and H. Kahtan, "Blockchain technology as a Fog computing security and privacy solution: An overview," *Computer Communications*, vol. 182, pp. 129–152, Jan. 2022, doi: 10.1016/j.comcom.2021.11.005.

[103] T. Nóbrega, C. E. S. Pires, and D. C. Nascimento, "Explanation and answers to critiques on: Blockchain-based Privacy-Preserving Record Linkage," *Information Systems*, p. 101935, Oct. 2021, doi: 10.1016/j.is.2021.101935.

[104] N. Pathak and A. Bhandari, *IoT, AI, and Blockchain for .NET: Building a Next-Generation Application from the Ground Up*. Berkeley, CA: Apress, 2018. doi: 10.1007/978-1-4842-3709-0.

[105] A. Wang, J. Shen, C. Wang, H. Yang, and D. Liu, "Anonymous data collection scheme for cloud-aided mobile edge networks," *Digital Communications and Networks*, vol. 6, no. 2, pp. 223–228, May 2020, doi: 10.1016/j.dcan.2019.04.001.

[106] J. Toledano and L. Janin, "Les enjeux des blockchains," p. 150.

[107] D. Zhang, A. Su, F. Xu, and J. Chen, "ARPA Whitepaper," *arXiv:1812.05820 [cs]*, Dec. 2018. [Online]. Available: http://arxiv.org/abs/1812.05820

[108] G. R. Carrara, L. M. Burle, D. S. V. Medeiros, C. V. N. de Albuquerque, and D. M. F. Mattos, "Consistency, availability, and partition tolerance in blockchain: a survey on the consensus mechanism over peer-to-peer networking," *Ann. Telecommun.*, vol. 75, no. 3, pp. 163–174, Apr. 2020, doi: 10.1007/s12243-020-00751-w.

[109] Y. Chen, H. Chen, Y. Zhang, M. Han, M. Siddula, and Z. Cai, "A survey on blockchain systems: Attacks, defenses, and privacy preservation," *High-Confidence Computing*, vol. 2, no. 2, p. 100048, Jun. 2022, doi: 10.1016/j.hcc.2021.100048.

[110] P. Christen, R. Schnell, T. Ranbaduge, and A. Vidanage, "A critique and attack on 'Blockchain-based privacy-preserving record linkage,'" *Information Systems*, p. 101930, Oct. 2021, doi: 10.1016/j.is.2021.101930.

[111] Y. Wang and A. Kogan, "Designing confidentiality-preserving Blockchain-based transaction processing systems," *International Journal of Accounting Information Systems*, vol. 30, pp. 1–18, Sep. 2018, doi: 10.1016/j.accinf.2018.06.001.

[112] N. Sfetcu, *La philosophie de la technologie blockchain - Ontologies*. Nicolae Sfetcu.

[113] Fondation pour l'Innovation Politique, "La blockchain ou la confiance distribuée", Juin 2016. [Online]. Available: https://www.ledecodeur.ch/wp-content/uploads/2016/09/la-blockchain-ou-la-confiance-distribu%C3%A9e-source-fondapol.pdf

[114] P. Genestier, L. Letondeur, S. Zouarhi, A. Prola, and J.-M. Temerson, "Blockchains and smart contracts: Prospects for the Internet of things and e-health," *Annales des Mines - Realites industrielles*, vol. 2017, no. 3, pp. 70–73, Jul. 2017.

[115] Q. Miao, H. Lin, J. Hu, and X. Wang, "An intelligent and privacy-enhanced data sharing strategy for blockchain-empowered Internet of Things," *Digital Communications and Networks*, p. S2352864821001048, Jan. 2022, doi: 10.1016/j.dcan.2021.12.007.

[116] "The Size and Quality of a Data Set | Data Preparation and Feature Engineering for Machine Learning," *Google Developers*. https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality

[117] G. Miklau and D. Suciu, "A formal analysis of information disclosure in data exchange," *Journal of Computer and System Sciences*, vol. 73, no. 3, pp. 507–534, May 2007, doi: 10.1016/j.jcss.2006.10.004.

[118] T. Ko, J. Lee, and D. Ryu, "Blockchain Technology and Manufacturing Industry: Real-Time Transparency and Cost Savings," *Sustainability*, vol. 10, no. 11, Art. no. 11, Nov. 2018, doi: 10.3390/su10114274.

[119] R. Gennaro and D. Wichs, "Fully Homomorphic Message Authenticators," in *Advances in Cryptology - ASIACRYPT 2013*, Berlin, Heidelberg, 2013, pp. 301–320. doi: 10.1007/978-3-642-42045-0_16.

[120] C. Regueiro, I. Seco, S. de Diego, O. Lage, and L. Etxebarria, "Privacy-enhancing distributed protocol for data aggregation based on blockchain and homomorphic encryption," *Information Processing & Management*, vol. 58, no. 6, p. 102745, Nov. 2021, doi: 10.1016/j.ipm.2021.102745.

[121] Z. Qiao, S. Liang, S. Davis, and H. Jiang, "Survey of attribute based encryption," in *15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking*

and Parallel/Distributed Computing (SNPD), Jun. 2014, pp. 1–6. doi: 10.1109/SNPD.2014.6888687.

[122] A. Lewko and B. Waters, "Decentralizing Attribute-Based Encryption," in *Advances in Cryptology – EUROCRYPT 2011*, Berlin, Heidelberg, 2011, pp. 568–588. doi: 10.1007/978-3-642-20465-4_31.

[123] B. Waters, "Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization," in *Public Key Cryptography – PKC 2011*, Berlin, Heidelberg, 2011, pp. 53–70. doi: 10.1007/978-3-642-19379-8_4.

[124] N. OUALHA and C. JANNETEAU, "MÉTHODE DE CHIFFREMENT BASÉE SUR LES ATTRIBUTS COMPRENANT UNE PHASE DE PRÉ-CALCUL," EP 3 371 929 B1. Sept 2018

[125] S. Wang, W. Ding, J. Li, Y. Yuan, L. Ouyang, and F.-Y. Wang, "Decentralized Autonomous Organizations: Concept, Model, and Applications," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 5, pp. 870–878, Oct. 2019, doi: 10.1109/TCSS.2019.2938190.

[126] "B-Box - A Decentralized Storage System Using IPFS, Attributed-based Encryption, and Blockchain." https://ieeexplore.ieee.org/abstract/document/9140747/

[127] J. Ikherbane, M. Maouche, and I. Lyon, "Calcul multipartite sécurisé basé sur un environnement d'exécution sécurisé," p. 8, 2018.

[128] A. C.-C. Yao, "How to generate and exchange secrets," in *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, Oct. 1986, pp. 162–167. doi: 10.1109/SFCS.1986.25.

[129] O. Goldreich, "Secure Multi-Party Computation." 1998.

[130] M. Andrychowicz, S. Dziembowski, D. Malinowski, and L. Mazurek, "Secure Multiparty Computations on Bitcoin," in *2014 IEEE Symposium on Security and Privacy*, May 2014, pp. 443–458. doi: 10.1109/SP.2014.35.

[131] G. Zyskind, O. Nathan, and A. Pentland, "Enigma: Decentralized Computation Platform with Guaranteed Privacy," *arXiv:1506.03471 [cs]*, Jun. 2015. [Online]. Available: http://arxiv.org/abs/1506.03471

[132] S. Goldwasser, S. Micali, and C. Rackoff, "The Knowledge Complexity of Interactive Proof Systems," *SIAM Journal on Computing*, Jul. 2006, doi: 10.1137/0218012.

[133] M. Blum, P. Feldman, and S. Micali, "Non-interactive zero-knowledge and its applications," in *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*, New York, NY, USA: Association for Computing Machinery, 2019, pp. 329–349. [Online]. Available: https://doi.org/10.1145/3335741.3335757

[134] N. Bitansky, A. Chiesa, Y. Ishai, O. Paneth, and R. Ostrovsky, "Succinct Non-interactive Arguments via Linear Interactive Proofs," in *Theory of Cryptography*, Berlin, Heidelberg, 2013, pp. 315–333. doi: 10.1007/978-3-642-36594-2_18.

[135] Q. Li and Z. Xue, “A Privacy-Protecting Authorization System Based on Blockchain and zk-SNARK,” in *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies*, New York, NY, USA, Dec. 2020, pp. 439–444. doi: 10.1145/3444370.3444610.

[136] “Zerocash: Decentralized Anonymous Payments from Bitcoin | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/abstract/document/6956581

[137] L. Peng, W. Feng, Z. Yan, Y. Li, X. Zhou, and S. Shimizu, “Privacy preservation in permissionless blockchain: A survey,” *Digital Communications and Networks*, vol. 7, no. 3, pp. 295–307, Aug. 2021, doi: 10.1016/j.dcan.2020.05.008.

[138] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, “A survey on the security of blockchain systems,” *Future Generation Computer Systems*, vol. 107, pp. 841–853, Jun. 2020, doi: 10.1016/j.future.2017.08.020.

I- Chiffrement évolutionniste

Le chiffrement évolutionniste est un type de chiffrement qui repose sur les algorithmes évolutionnistes pour chiffrer les données sensibles. Inventé par Dr. F.Omary en 2006, SEC (Symetrical Evolutionary Ciphering) est le premier système de chiffrement qui propose ce type d'approche. Il s'agit d'un chiffrement symétrique qui cherche à trouver les meilleurs permutations capables de rendre le chiffré aléatoire. Le principe est simple, tout d'abord, il réalise un encodage du texte clair en liant chaque caractère à sa liste de positions. Ensuite, à travers les différentes itérations de l'algorithme évolutionniste, il essaie de trouver la combinaison des permutations la plus puissante de ces listes pour réaliser un chiffrement bien sécurisé [6], et ce en se basant sur une fonction d'évaluation bien définie.

1.1 Description de l'algorithme de chiffrement

Pour expliquer le chiffrement, nous considérons le texte en clair T comme une entrée de ce système, puis nous appliquons la première étape de l'algorithme comme suite :

Étape 1 : Encodage

C'est l'étape de codage ou de représentation du texte en clair sous forme de chromosome. On suppose que T contient les caractères suivants : c_1, c_2, c_3,..., c_m.

Chaque caractère apparaît au moins à une position dans le texte, puis on définit pour chaque caractère sa liste de positions dans ce texte appelée Li ($0 < i < m+1$). Ainsi, le texte en clair est représenté par le vecteur $T = \{(c_1, L_1) \ldots (c_m, L_m)\}$ qui sera le chromosome initial de toutes les populations.

A travers cette représentation, nous pouvons récupérer deux propriétés importantes de cette population, qui sont :

① $L_i \cap L_j = \emptyset$, pour $i, j \in [1, m]$, avec $i \neq j$.

② $L_1, L_2 .., L_m$ est une partition de l'ensemble $\{1, 2 ..., n\}$

Etape 2 : Génération de la population initiale

Soit:

- q est la taille de la population.
- CHj (j ∈ [1, q]) est la représentation de chaque chromosome.
- Et P1 est la représentation de la population initiale.

Ensuite,

- la première population sera représentée par : $P_1 = \{CH_{11}, CH_{12}, CH_{13},\ldots, CH_{1q}\}$
- La seconde est : $P_2 = \{CH_{21}, CH_{22}, CH_{23},\ldots, CH_{2q}\}$,
- et ainsi de suite, chaque chromosome CHj (j ∈ [1, q]) est défini comme une nouvelle combinaison entre les c_i et L_i.

La première population générée ne doit pas suivre une fonction bien définie mais elle doit s'appuyer sur des événements aléatoires pour la générer car plus la génération initiale de la population est aléatoire, plus l'algorithme est efficace.

Étape 3 : Évaluation

Dans cette étape, il faut évaluer la partition aléatoire Ej construite à partir de chaque chromosome CHj telle que : $Ej = \{e_{j1}, e_{j2}, \ldots, e_{jm}\}$.

Ensuite, il faut attribuer une valeur à chaque partition afin d'évaluer son efficacité à l'aide de la formule suivante :

$$F(\mathrm{Xj}) = \sum_{i=1}^{m} |Card\,(eji) - [n/m]|$$

Grâce à cette fonction, nous essayons de trouver la partition dont tous les éléments ont une cardinalité similaire.

Étape 4 : Sélection

La méthode de sélection utilisée est la sélection de la roulette. Comme son nom l'indique, le principe de cette fonction est basé sur les performances de la roulette de casino. Il peut transformer la performance de chaque parent en une probabilité qui sera distribuée plus tard sur la roulette du jeu. On choisit au hasard la valeur du paramètre « r » que l'on peut considérer comme la boule à lancer sur la roue pour choisir le chromosome élu.

Étape 5 : Opérateurs génétiques

Deux opérateurs sont à appliquer dans cette étape : l'opérateur du croisement et l'opérateur de la mutation.

- Le croisement:

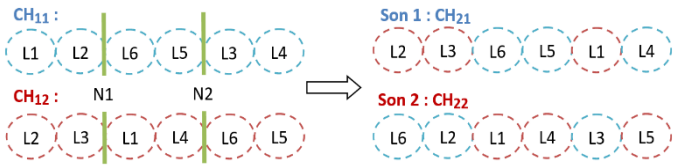
Le croisement utilisé dans cet algorithme est le MPX dans lequel le codage du fils a une forte analogie avec celui des parents.

Le choix de cet algorithme est dû au fait qu'il maintient les caractéristiques de la population d'origine qui sont :

① $Li \cap Lj = \emptyset$, for $i, j \in [1, m]$, with $i \neq j$.

② L1, L2, ..., Lm est une partition de l'ensemble {1, 2, ..., n}

Ce croisement peut être illustré par l'exemple ci-dessous :



- La Mutation:

L'opérateur de la mutation est appliqué avec une probabilité très faible. Il consiste en une permutation entre quelques listes de la population.

1.2 Les variantes de SEC

Plusieurs variantes ont été créées à la base de ce chiffrement, on site parmi elles :

- SEC binaire : cette version permet de considérer le codage binaire du message et réalise l'encodage en utilisant les positions de différents blocs de bits successive de taille k tel que $k \neq 8*x$
- SEC avec Fusion : dans cette variante l'auteur cherche à fusionner les listes ayant une taille très petite dans une seul liste et l'associer à un nouveau caractère différents de ceux du message d'origine.
- SEC avec Fragmentation : à l'encontre de la variante utilisant la Fusion pour équilibrer la taille des listes, la méthode de la fragmentation permet de diviser les

listes ayant une grande taille en des listes de même taille proche de la moyenne des positions.

- SEC avec partition : dans cette variante, les auteurs proposent de modifier dans le contenu des listes des positions afin d'équilibrer leurs tailles en utilisant le problème de Partition. Ce qui augmente sa résistance.