

**Đại học Quốc gia Tp.Hồ Chí Minh**

Trường Đại học Bách Khoa  
Khoa Khoa học và Kỹ thuật Máy tính



**Mật mã an ninh mạng**

# Bài tập lớn số 1

Nhóm	Họ và tên	MSSV
1	Nguyễn Hoàng Phúc	1927030
	Đoàn Thị Mỹ Hằng	1827013
	Huỳnh Thị Kim Phượng	1827026

HK 2 – Năm học 2019- 2020

## **Chương 1: Giới thiệu**

Yêu cầu: Bạn cần phải xây dựng một chương trình mã hóa bằng một ngôn ngữ lập trình bất kỳ mà bạn chọn. Sau đó bạn sẽ tạo ra một số bản mã. Cuối cùng bạn sẽ cố gắng phân tích các bản mã trên. Cho đơn giản, chúng ta giả sử rằng đầu vào là các ký tự trong tập {a-z, A-Z, 0-9} và ký tự khoảng trắng

Nhóm sẽ đi sâu vào trình bày các bước phân tích và thiết kế hệ thống dựa trên các yêu cầu đặt ra ở chương 2 - Phân tích và thiết kế hệ thống

Sau đó dựa vào những bước phân tích thiết kế hệ thống ở chương 2, chúng ta sẽ hiện thực hệ thống. Nhóm trình bày chi tiết phần hiện thực hệ thống, cách thức và kết quả đánh giá hệ thống đã xây dựng ở chương 3 - Hiện thực và kết quả.

Cuối cùng, nhóm xin trình bày những gì hệ thống đã làm được và không làm được theo các yêu cầu đã đặt ra, đánh giá ưu và nhược điểm của hệ thống, hướng phát triển của hệ thống ở chương 4 – Đánh giá hệ thống và kết luận

## **Chương 2: Phân tích và thiết kế hệ thống**

Trong bài tập lớn này, nhóm sử dụng ngôn ngữ lập trình Python để xây dựng chương trình mã hóa. Và nhóm sử dụng tiếng anh là ngôn ngữ của bản rõ plaintext dùng để mã hóa và giải mã.

### **Chương trình sẽ diễn tả các chức năng như sau:**

- Từ một bản rõ tạo ra một bản mã dùng mã hóa thay thế đơn ký tự. Nói cách khác khóa sử dụng là một quy tắc thay thế cho mỗi ký tự đầu vào.
- Từ một bản rõ tạo ra một bản mã dùng mã hóa hoán vị. Chúng ta giả sử rằng hệ mã này sẽ làm việc trên một khối có 8 ký tự. Nói cách khác, nó luôn luôn chuyển vị các ký tự trong một khối có 8 ký tự.
- Từ một bản rõ tạo ra một bản mã dùng mã hóa nhân dựa trên hai hàm nói trên. Giả sử mã hóa thay thế được dùng trước sau đó mã hóa hoán vị được dùng để mã hóa kết quả và lấy

ra được bản mã cuối cùng.

- Tạo ba phương thức với ba hàm nói trên. Mỗi phương thức (với khóa cố định), mã hóa bản rõ có chiều dài tùy ý và ít nhất là 1000 ký tự.
- Sau đó tạo ra các bản mã dùng một trong ba phương thức trên (với khóa khác nhau) và bắt đầu phải thiết kế các phương pháp để tìm ra bản rõ tương ứng với bản mã đã có. Bắt đầu với (1), tiếp tục với (2) và (3).
  1. Xây dựng công cụ hay dùng công cụ có sẵn để phân tích mã các hệ mã nói trên.
  2. Cố gắng lấy các bản rõ từ các bản mã.
  3. Cố gắng lấy khóa đã sử dụng.

### **Chương 3: Hiện thực và kết quả**

#### **a. Hiện thực**

Source code gồm 3 file chương trình chính là encode.py, decode.py và fuction.py. Và 2 file text là plaintext.txt chứa đoạn văn bản dùng để mã hóa và common\_words\_list.txt có chức năng như 1 cuốn từ điển chứa các từ tiếng anh thông dụng.

#### **File fuction.py chứa 5 method sau:**

**DocFile(path)** với tham số đầu vào là path chứa file input để read file.

**LuuFile(path, data)** với 2 tham số đầu vào là path, data dùng để ghi file trả về 1 mảng với mỗi phần tử là từng dòng của file.

**Caesar\_cipher\_encode(plaintext, n)** với 2 tham số đầu vào là plaintext: một chuỗi ký tự cần mã hóa và n: là bước dịch chuyển để mã hóa theo giải thuật Caesar, kết quả của hàm trả về chuỗi ký tự đã được mã hóa.

**cut\_block\_8words(s)** với tham số s: 1 chuỗi ký tự về 1 mảng với mỗi phần tử là 1 block chứa 8 ký tự.

**permutation\_block(arr, permutation\_Table)** với tham số đầu vào là arr : mảng có mỗi phần tử chứa 8 ký tự và permutation\_Table: chuỗi hoán vị 8 số từ 1->8 trả về 1 chuỗi ký tự đã được mã hóa.

**File encode.py** là chương trình mã hóa gồm 3 giải thuật mã hóa là Caesar, Permutation, Mix (kết hợp Caesar và Permutation) để người dùng lựa chọn.

Giải thuật Caesar được hiện thực bằng cách gọi hàm DocFile để đọc vào file plaintext.txt và lấy độ dịch chuyển để mã hóa từ user input nhập vào (đây cũng chính là key mã hóa), rồi gọi method Caesar\_cipher\_encode để mã hóa từng dòng của file plaintext để trả về cipher là từng dòng mã hóa được lưu ở ciphertext.txt, key dùng để mã hóa được lưu ở file key.txt.

Giải thuật Permutation được hiện thực bằng cách lấy permutation\_Table từ user input là chuỗi 8 ký tự 1-8, dùng method cut\_block\_8words để tạo thành những block 8 ký tự và gọi hàm permutation\_block để thực hiện hoán vị các ký tự của file input nhập vào .

Giải thuật nhân kết hợp 2 giải thuật trên thực hiện mã. Đầu tiên file input sẽ được mã hóa bằng method Caesar\_cipher\_encode theo độ dịch chuyển Caesar thành những dòng cipher sau đó tiếp tục dùng hàm cut\_block\_8words để chuyển cipher thành những block 8 ký tự và cuối cùng dùng method permutation\_block để hoán vị các dòng cipher vừa tạo ra được lưu vào file ciphertext.txt

### **Chọn phương pháp mã hóa nhân để phân tích và thiết kế hàm giải mã:**

Sau đó nhóm tạo ra các bản mã dùng phương thức kết hợp giữa giải thuật Caesar và Permutation (với khóa khác nhau) để thiết kế các phương pháp để tìm ra bản rõ tương ứng với bản mã đã có và việc giải mã được hiện thực ở file decode.py.

File decode.py là chương trình thực hiện phân tích mã dựa vào đặc điểm ngôn ngữ tiếng anh, đặc biệt phân tích dự theo ký tự khoảng trắng.

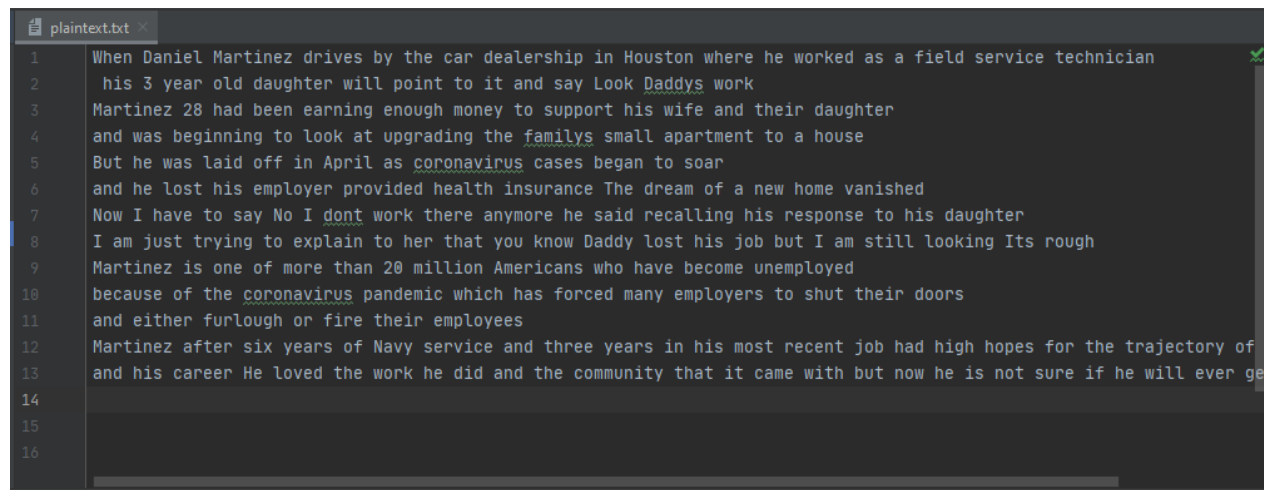
Phạm vi mã hóa bản rõ có chiều dài ít nhất là 1000 ký tự.

Trung bình thì mỗi từ (word) tiếng anh sẽ không quá 10 ký tự ,vậy 1000 ký tự sẽ tương ứng với 100 word. Vậy nên mỗi bản rõ plaintext sẽ có ít nhất 99 khoảng trắng. Vậy chúng ta cho ký tự có tần số xuất hiện nhiều nhất > 99 lần là ký tự khoảng trắng và lấy key đó lưu vào danh sách key ở hack\_key.txt. Giá trị key có thể không duy nhất do tần số xuất hiện của các nguyên âm cũng khá nhiều, nên dựa vào bước giải mã tiếp theo để lấy được key Caesar trong trường hợp có nhiều hơn 1 key Caesar.

Cuối cùng thực hiện hoán vị cho chỉnh hợp 8 số và dựa vào danh sách những từ tiếng anh thông dụng ở file `common_word_list.txt` để tách thành từng từ tiếng anh có nghĩa. Nếu số lượng từ tiếng anh có nghĩa lớn hơn 7,5% của list thì ta lưu kết quả đó ở file `hack_plaintext.txt`.

## b. Kết quả hiện thực

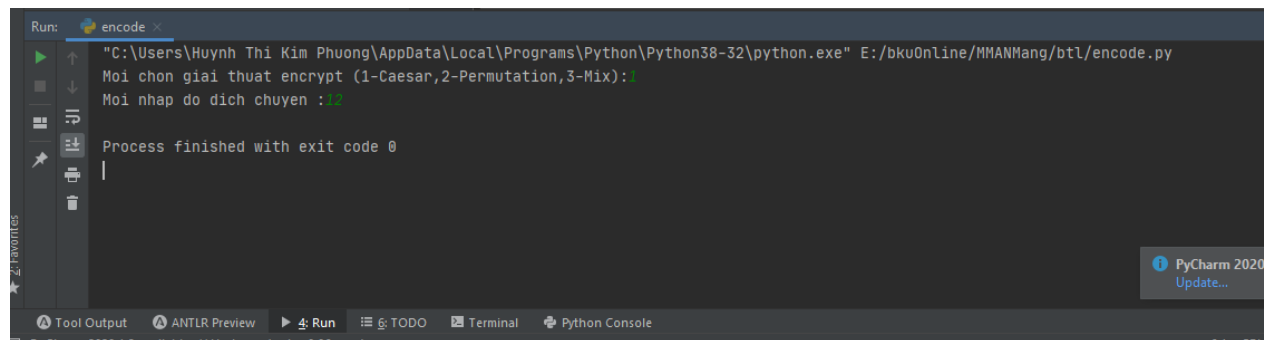
Với bản rõ ban đầu như sau:



```
plaintext.txt
1 When Daniel Martinez drives by the car dealership in Houston where he worked as a field service technician
2 his 3 year old daughter will point to it and say Look Daddys work
3 Martinez 28 had been earning enough money to support his wife and their daughter
4 and was beginning to look at upgrading the familys small apartment to a house
5 But he was laid off in April as coronavirus cases began to soar
6 and he lost his employer provided health insurance The dream of a new home vanished
7 Now I have to say No I dont work there anymore he said recalling his response to his daughter
8 I am just trying to explain to her that you know Daddy lost his job but I am still looking Its rough
9 Martinez is one of more than 20 million Americans who have become unemployed
10 because of the coronavirus pandemic which has forced many employers to shut their doors
11 and either furlough or fire their employees
12 Martinez after six years of Navy service and three years in his most recent job had high hopes for the trajectory of
13 and his career He loved the work he did and the community that it came with but now he is not sure if he will ever ge
14
15
16
```

## ❖ Kết quả quá trình encryption sau khi run `encode.py`

Chọn giải thuật Caesar với mode 1



```
Run: encode
"C:\Users\Huỳnh Thị Kim Phượng\AppData\Local\Programs\Python\Python38-32\python.exe" E:/bkuOnline/MMANMang/btl/encode.py
Moi chon giai thuat encrypt (1-Caesar,2-Permutation,3-Mix): 1
Moi nhap do dich chuyen : 12
Process finished with exit code 0
```

Với key là độ dịch chuyển 12

```
key.txt x
1  Đo dịch chuyen Caesar cipher là: 12
2
```

Ta được bản mã cipher sau :

```
ciphertext.txt x
1  h1tqzLPmzuqxLYm35uzqALp3u7q4Ln L5tqLom3Lp9mxg34tu1LuzLT06450zL8tq3qLtqL803wqpLm4LmLruqxpL4q37uoqL5qotzuoumz
2  t04LEL qm3L0xplpm6st5q3L8uxxL10uz5L50Lu5Lmzpl4m LX00wLPmp 4L803w
3  Ym35uzqALDjLtmplnqqzLqm3zuzsLqz06stLy0zq L50L4611035Ltu4L8urqLmzpl5tqu3Lpm6st5q3
4  mzpL8m4LmqsuuzzsL50Lx00wLm5L61s3mpuzsL5tqLrmyux 4L4ymxxLm1m35yqz5L50LmLt064q
5  N65LtlqL8m4LxmupL0rrLuzLM13uxLm4Lo030zm7u364Lom4q4LnqsmzL50L40m3
6  mzpLtlqLx045Ltu4Lqy1x0 q3L1307upgpLtgmx5tLuz463mzoqLftqLp3qmyL0rLmLzq8Lt0yqL7mzu4tqp
7  Z08LULtm7qL50L4m LZ0LULp0z5L803wL5tq3qLmz y03qLtlqL4mupL3gomxxuzsLtu4L3q410z4qL50Ltu4Lpm6st5q3
8  ULmyLv645L53 uzsL50Lq91xmuzL50Ltlq3L5tm5L 06Lwz08LPmp Lx045Ltu4Lv0nLn65LULmyL45uxxLx00wuzsLU54L306st
9  Ym35uzqALu4L0zqL0rLY03qL5tmzLDBLyuxxu0zLMyq3uomz4L8t0Ltm7qLnqo0yqL6zqy1x0 qp
10 nqom64qL0rL5tqLo030zm7u364L1mzpqyuoL8tuotLtm4Lr03oqpLymz Lqy1x0 q34L50L4t65L5tqu3Lp0034
11 mzpLqu5tq3Lr63x06stL03Lru3qL5tqu3Lqy1x0 qq4
12 Ym35uzqALmr5q3L4u9L qm34L0rLZm7 L4q37uoqLmzpl5t3qqL qm34LuzLtu4Ly045L3gqgz5Lv0nLtmpltu5tL01q4Lr03L5tqL53mygo503 L0rL
13 mzpLtu4Lom3qq3LtlqLx07qpL5tqL803wLtlqLpupLmzpl5tqLo0yy6zu5 L5tm5Lu5LomyqL8u5tLn65Lz08LtlqLu4Lz05L463qLurLtlqL8uxxLq7q3Lsq
14
15
16
```

Chọn giải thuật Permuation với mode 2

```

Run: encode
"C:\Users\Huỳnh Thị Kim Phượng\AppData\Local\Programs\Python\Python38-32\python.exe" E:/bkuOnline/MMANMang/bt1/encode.py
Moi chon giai thuat encrypt (1-Caesar,2-Permutation,3-Mix): 3
Moi nhap day hoan vi cho chuoi 8 ki tu (Vi dụ: 81726354): 87654321
Process finished with exit code 0

```

Với key là dãy hoán vị 87654321:

```

key.txt
1 Day hoan vi block 8 ki tu la: 87654321
2

```

Ta được bản mã cipher sau :

```

ciphertext.txt
1 haD nehWtraM leiird zenit yb sevd rac ehhsrelaeuoH ni pehw notsow eh er sa dekr dleif a ecivresicinhcoetan
2 ey 3 sihd dlo ra rethguaioP lliwti ot tnyas dna aD kool row syddk
3 zenitraM dah 82 rae neebone gninenom hgupus ot ysih tropna efiw rieht drethguad
4 saw dnaninnigebool ot ggpu ta kt gnidarlimaf ehllams syemtrapa a ot tnhouse
5 w eh tuB dial saA ni ffo sa lirpivanorocesac sur nageb sto soar
6 l eh dna sih tsoreyolpmeedivorp htlaeh dnarusni d ehT ec fo maeroh wen asinav emhed
7 ah I woNas ot evd I oN ykrow tnoa ereht h eromynr dias egnillaceser sih ot esnopuad sih ghter
8 suj ma Igniyr tlpxe ot h ot nia taht rewonk uoyl yddaD sih tso tub bojits ma Iikool llr stI gnough
9 zenitraM eno si erom fo 02 naht noillimnaciremAah ohw smoceb evlpmenu eoyed
10 esuacebc eht forivanoroednap sucihw cimof sah hnam decryolpme ys ot sreieht tuhr doors
11 htie dnaolruf ref ro hguieht eriyolpme rees
12 zenitraMs retfa sraey xiyyaN fo ecivres rht dna sraey ee sih ni cer tsom boj tnehgin dahf sepon t eht rorotcejarsih f
13 sih dnaH reerac devol ekrow eht did eh eht dnatinummoci taht yw emac t tub htii eh wonus ton seh fi erve lliw i teg
14
15
16

```

**Chọn giải thuật Mix kết hợp Caesar và Permutation với mode 3**

```

Run: encode x
"C:\Users\Huỳnh Thị Kim Phượng\AppData\Local\Programs\Python\Python38-32\python.exe" E:/bkuOnline/MMANMang/bt1/encode.py
Moi chon giai thuat encrypt (1-Caesar,2-Permutation,3-Mix): 1
Moi nhap do dich chuyen : 12
Moi nhap day hoan vi cho chuoi 8 ki tu (Vi du: 81726354): 12365487
Process finished with exit code 0

```

Với key1 là độ dịch chuyển 12 và key2 là dãy hoán vị 12365487

```

key.txt
1  Do dich chuyen Caesar cipher la: 12; Day hoan vi block 8 ki tu la: 12365487
2

```

Ta được bản mã cipher sau :

```

ciphertext.txt
1 itqPLzzmuqxMYL53uzqpLAu37q4 nL5LtqL3mopLqmx43qut1LuTLz604508Lzqt3qLLqt083wqmLpL4mLrxquLp4q3ou7Lq5gouztuomz
2 tu4LElq m3Lpx0pLm6sq5tL38ux1Lxu0z5LL055uLmz4Lp mX0Lw0mPpp 8L430w
3 Ym3zu5AqLDJmtLLpnqqgLz3mzvzqLs0z6st0yLqz L54L016103tL54uL8vLqzmpL5uqtL3pm65ts3q
4 mzp8LL4ngszuzusL5xL000wLm6L5s13mpszu5LtqLymrxu 4Lmy4xxLm153mgyz5LL05Lmt064q
5 N65qtL8Lm4LumxLp0rrzuLML13umLxL4o03mz0v7364moLq44LnmsqLz50L40m3
6 mzpqtLxL045utLL4qy1 0x3qL13u70qppLtxmq5Luz364zmoqLqtfpL3qm0LyLrmLzL8q0tyqLzm74utgp
7 Z08LUmt7qLL05m4 LZUL0pL0z508Lw3L5tq3qmlz yq30tLqL4pum3LqomuxszLtu3L44q10zLq405LtuPL46mst5q3
8 ULMvLy465L5u 3szL509qLx1muz05LtLq3Lmt5L5 06zwL80LPm ppxL045utLL4v0n6nLL5ULm4Lyv5xxL00xuwzsL45U3L06st
9 Ym3zu5AqLu4z0LLq0rL30yLq5tmDLzL8yux0uxLzMyqqu3zm4L8L0tmt7qLqny0qL6yqzx10 qp
10 nqo46mLq0rLqt50L0307mz3u64Lzm1gpyuot8LoutLtL4m0r3oqyLpzm Lqx1y 0q3405L4Lt65t5Luq3Lp0034
11 mzpugt5q3L36r0x6st30LrLu3qt5Luq3Lqx1y 0qq4
12 Ym3zu5AqLmr3q54Lu9Lmq 43L0rmZL 7L4qu73qoLmz5Lp3tqqLmq 43LuzutLL4y043L5oqqz50vLLntmpuLtLtsLt04q1rL03Lqt55L3mv5oq30 L0tL
13 mzpugtLL4om33qqTLqLxq70Lp5tq08Lw3LtuqLpLpmzpt5LLq0yz6y5u L55mtuL5Logym8Lu5t6nLL5z08qtLuL4LzL50643qLLruqtL8uLxx7qq3L5q
14
15
16

```

### ❖ Kết quả của quá trình decryption sau khi run file encode.py

Ta được danh sách key ở file hack\_key.txt như sau :



```

hack_key.txt
1  Danh sách key Caesar key có thể:
2  4
3  12
4

```

Output plaintext được giải mã từ cipher :

```

hack_plaintext.txt
1  2C Ifj6ID Gfs6MODI Uf9MDQ Nf7TfOC f86Mf9 6G MNCDFDIfnJPN0JIfRC M fC fRJMF 9f6Nf6fAD G9fN MQD8 fO 8CID8D6I
2  CNDfYfT 6MfJG9f96PBCQ MfRDGGfKJDIOf0JfD0f6I9fN6TfrJJFfj699TNfRJMF
3  s6MODI UfxdfC69f7 If 6MIDIBf IJPBCfHJI Tf0JfNPKKJMOfCDNfRDA f6I9fOC DMf96PBCO M
4  6I9fR6Nf7 BDII0IBf0JfGJJf60fPKBM69DIBfOC fA6HDGfNfNH6GGf6K6MDH IOf0Jf6fCJPN
5  hPOfC fR6NfG6D9fJAAfDfGfKMDGf6Nf8JMJi6QDMPNf86N Nf7 B6If0JfNJ6M
6  6I9fC fGJNOfCDNf HKGJT MfKMJQD9 9fC 6GOCfDfINPM6I8 fZC f9M 6HfJAf6fI RfCJH fQ6IDNC 9
7  tJRf0fC6Q f0JfN6TftJf0f9JIOfRJMFfOC M f6ITHJM fC fN6D9fM 86GDIbfcDNfM NKJIN f0JfCDNf96PBCO M
8  of6HfEPNOfQMTDIBf0Jf SKG6DI0JfC MfOC60fTJPFfIJRfj699fGJNOfCDNfEJ7f7P0f0f6HfNODGGf6JJfDIBf0fNfMJPBC
9  s6MODI UfDNfJI fJAfHJM fOC6IfXVfHDGGDIIfGH MD86INfRCJfC6Q f7 8JH fPI HKGJT 9
10 7 86PN fJAfOC f8JMJi6QDMPNfK6I9 HD8fRCD8CfC6NfAJM8 9fH6ITf HKGJT MNf0JfNCP0fOC DMf9JJMN
11 6I9f DOC MfAPM6JPBCfJHfADM fOC DMf HKGJT N
12 s6MODI Uf6AO MfNDSfT 6MnfJAft6QTfN MQD8 f6I9fOCM fT 6MnfDfCDNfHJNOfM 8 IOfEJ7fC69fCDBCfCJk NfAJMfOC fOM6E 8OJMTfJA
13 6I9fCDNf86M Mfn f6JQ 9fOC fRJMFfC f9D9f6I9fOC f8JHHPIDQfTfOC60fD0f86H fR0Cf7P0fIJRfC fDNfIJOfNPM fDAfC fRDGGf Q MfB
14
15
16 When Daniel Martinez drives by the car dealership in Houston where he worked as a field service technician
17 his 3 year old daughter will point to it and say Look Daddys work
18 Martinez 28 had been earning enough money to support his wife and their daughter
19 and was beginning to look at upgrading the familys small apartment to a house
20 But he was laid off in April as coronavirus cases began to soar
21 and he lost his employer provided health insurance The dream of a new home vanished
22 Now I have to say No I dont work there anymore he said recalling his response to his daughter
23 I am just trying to explain to her that you know Daddy lost his job but I am still looking Its rough
24 Martinez is one of more than 20 million Americans who have become unemployed
25 because of the coronavirus pandemic which has forced many employers to shut their doors
26 and either furlough or fire their employees
27 Martinez after six years of Navy service and three years in his most recent job had high hopes for the trajectory of

```

## **Chương 4: Đánh giá hệ thống và kết luận**

Chương trình đã đáp ứng tất cả các yêu cầu requirements đã đặt ra gồm mã hóa file đầu vào plaintext chứa ít nhất 1000 ký tự bằng 3 giải thuật Caesar, Permutation và giải thuật kết hợp giữa Caesar và Permutation và giải mã bằng phương pháp phân tích mã dựa vào đặc tính ngôn ngữ.

Chương trình có thể mở rộng trên những ngôn ngữ khác vì giải thuật phân tích mã là dựa vào phân tích đặc tính của ngôn ngữ kết hợp tra từ điển của danh sách những từ thông dụng ở file `common_words_list.txt` để kết hợp từ tìm ra bản rõ ban đầu.

Tuy nhiên chương trình cũng có những hạn chế và giới hạn phạm vi mã hóa như sau:

- Giải thuật dùng để mã hóa như Caesar hay Permutation khá đơn giản có thể dễ dàng giải mã bằng phương pháp tấn công vét cạn (brute-force).
- Phạm vi bản rõ sử dụng giới hạn ở ngôn ngữ tiếng anh dễ dàng nhận biết và dựa vào đặc điểm, đặc tính ngôn ngữ khá dễ dàng để phân tích mã.
- Thời gian thực thi của chương trình sẽ lâu hơn nếu file đầu vào lớn .
- Để tăng tính chính xác ta có thể tăng số lượng danh sách từ phổ biến nhưng bù lại thời gian giả mã sẽ tăng do tốn thêm thời gian quét các từ mới này. Do đó ta phải chọn giữa tính chính xác và thời gian để ưu tiên theo nhu cầu.