

REPUBLIQUE DEMOCRATIQUE DU CONGO

*MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET
UNIVERSITAIRE*

UNIVERSITE DE KINSHASA

CAMPUS DE KINSHASA

Projet d'examen Python Avancée

Rapport :

Thème : Sécurité Informatique

Implémentation des Techniques de Cryptographie en
Python



Roger MAWETE
DEA1 en Informatique

Dirigé par :Prof Jordan F. Masakuna

Année Académique 2024-2025

TRAVAIL DEMANDE A FAIRE :

- Selon Le Sujet, Trouver Un Data Set Approprié.
- Élaborer La Problématique Du Sujet (Améliorer La Description Du Projet).
- Élaborer Une Brève Revue De Littérature Sur Le Sujet.
- Identifier Quelques Questions De Recherche (Au Moins Cinq Questions).
- Créer Un Fichier Texte (Requirements.Txt) Contenant Toutes Les Dépendances Requises.
- Installer Des Nouvelles Librairies Si Nécessaires Répondre Aux Questions En Utilisant Python (Les Paramètres De La Solution À Garder Dans Un Dictionnaire).
- Discuter/Commenter Des Résultats Obtenus.

SUJET CHOISI***Sécurité Informatique***

Implémentation des Techniques de Cryptographie en Python

PROBLEMATIQUE

Les communications ont toujours constitué un aspect important dans l'acquisition de nouvelles connaissances et l'essor de l'humanité. Le besoin d'être en mesure d'envoyer un message de façon sécuritaire est probablement aussi ancien que les communications elles-mêmes.

D'un point de vue historique, c'est lors des conflits entre nations que ce besoin a été le plus vif.

Dans notre monde moderne, où diverses méthodes de communication sont utilisées régulièrement, le besoin de confidentialité est plus présent que jamais à une multitude des niveaux.

La cryptographie s'intéresse à la transmission et réception confidentielles des messages et des données.

L'importance croissante de la sécurité des informations dans le monde numérique actuel, avec l'augmentation du cyber attaques et des violations de données, il est crucial de sécuriser les communications.

Python est un langage généraliste de programmation interprété qui a la particularité d'être très lisible et pragmatique. Il dispose d'une très grosse base de modules externes, notamment scientifiques, qui le rend particulièrement attractif pour programmer des problèmes mathématiques et scientifiques voir en sécurité informatique. Le fait que Python soit un langage interprété le rend plus lent que les langages compilés, mais il assure en revanche une grande rapidité de développement qui permet à l'humain de travailler un peu moins tandis que l'ordinateur devra travailler un peu plus. Cette particularité a fait que Python est devenu l'un des principaux langages de programmation utilisés par les scientifiques.

Ce projet d'examen intervient après avoir suivi le cours de Python avancée qui selon l'animateur a bien voulu nous mettre en exercice, ainsi nous étions motivés de travailler sur la sécurité informatique précisément en implémentant des techniques de cryptographie en Python.

QUESTIONS DE RECHERCHE

1. La Cryptographie, Qu'est-ce ? les origines et importances ?
2. Quels sont les systèmes cryptographiques les plus utilisés ?
3. Comment implémenter ces techniques de cryptographie en python ?
4. Quels outils performants qui peuvent faciliter l'implémentation des techniques cryptographiques en Python ?
5. Quels sont les modules nécessaires pour implémenter ces techniques de cryptographie en python ?

REVUE DE LA LITTÉRATURE

Clarification du concept¹

La cryptographie est donc l'étude des méthodes d'envoi de messages codés de telle sorte que seule le destinataire puisse le décoder. Le message qu'on veut envoyer s'appelle le texte clair et le message codé, ou encryté, s'appelle aussi cryptogramme.

Le processus de conversion d'un texte clair en message codé s'appelle chiffrement, ou codage ; et le processus inverse s'appelle déchiffrement, ou décodage. Pour effectuer un codage, on suit une méthode précise appelée système de codage, ou système cryptographique, ou même encore crypto système. Un codage se fait donc à l'aide d'un système cryptographique, et celui-ci nécessite très souvent l'utilisation d'une clé de codage. Cette clé (un mot, un nombre, une grille) est nécessaire pour décoder le message chiffré. En d'autres termes, la clé modifie le comportement du mécanisme de codage et de décodage.

Les symboles utilisés dans un message sont appelés des lettres et l'ensemble des symboles possibles s'appelle l'alphabet. On désigne souvent l'alphabet par A. L'alphabet du texte clair peut être différent de l'alphabet du message codé. Le texte clair et le texte chiffré sont souvent découpés en blocs. L'intention derrière le découpage en blocs est habituellement d'envoyer le texte comme une succession de blocs qui sont encodés et décodés séparément.

La cryptanalyse est l'étude des méthodes qui permettent de découvrir le sens d'un message codé, sans connaître le message original. Il y a plusieurs situations possibles. On peut vouloir simplement trouver le sens du message codé, sans chercher à trouver la clé de codage. Mais, en général on voudra trouver d'abord quel est le système de codage, puis la clé de codage utilisée. Lorsqu'on a trouvé tous les éléments de la méthode utilisée pour coder des

¹ Bergeron F. et all (2014). La cryptographie de l'antiquité à l'internet. UQAM. Québec

messages, on dit qu'on a cassé, ou brisé, le système cryptographique utilisé. Plus un système difficile à briser, plus il est sûr.

Histoire de la cryptographie

Le terme (cryptographie) vient en effet des mots grecs anciens : *kruptos* (κρυπτος) qu'on peut traduire comme (secret) ou (cachée) ; et *graphein* (γραφειν) pour « écriture ». Plus précisément, la cryptographie est l'étude des codes secrets, et non celle des messages simplement voilés (comme avec de l'encre invisible, par exemple).

Les origines de la cryptographie semblent remonter à plus de 4000 ans. On a trouvé, sur une tombe égyptienne de cette époque, des inscriptions contenant des hiéroglyphes modifiés, et il semblerait qu'on ait cherché par ces modifications à obscurcir le sens des inscriptions. Quoi qu'il en soit, plusieurs indications archéologiques tendent à montrer que les (écritures secrètes) sont en fait aussi anciennes que l'invention de l'écriture elle-même.

Le premier exemple indéniable de cryptographie remonte au moins au 5^{ème} siècle avant notre ère. En effet, les Spartiates (Grèce) du temps avaient développé une méthode originale pour l'échange de messages secrets. Celle-ci est basée sur le fait que deux copies identiques d'un bâtonnet, appelée scytale, soient en possession de l'envoyeur et du récepteur du message. Pour préparer un message, on enroule en spirale autour de la scytale une bandelette de parchemin (ou de cuir), pour ensuite écrire le message le long de la scytale. Une fois déroulée la bandelette ne contient plus qu'une suite apparemment incompréhensible de lettres. Cependant, pour décoder le message il suffit simplement d'enrouler la bandelette sur la scytale jumelle. Comme la méthode est assez simple, il leur fallait bien entendu la conserver secrète.

En 44 avant notre ère, Jules César utilisait une simple méthode de substitution de lettres pour communiquer secrètement avec ses généraux. Dans son système cryptographique, connu comme le code de César, on place les 26 lettres de l'alphabet dans l'ordre habituel et le message codé est obtenu en décalant circulairement chaque lettre du message clair de trois positions.

Autrement dit, on 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 |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| 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 |

Pour illustrer, le mot POURQUOI devient SRXUTXRL. Bien que le résultat semble tout à fait incompréhensible, il faut chercher à casser le code de César ici le terme casser veut tout simplement dire décoder les messages secrets.

Du côté européen, en 1379, Gabriel de Lavinde fait de la cryptographie une science mieux comprise, en publiant le premier manuel sur le sujet. Il y présente sa compilation des systèmes de codage connus. Plusieurs ouvrages d'autres auteurs suivront. Plusieurs se mettent à décrire de nouveaux systèmes de codage, ainsi que des mécanismes pour faciliter ces codages. Ainsi, dans un traité publié en 1466, l'italien Leon Battista Alberti décrit la construction d'outils de codage comme son cadran, qui facilitent les codages poly alphabétiques. On attribue souvent au français Blaise de Vigenère le développement, en 1586, de ce qui fut longtemps considéré comme un « chiffre indéchiffrable ». Cependant, la paternité de ce système reviendrait plutôt à Giovan Batista Belaso, en 1553, et Vigenère en aurait simplement clarifié certains aspects.

On trouve apparemment bien moins de détails sur les efforts cryptographiques durant la guerre froide, probablement parce que ces informations sont encore trop secrètes. Il est possible que certains des systèmes plus modernes aient été considérés plus secrètement à cette époque.

On en est maintenant à l'époque moderne des codes à clés publiques, basés sur diverses notions mathématiques, avec toutes les applications nouvelles suscitées par l'utilisation de l'Internet, sans mentionner les utilisations potentiellement plus problématiques comme celles liés au terrorisme. On a aussi des indications claires de tendances à venir comme la cryptographie quantique, basée sur les principes de la théorie des quanta.

Au final, les applications potentielles des outils développés pour la cryptanalyse sont aujourd'hui encore plus variées, incluant par exemple celles dans le domaine de l'étude de génomes.

Utilisation courante de la cryptographie

Des systèmes cryptographiques de toute sorte sont utilisés de façon courante. Tous ne nécessitent pas le même niveau de sécurité, et les systèmes cryptographiques utilisés varient beaucoup en complexité. Dans certains cas les codages sont très simples, et dans d'autres on cherche à assurer la meilleure sécurité disponible. Les utilisations vont de la téléphonie cellulaire, aux transactions bancaires, en passant par le cryptage de certaines chaînes de télévision sans mentionner les communications diplomatiques, militaires, ou encore terroristes ou criminelles. On comprend donc que, dans certains cas, les impératifs de facilité et de rapidité de codage l'emportent sur l'assurance d'une sécurité absolue.²

LES ALGORITHMES CRYPTOGRAPHIQUES UTILISES

Chiffrement de Hill (codage alpha-numérique-matrice transposée)

Cette méthode reste emblématique des systèmes basés sur des matrices utilisant les transformations linéaires de blocs de texte.

On commence par regrouper les lettres du texte en clair en blocs de m caractères. On numérise ces blocs, puis on les code au moyen d'une certaine matrice. Pour numériser les lettres, on procède très souvent de la façon suivante :

² Schneier B. (1996). Cryptographie appliquée, 2ème edition. John Wiley and sons. London

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 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 | , | ! | ? |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |

Tous les calculs s'effectuent ici modulo le nombre de lettres de l'alphabet de référence.

On travaillera dans ce mémoire donc avec 28 lettres, tout simplement en ajoutant le caractère d'espacement () et de ponctuation apostrophe (')

Après avoir chiffré le message, on obtient une matrice composée des chiffres issus du message puis on multiplie cette matrice par la matrice mot de passe choisie.

Si les dimensions de la matrice issue du message ne sont pas compatibles pour la multiplication avec la matrice mot de passe, on ajoute au message un caractère de l'alphabet choisie au Hazard autant de fois que possible pour le rendre compatible.

Après multiplication, on obtient ainsi une matrice composée des chiffres qui après déchiffrement représente le message crypté.

En cas des nombres négatifs après chiffrement, pour déchiffrer, on ajoute au nombre négatif le nombre de lettres de l'alphabet autant de fois jusqu'à trouver un nombre positif et ensuite le déchiffrer selon l'alphabet.

Code de Vigenère

En 1550, le diplomate français Blaise de Vigenère (1523-1596), secrétaire de Charles IX, voyage partout en Europe et se familiarise avec les méthodes cryptographiques connues pour des raisons professionnelles. Dix ans plus tard, il abandonne sa carrière de diplomate, et se consacre exclusivement à l'étude détaillée des écrits d'Alberti, de Trithème et de Porta. Il donne la forme finale à un nouveau chiffrement puissant auquel on donne plus tard son nom.

La force du chiffre de Vigenère vient du fait qu'une même lettre en clair peut être chiffrée de différentes manières. A partir de l'invention de l'imprimerie, vers 1450, l'emploi des messages codés se généralise dans les relations diplomatiques entre les états européens et chez les militaires. L'art du chiffrement et du déchiffrement évolue plus rapidement et des scientifiques renommés contribuent au développement des techniques utilisées. Le mathématicien Cardano, de Milan (célèbre pour sa résolution des équations du troisième degré), l'architecte Alberti de Florence, et l'Abbé Trithème font évoluer la science du chiffre.

Dans le traité « des secrètes manières d'écrire », Vigenère décrit le chiffre qu'il qualifie d'indéchirable. Le procédé de Vigenère, fondé sur la tabula recta de Trithème, consiste à changer l'alphabet de substitution à chaque chiffrement d'une lettre, ce qui fait qu'on ne peut tenter de décrypter le message en utilisant simplement un calcul de fréquence des lettres.

DESCRIPTION DU CODE DE VIGENERE

Le chiffrement dépend d'un mot clé, dans l'exemple ci-dessous c'est le mot PERMUTE. Pour coder un mot en clair, comme SECURITE, on consulte la une table appelé tabula recta a la ligne commençant par la première lettre, P, du mot clé. On remplace alors la première lettre, S, du mot en clair par son correspondant H sur cette ligne. On procède de même pour la seconde lettre E du mot en clair, mais on utilise maintenant la ligne commençant par la seconde lettre E du mot clé. On continue ainsi pour les autres lettres, en recommençant au début du mot clé si nécessaire. Le chiffrement du mot en clair SECURITE donne donc le mot codé HITGLBXT.

Le système de Vigenère ne sera déchiffré qu'au milieu du XIXème siècle, et demeurera à la base de la plupart des machines à chiffres, jusqu'au début du XXe siècle.

Algorithme AES (Advanced Encryption Standard)

AES est un algorithme de chiffrement symétrique, choisi en octobre 2000 pour être le nouveau standard de chiffrement pour les organisations du gouvernement des Etats-Unis.

Il est issu d'un appel à candidatures international lancé en 1997 et ayant reçu 15 propositions. Au bout de cette évaluation, ce fut le candidat Rijndael (« prononcé Rayndal »), du nom de ses deux concepteurs Joan Daemen et Vincent Rijmen (tous les deux Belge) qui a été choisi.

Principe

L'algorithme prend en entrée un bloc de 128 bits (la clé fait 128, 192) ou 256 bits. Les 128 bits en entrée sont « mélangés » selon une table définie au préalable.

- Ces octets sont ensuite placés dans une matrice de 4X4 éléments et ses lignes subissent une rotation vers la droite
- L'incrément pour la rotation varie selon le numéro de la ligne.
- Une transformation est ensuite appliquée sur la matrice par un XOR avec une matrice clé.
- Finalement, un XOR entre la matrice et une autre matrice permet d'obtenir une matrice intermédiaire.
- Ces différentes opérations sont répétées plusieurs fois et définissent un « tour ».
- Pour une clé de 128, 192 ou 256, AES nécessite respectivement 10, 12 ou 14 tours.

L'algorithme AES n'est pas vraiment cassé à la date d'aujourd'hui.

DATASET

Afin de répondre aux consignes demandés par l'enseignant, nous avons utilisé SMSSpamCollection dataset classique contenant des SMS étiquetés comme "spam" ou "ham" (non-spam). Il est disponible sur Kaggle ou UCI Machine Learning Repository

RESULTATS ET DISCUSSIONS

1. TECHNIQUE DE CRYPTOGRAPHIE IMPLEMENTE

- Chiffrement de César
- Chiffrement de Vigenère
- Chiffrement AES (Advanced Encryption Standard)

Il convient de noter que nous avons adapter par la suite, le crypto système de Hill Cipher à notre propre CryptoSystème dont nous avons implémenter dans cette application.

2. LES MODULES

A cet effet, trois modules ont été implémentés : Cryptographie.py ; hill_cipher.py et gui.py

Cryptographie

Nous avons créé des fonctions ayant les fonctionnalités suivantes :

Chargement du Dataset : La fonction 'load_dataset' utilise pandas pour charger les données à partir d'un fichier xlsx.

Chiffrement de César : La fonction cesar_cipher prend un texte et un décalage en entrée, puis chiffre ou déchiffre le texte en décalant chaque lettre.

Chiffrement de Vigenère : La fonction vigenere_cipher prend un texte, une clé et un indicateur de

chiffrement/déchiffrement. Elle chiffre ou déchiffre le texte en utilisant la clé de Vigenère.

Chiffrement AES :

`generate_key` génère une clé AES aléatoire.

`encrypt_message` chiffre le message avec la clé AES.

`decrypt_message` déchiffre le message chiffré avec la même clé.

Hill_Cipher

Nous avons adapté l'algorithme de Hill Cipher à notre manière pour crypter et décrypter des messages.

GUI

GUI et le module des interfaces graphiques utilisateurs, dans ce module nous avons créé trois interfaces :

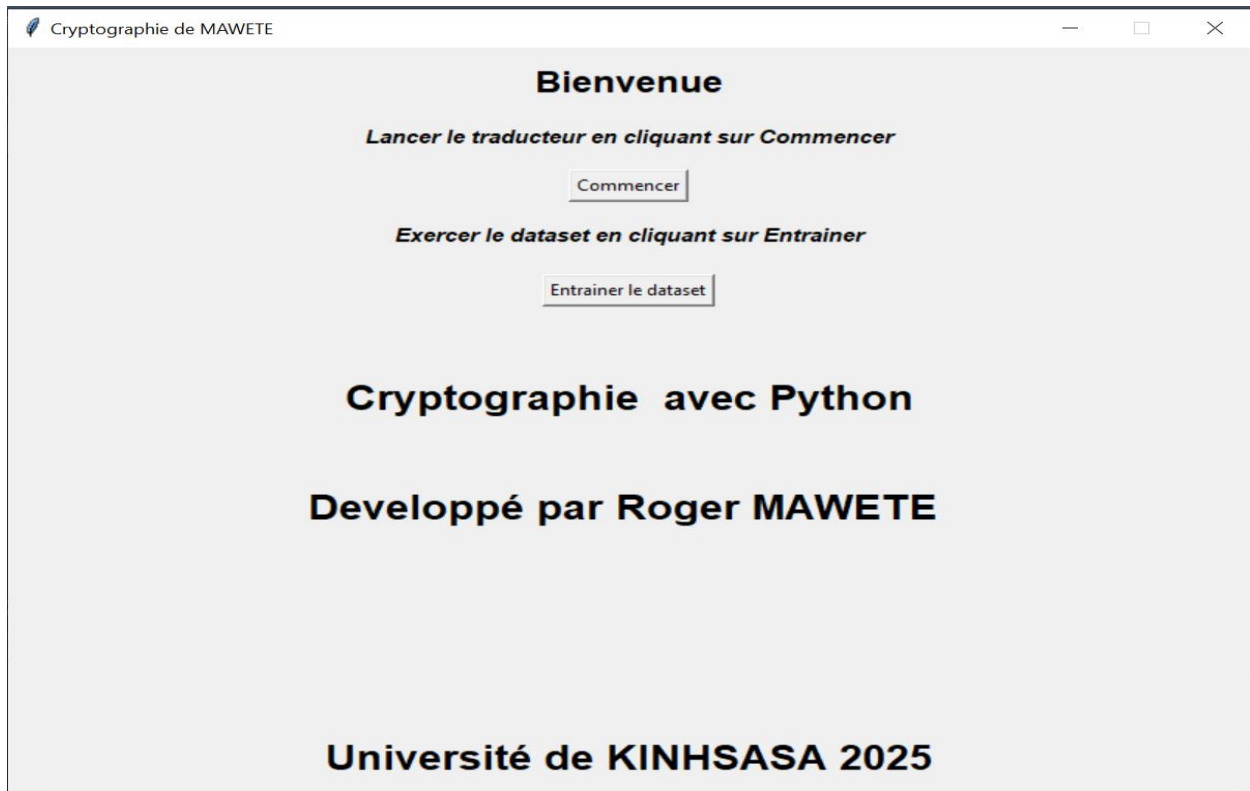
Le menu principal

L'interface pour entrainer le dataset qui génère de façon aléatoire les messages dans notre dataset et crypte ce message selon les algorithmes implémenter dans le module Cryptographie et représenter par les boutons des commandes présents.

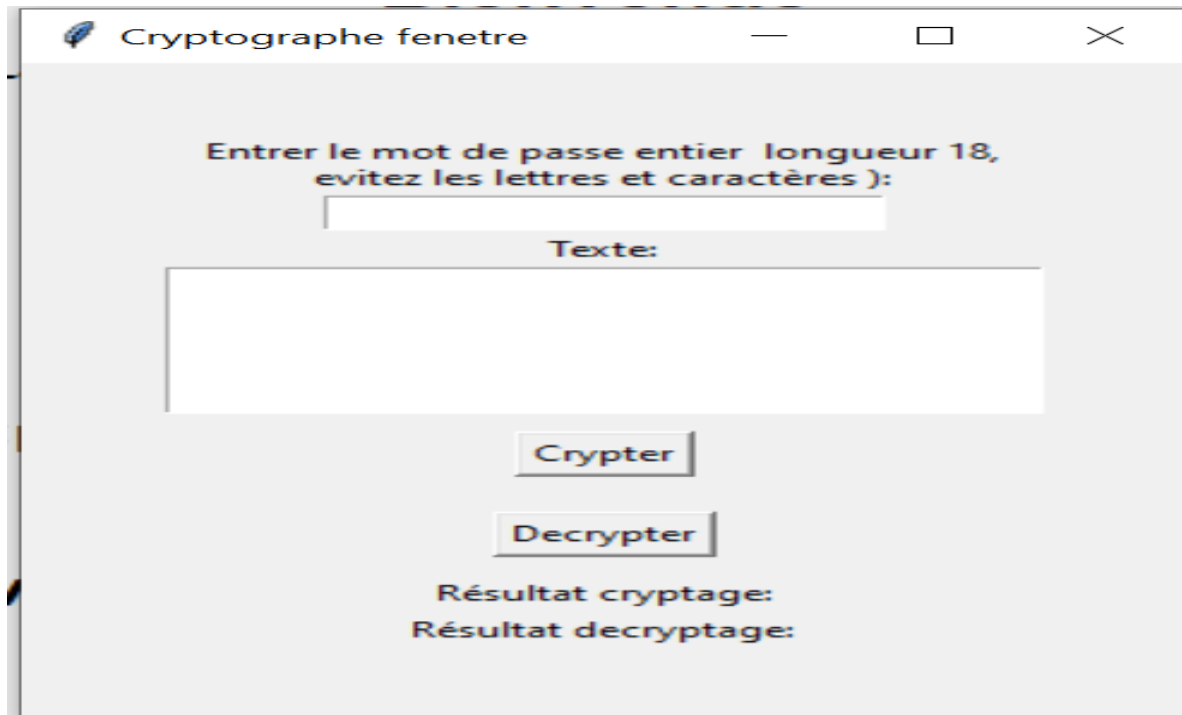
L'interface pour crypter et décrypter le message selon l'algorithme adapté de Hill_Cipher qui laisse à l'utilisateur la possibilité d'entrer un message, une clé et l'application code ou décode le message entré selon Hill.

PRESENTATION DES INTERFACES UTILISATEURS

MENU PRINCIPAL

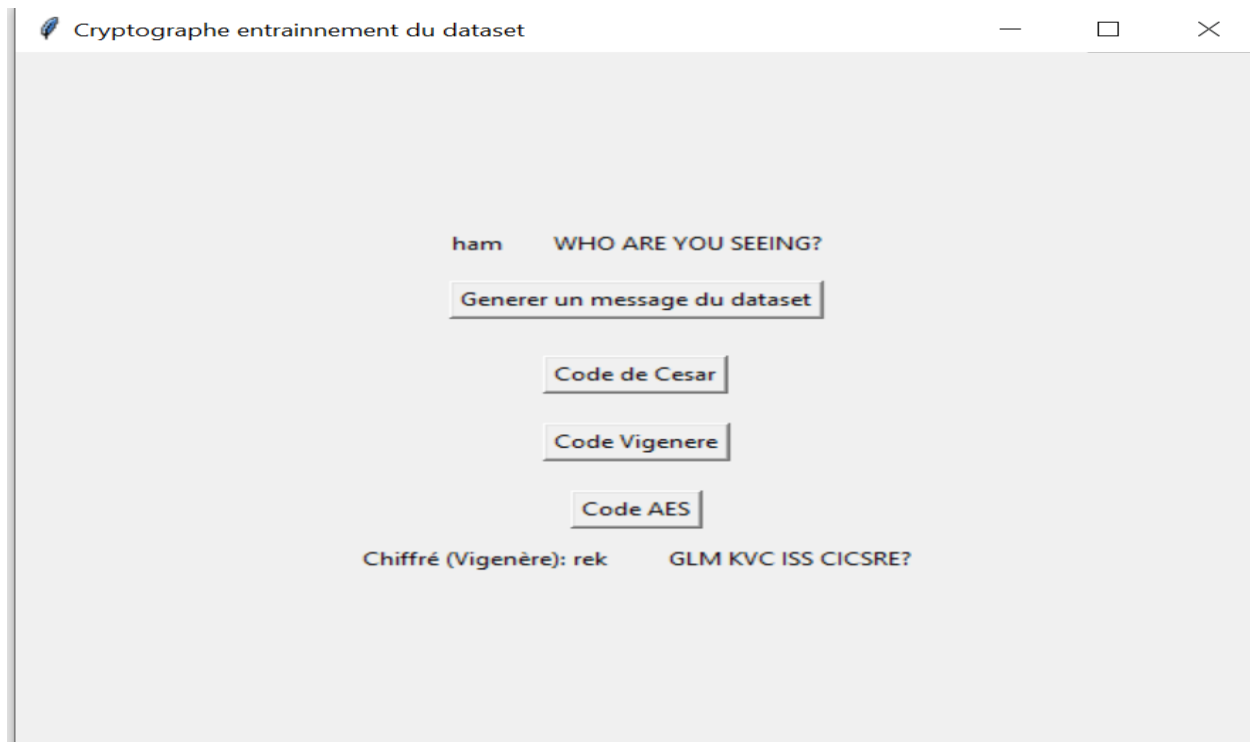


FENETRE DE CRYPTOGRAPHIE



A Java window titled "Cryptographe fenetre" with standard window controls. The window contains a text label "Entrer le mot de passe entier longueur 18, evitez les lettres et caractères):" followed by a small text input field. Below this is a larger text area labeled "Texte:". At the bottom, there are two buttons labeled "Crypter" and "Decrypter", followed by two labels "Résultat cryptage:" and "Résultat decryptage:".

ENTRAINNEMENT DANS LE DATASET



A Java window titled "Cryptographe entraînement du dataset" with standard window controls. The window displays a sample message: "ham WHO ARE YOU SEEING?". Below the message is a button labeled "Generer un message du dataset". Underneath the button are three buttons labeled "Code de Cesar", "Code Vigenere", and "Code AES". At the bottom, there are two labels: "Chiffré (Vigenère): rek" and "GLM KVC ISS CICSRE?".