|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  \_\_\_\_\_\_\_\_\_\_\*\_\_\_\_\_\_\_\_\_  D:\IT\20161\Project I\Báo cáo\anh.png  BÁO CÁO BÀI TẬP LỚN  MÔN: AN TOÀN VÀ BẢO MẬT THÔNG TIN  Đề tài: Giải mã hệ mã mật Vigenere  Giảng viên hướng dẫn: Đỗ Văn Uy  Nhóm sinh viên thực hiện:   |  |  |  | | --- | --- | --- | | STT | Họ và tên | MSSV | | 1 | Chu Xuân Vĩnh | 20145278 | | 2 | Khúc Trọng Ngọc | 20143205 | | 3 | Hà Văn Quang | 20143578 |       Hà Nội, 10/4/2017 |

**MỤC LỤC**

[**LỜI MỞ ĐẦU** 3](#_Toc481878716)

[**PHẦN I. GIỚI THIỆU CHUNG VỀ HỆ MÃ MẬT VIGENERE** 4](#_Toc481878717)

[**1** **Lịch sử hệ mã mật Vigenere** 4](#_Toc481878718)

[**2** **Mã hóa và giải mã trong hệ mã Vigenere** 5](#_Toc481878719)

[**2.1** **Mã hóa bản rõ** 6](#_Toc481878720)

[**2.2** **Giải mã bản mã** 8](#_Toc481878721)

[**3** **Tính chất đại số của mật mã Vigenere** 9](#_Toc481878722)

[**PHẦN II. ĐÁNH GIÁ HỆ MÃ MẬT VIGENERE** 10](#_Toc481878723)

[**PHẦN III: KĨ THUẬT PHÁ MÃ HỆ MÃ VIGENERE** 11](#_Toc481878724)

[**1** **Ước lượng độ dài của khóa** 11](#_Toc481878725)

[**1.1** **Phương pháp Kasiski** 11](#_Toc481878726)

[**1.2** **Phương pháp sử dụng chỉ số trùng hợp (Index of Coincidence - IC)** 14](#_Toc481878727)

[**2** **Khôi phục khóa** 17](#_Toc481878728)

[**2.1** **Phương pháp χ2** 17](#_Toc481878729)

[**2.2** **Phân tích tần số** 25](#_Toc481878730)

[**PHẦN IV. KẾT LUẬN CHUNG** 30](#_Toc481878731)

[**TÀI LIỆU THAM KHẢO** 31](#_Toc481878732)

**LỜI MỞ ĐẦU**

Trong lịch sử phát triển của nhân loại, từ khi con người xuất hiện nhu cầu trao đổi thông tin với nhau thì nhu cầu giữ bí mật và đảm bảo an toàn của những thông tin đó cũng xuất hiện theo. Sự cần thiết của việc tìm ra một phương pháp đơn giản, hiệu quả để đảm bảo cho thông điệp gửi đi chính là nguyên nhân của sự hình thành mật mã. Và cuộc chiến giữa một bên luôn muốn che giấu thông tin của mình và một bên luôn muốn đọc được những thông tin đó đã thúc đẩy ngành mật mã ngày càng phát triển. Ngành mật mã học phát triển qua nhiều giai đoạn từ các hệ mã cổ điển như: mật mã một bảng thế, mật mã đa bảng thế,… đến các hệ mã hiện đại như các hệ mã khối, hê mã khóa công khai,... Mong muốn hiểu biết hơn về các hệ mã cổ điển và các điểm yếu của hệ mã này nên chúng em xin chọn đề tài “Giải mã hệ mã mật Vigener”.

Bài báo cáo được chia thành bốn phần:

**Phần 1: Giới thiệu chung về hệ mã mật Vigener**

**Phần 2: Kĩ thuật phá mã hệ mã mật Vigenere**

**Phần 3: Đánh giá hệ mã Vigenere**

**Phần 4: Kết luận chung**

Do kiến thức còn hạn hẹp nên bài báo cáo và phần trình bày khong tránh khỏi những sai sót, hạn chế. Em mong thầy cô góp ý để chúng em làm tốt hơn trong những bài tập kế tiếp.

Chúng em xin chân thành cảm ơn thầy Đỗ Văn Uy đã hướng dẫn và giúp đỡ để chúng em hoàn thành đề tài này!

**PHẦN I. GIỚI THIỆU CHUNG VỀ HỆ MÃ MẬT VIGENERE**

1. **Lịch sử hệ mã mật Vigenere**

Hệ mã mật Vigenere lần đầu tiên xuất hiện vào năm 1585 trong cuốn sách ***Traicté des Chiffres*** (***A Treatise on Secret Writing***) được viết bởi Blaise de Vigenère.



*Hình 1. Blaise de Vigenère (April 5, 1523 - February 19, 1596)*

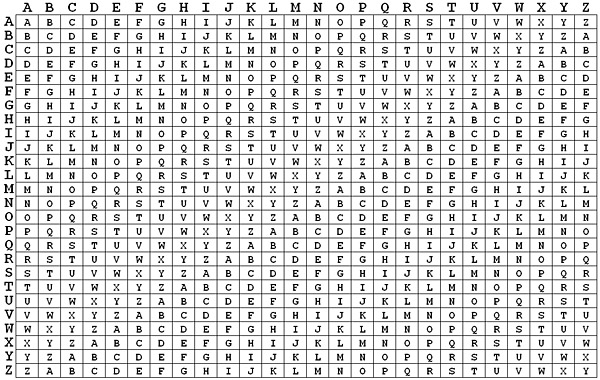
Tuy nhiên, năm 1553 Giovan Batista Belaso cũng đã nói về kĩ thuật tương tự trong cuốn sách nhỏ của ông ***La cifra del. Sig. Giovan Batista Belaso*** [KAHN1967, page 137]. Trong cuốn Singh [SINGH1999, pp. 45--51, Chapter 2] có cuộc thảo luận ngắn và thú vị về Vigenere, cùng với đó Kahn [KAHN1967, page 137, Chapter 4] có bài trình bày chi tiết hơn. Dưới đây là một và trích dẫn nói về Vigenere:

* Vigenère đã trở nên quen thuộc với những bài viết của Alberti, Trithemius và Porta khi ông 26 tuổi, ông được đưa đến Rome trong một nhiệm kỳ ngoại giao kéo dài hai năm. Vigenère quan tâm tới mật mã hoàn toàn mang tính thực tiễn và liên quan tới công việc ngoại giao của ông. Sau đó, ở tuổi ba mươi chín, Vigenère quyết định rằng ông đã tích lũy đủ tiền cho ông để có thể từ bỏ sự nghiệp của mình và tập trung vào nghiên cứu mã mật. Khi đó, ông kiểm tra chi tiết những ý tưởng của Alberti, Trithemius, và Porta, kết hợp chúng thành một mật mã mới mạch lạc và mạnh mẽ [SINGH1999, trang 46].
* …
* Mặc dù Alberti, Trithemius và Porta đều có những đóng góp quan trọng, nhưng mật mã được gọi là mật mã Vigenère để tôn vinh người đã phát triển nó thành hình thức cuối cùng. Sức mạnh của mật mã Vigenère nằm ở việc sử dụng không phải là một, nhưng có 26 bảng chữ cái mật mã để mã hóa một tin nhắn [SINGH1999, 48].
* …
* Để giải mã thông điệp, người nhận cần biết hàng nào của bảng mã Vigenère đã được sử dụng để mã hoá từng chữ cái, vì vậy phải có một hệ thống chuyển đổi giữa các hàng. Điều này đạt được bằng cách sử dụng từ khóa [SINGH1999, trang 49].
* …

Trong những phần tiếp theo của báo cáo sẽ trình bày về một số vấn đề cơ bản của hệ mã Vigenere như: mã hóa và giải mã; đánh giá hệ mã; phá mã Vigenere; …

1. **Mã hóa và giải mã trong hệ mã Vigenere**

Hệ mã Vigenere sử dụng bảng kích thước 26x26 với tên cột và tên hàng là các kí hiệu từ A đến Z. Hàng đầu tiên trong bảng ta điền 26 chữ cái tiếng anh từ A đến Z. Bắt đầu từ hàng thứ hai, mỗi hàng được hình thành bởi cách quay trái hàng trên nó một vị trí. Ví dụ, khi B được dịch chuyển thành kí tự đầu tiên của hàng thứ hai, thì kí tự A được quay về cuối hàng. Hình 2 dưới đây là bảng mã Vigenere:



*Hình 2. Bảng mã Vigenere*

Cùng với bản rõ cần mã hóa, hệ mã Vigenere cần thêm từ khóa, từ khóa được lặp lại với tổng độ dài bằng tổng độ dài bản rõ. Ví dụ, giả sử bản rõ là **MICHIGAN TECHNOLOGICAL UNIVERSITY** và từ khóa là **HOUGHTON**. Khi đó, từ khóa phải lặp lại như sau:

**MICHIGAN TECHNOLOGICAL UNIVERSITY**

**HOUGHTON HOUGHTONHOUGH TONHOUGNTO**

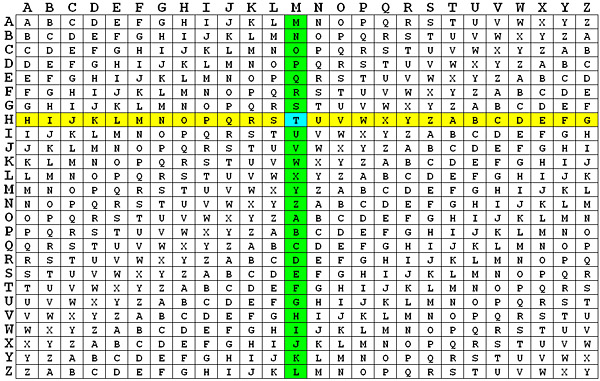
Chúng ta loại bỏ các dấu cách và dấu chấm câu, chuyển đổi các chữ cái sang chữ hoa và chia kết quả thành khối năm chữ cái. Ta được như sau:

**MICHI GANTE CHNOL OGICA LUNIV ERSIT Y**

**HOUGH TONHO UGHTO NHOUG HTONH OUGHT O**

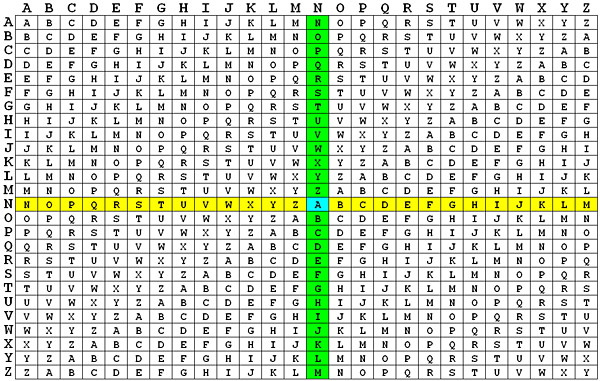
* 1. **Mã hóa bản rõ**

Để mã hóa bản rõ, chọn một kí tự trong bản rõ và một kí tự trong từ khóa tương ứng với nó. Sử dụng chữ chữ cái ở bản rõ làm chỉ số cột, chữ cái ở bản mã làm chỉ số hàng tương ứng, và chữ cái xuất hiện ở giao hàng với giao cột là chữ cái xuất hiện trong bản mã. Ví dụ, kí tự đầu tiên trong bản rõ là **M** và kí tự trong từ khóa tương ứng là **H**. Nghĩa là cột M, hàng H được sử dụng để tạo mã. Giao giữa cột M và hàng H là chữ cái T, do vậy T là kết quả của việc mã hóa. Hình 3 dưới đây minh hoa cho quá trình trên:



*Hình 3. Mã hóa kí tự M với từ khóa là H*

Tương tự, kí tự **N** trong **MICHIGAN** tương ứng với kí tự **N** trong từ khóa (**HOUGHTON**), giao giữa cột N với hàng N là kí tự A chính là kí tự trong bản mã. Hình 4 minh họa cho mã hóa N bởi từ khóa N:



*Hình 4. Mã hóa kí tự N bởi từ khóa N*

Lặp lại quá trình trên cho tới khi tất cả các kí tự trong bản rõ được mã hóa, bản mã thu được là **TWWNPZOA ASWNUHZBNWWGS NBVCSLYPMM.** Dưới đây là bản mã, từ khóa và bản rõ được căn thẳng với nhau.

**MICHI GANTE CHNOL OGICA LUNIV ERSIT Y**

**HOUGH TONHO UGHTO NHOUG HTONH OUGHT O**

**TWWNP ZOAAS WNUHZ BNWWG SNBVC SLYPM M**

* 1. **Giải mã bản mã**

Để giải mã, chọn kí tự trong bản mã và kí tự tương ứng trong từ khóa. Dùng kí tự trong từ khóa để tìm hàng, kí tự trong bản mã là giao giữa hàng của từ khóa và cột của kí tự bản rõ. Ví dụ, để giải mã kí tự **T** trong bản mã, ta tìm kí tự tương ứng trong từ khóa là kí tự **H.** Trong hàng H tìm kí tự T và cột có chứa T cho ta kí hiệu của bản rõ là **M.** Hình 3 ở trên cho ta thấy quá trình này. Xét kí tự thứ năm **P** trong bản mã. Kí tự tương ứng trong từ khóa là **H** và hàng **H** được dùng để tìm kí tự **P** trong hàng. Từ kí tự **P** tìm tiêu đề cột tương ứng ta được cột **I**, kí tự bản rõ tương ứng là **I**

1. **Tính chất đại số của mật mã Vigenere**

Ở phần trên, ta đã dùng bảng mã Vigenere với kích thước 26x26 để mã hóa và giải mã hệ mã mật Vigenere. Hàng bên dưới được hình thành bằng cách quay trái hàng bên trên một vị trí. Thay vì dùng bảng mã để mã hóa và giải mã ta có thể thực hiện như sau.

Gán mỗi kí tự A, B, C, …, Z ứng với một giá trị 0, 1, 2, …, 25, khi đó giá trị tương ứng với mỗi kí tự chính là khoảng cách từ kí tự đó đến kí tự A. Kí tự **P** (plaintext ) của bản rõ được mã hóa thành kí tự **C** (Ciphertext) với từ khóa có giá trị tương ứng là **d** như sau:

***C****=(****P****+****d****) mod 26*

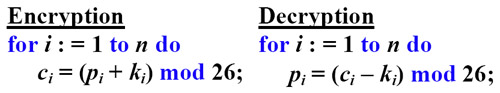
Tổng quát, cho bản rõ là p1p2…pn, từ khóa là k1k2…kn và bản mã thu được là c1c2…cn ta có:

*ci= (pi+ki) mod 26*

Để giải mã, ta có phép đảo ngược như sau:

*pi = (ci - ki) mod 26*

Từ tính chất đại số trên, dễ dàng lập trình để mã hóa (Encryption) và giải mã (Decryption) như sau:



**PHẦN II. ĐÁNH GIÁ HỆ MÃ MẬT VIGENERE**

**PHẦN III: KĨ THUẬT PHÁ MÃ HỆ MÃ VIGENERE**

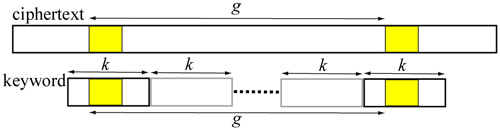
1. **Ước lượng độ dài của khóa**
   1. **Phương pháp Kasiski**

Friedrich W. Kasiski là một sĩ quan quân đội Đức, ông phát hành cuốn sách Cryptography and the Art of Decryption vào năm 1863. Cuốn sách dài hơn 100 trang này là cuốn sách đầu tiên đề cập đến việc phá mã Vigenere.

Ý tưởng của phương pháp là tìm các đoạn mã lặp lại trong bản mã và khoảng cách giữa chúng. Kasiski có các quan sát sau:

* Nếu một xâu con trong bản rõ được mã hóa bởi cùng một xâu con trong khóa thì trong bản mã sẽ chứa lặp lại các xâu con được mã hóa tương ứng với khoảng cách giữa chúng là bội của độ dài khóa
* Tất nhiên không phải mọi xâu con trong bản mã đều lặp lại theo cách này, có trường hợp hai xâu con khác nhau trong bản rõ sau khi mã hóa có thể tạo ra xâu con giống nhau trong bản mã, tuy nhiên xác suất xảy ra trường hợp này là rất nhỏ.

Trong hình minh họa dưới đây, khoảng các giữa hai xâu con trùng nhau trong bản mã là g, k là độ dài khóa. Nếu hai xâu con trùng nhau trong bản mã được sinh ra bởi việc mã hóa một xâu con trong bản rõ bởi cùng một xâu con trong khóa thì một trong các ước chung của g là chiều dài của khóa.



Nhận xét: một bản mã dài sẽ xuất hiện nhiều hơn những xâu con lặp lại, ngược lại một bản mã ngắn được mã hóa với một khóa dài sẽ khó xuất hiện những xâu con lặp lại trong bản mã.

Ví dụ: Một đoạn văn bản bằng tiếng anh:

TODAY I AM GOING TO GIVE YOU SOME ADVICE ABOUT RECORD KEEPING KEEPING ACCURATE RECORDS IS VERY IMPORTANT BECAUSE IF A PATIENT IS UNHAPPY ABOUT HIS TREATMENT HE MAY WANT TO MAKE A COMPLAINT TO THE SURGEON THESE RECORDS WILL HELP US TO MAKE A CASE ALWAYS WRITE YOUR NAME THE DATE AND THE TIME ON ALL THE RECORDS DO NOT USE PENCIL OR COLORED PENS USE BLACK INK IF YOU MAKE A MISTAKE JUST CROSS IT OUT WITH A SINGLE LINE DO NOT USE CORRECTION FLUID DO NOT TRY TO FILE THE RECORDS AWAY YOURSELF OR THEY MAY GET LOST PUT THEM IN THIS BOX HERE THE ADMINISTRATION MANAGER WILL MAKE SURE THEY ARE FILED AWAY CORRECTLY

Ta sẽ mã hóa đoạn văn bản trên với khóa “MUST” có độ dài là 4. Ta có bản mã:

FIVTK CSFSI AGSNG ZUPWR AOKHY YSWHC UXMVG NFLWV ALVDQ YHBZA CXQJA GSUUV GLSMQ LWVAL VLUMN XDSAF BIJMM HLUQW SNEYA YMJSM UYFMU MMGTU HIKUT HGNZB ENJXM NEXZN ZXYUQ PMHLM AGSDQ UUHYJ DTUHL MANZX EOJZQ IFMTY KXDYU HDXKP UFDAQ FHNEN GFMEW TOUKX MFOTK MOKUN WRAOJ GMGWM TYVTF YSGPN ZXFCE XAHSE XNZXD YUHDX KWAHG MGMWI QHUBX IJVAF GKQXH XZMML QVDTO EAGWC XRAOE TWYSF UMLTW YBNEN UKAMK BFIMM ICLAM MAGSF WEUHW WAHGM GMWVA LJXON AHZZD NUXVH ZILMD SLHRC DXFBW KQWGK PMSPM SQHGL KXXZG KFBWR YUQZQ NDHEN HNFNZ XYCFM TCKUA RZXDY LAQUV FUHAL FLSMU IFFMH SZQLO BXFET WYKND YLAQS SKQZA EQXSP MSUHD LWVFF Q

Phân tích đoạn mã trên được:

Xâu con lặp lại có độ dài 8

XDYUHDXK:

171--243 (72)

WAHGMGMW:

251--335 (84)

Xâu con lặp lại có độ dài 6

LWVALV:

37--65 (28)

Xâu con lặp lại có độ dài 5

DYLAQ:

428--464 (36)

Xâu con lặp lại có độ dài 4

WRAO:

18--210 (192)

WVAL:

38--342 (304)

NZXY:

129--413 (284)

HLMA:

137--153 (16)

MAGS:

139--325 (186)

ZXDY:

242--426 (184)

ETWY:

294--458 (164)

SPMS:

382--478 (96)

Nhận xét: Ta thấy khoảng các giữa các xâu con lặp lại: 72, 84, 28,… đều có ước chung là 2, 4, 8.

* 1. **Phương pháp sử dụng chỉ số trùng hợp (Index of Coincidence - IC)**

1. **Tính IC**

Cho một văn bản, chỉ số trùng hợp IC là xác suất ta chọn hai lần được cùng một ký từ cùng văn bản. Chỉ số trùng hợp được sử dụng đầu tiên bởi William F. Friedman vào năm 1992.

Độ dài văn bản N (có N ký tự) và kích thước bảng chữ cái là n, ai là ký tự thứ i trong bảng chữ cái. Số lần ai xuất hiên trong văn bản là Fi lần.

Xác xuất để hai lần chọn ngẫu nhiên đều được ai là:

http://www.cs.mtu.edu/~shene/NSF-4/Tutorial/VIG/EQN-IOC-1.jpg

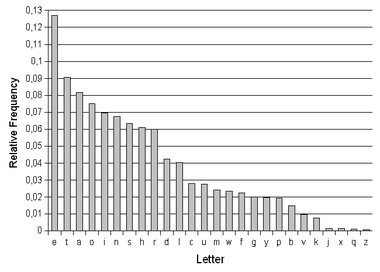
và xác xuất để hai lần chọn ngẫu nhiên được cùng ký tự là:

http://www.cs.mtu.edu/~shene/NSF-4/Tutorial/VIG/EQN-IOC-2.jpg

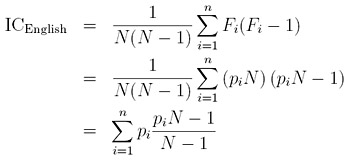
Vậy chỉ số trùng hợp là:

http://www.cs.mtu.edu/~shene/NSF-4/Tutorial/VIG/EQN-IOC-3.jpg

Xét bảng chữ cái tiếng anh (26 ký tự):



Chỉ số trùng hợp của bảng chữ cái tiếng anh là:



Nếu N đủ lớn thì (piN – 1)/(N – 1) sẽ xấp xỉ bằng pi, do đó:

http://www.cs.mtu.edu/~shene/NSF-4/Tutorial/VIG/EQN-IOC-5.jpg

Từ đồ thị thể hiện tần số xuất hiện của các ký tự trong bảng chữ cái tiếng anh ta tính được ICEnglish 0.0686.

Trong trường hợp văn bản là những xâu được sinh ngẫu nhiên thì tần suất xuất hiện của một ký tứ trong văn bản là pi = 1/n, n là kích thước bảng chữ cái. Từ đó tính được ICRandom 1/n. Nếu sinh ngẫu nhiên từ bảng chữ cái tiếng anh thì ICRandom = 1/26 0.038466.

1. **Sử dụng IC để ước lượng độ dài khóa**

Chỉ số trùng hợp sử dụng để ước lượng độ dài khóa. Ban đầu đoán độ dài khóa là *l* và chia C = C1C2…CN thành *l* đoạn:

S1 bắt đầu từ C1 và có cấu trúc: S1 = C1C1+*l*C1+2*l*…

S2 bắt đầu từ C2 và có cấu trúc: S1 = C2C2+*l*C2+2*l*…

Tổng quát, Si bắt đầu bằng Ci và có cấu trúc: Si = CiCi+*l*Ci+2*l*…và, S*l* bắt đầu bằng C*l* và có cấu trúc: S*l* = *ClC2lC3l*…

Với *l* là độ dài của khóa thì mỗi khối Si sẽ được thế bằng một bảng thế (hay mã hóa bằng một ký tự trong khóa)

Ví dụ: Ta có chuỗi mã:

**RSTCS JLSLR SLFEL GWLFI ISIKR MGL**

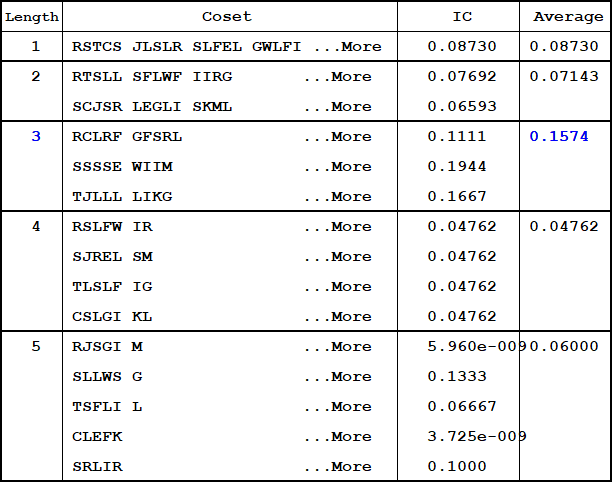
**RST CSJ LSL RSL FEL GWL FII SIK RMG L**

Nếu độ dài khóa là 3 thì ta sẽ chia đoạn mã trên thành 3 khối:

**RCLRFGFSRL  
SSSSEWIIM  
TJLLLLIKG**

Giả sử bản rõ được viết bằng tiếng anh và độ dài khóa k mà ta ước lượng là chính xác thì IC của các khối Si sẽ tiến gần đến ICEnglish theo độ dài của Si. Nếu bản rõ được sinh ngẫu nhiên từ bảng chữ cái tiếng anh thì ICSi sẽ tiến gần đến ICRandom. Dựa vào quan sát trên ta sẽ chia bản mã lần lươt thành 1 khối, 2 khối, 3 khối,…tương ứng với *l* = 1, 2, 3,… Tính IC trung bình của các khối và so với ICEnglish, nếu cách chia nào cho IC gần với ICEnglish nhất thì *l* tương ứng chính là độ dài khóa.

Xét lại ví dụ trên:



Thấy khi ước lượng *l* = 3 ta thu được IC lớn nhất, do đó xác suất bản rõ được mã hóa với khóa có độ dài 3 là lớn nhất.

1. **Khôi phục khóa**
   1. **Phương pháp χ2**

Nếu chiều dài của từ khoá ước lượng là chính xác, thì mỗi khối được xây dựng sẽ được mã hóa bằng cùng một chữ cái

Ví dụ dưới đây là việc mã hóa với từ khóa là **BOY** và bản rõ là:

**MICHIGAN TECHNOLOGICAL UNIVERSITY**

để mã hóa ta lặp lại từ khóa **BOY** sao cho độ dài bằng với bản mã như sau:

**MICHIGAN TECHNOLOGICAL UNIVERSITY  
BOYBOYBO YBOYBOYBOYBOY BOYBOYBOYB**

Sau khi mã hóa ta được bản mã:

**NWAIWEBB RFQFOCJPUGDOJ VBGWSPTWRZ**

Do từ khóa có độ dài là 3 nên ta sẽ chia bản mã thành 3 khối **N I B F O P D V W T Z  
W W B Q C U O B S W**  **A E R F J G J G P R**

Nếu ta coi chữ A đứng ở vị trí thứ 0 thì ta có bảng vị trí của 26 chữ cái trong tiếng anh như sau

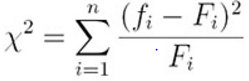
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G | H | I | J |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| K | L | M | N | O | P | Q | R | S | T |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| U | V | W | X | Y | Z |  |  |  |  |
| 20 | 21 | 22 | 23 | 24 | 25 |  |  |  |  |

Kể từ chữ cái đầu tiên trong từ khóa là B, các chữ trong bản rõ tương ứng với B được dịch sang bên phải một vị trí sao cho M trở nên N và A trở nên B. Kể từ khi chữ cái thứ hai trong từ khóa là O, các chữ trong bản rõ tương ứng với O được dịch sang phải 14 vị trí sao cho I trở thành W, O trở thành B, và vân vân. Tương tự, khối thứ ba cũng thu được bằng cách dịch chuyển các chữ trong bản rõ tương ứng với Y sang phải 24.

Trong trường hợp chỉ biết ba khối, chúng ta cần phải dich chuyển mỗi trong số chúng sang bên trái một số vị trí để có được bản rõ trở lại. Chính xác hơn, mỗi khối được chuyển sang bên trái 1 vị trí , 2 vị trí, ..., và 25 vị trí. Lưu ý rằng chúng ta không cần phải thay đổi 0 vị trí bởi vì nó là khối riêng của mình. Do mỗi lần dịch chuyển sẽ tạo ra một giải mã có thể có của các khối, nên có 26 khả năng khác nhau. Nếu chúng ta có k khối, tổng số kết hợp sự dịch chuyển là 26k, mà có thể rất lớn thậm chí nếu k là nhỏ. Ví dụ, nếu độ dài từ khóa có thể là 8, có 268 = 208.827.064.576 kết hợp sự dịch chuyển có thể (hoặc từ khóa có thể). Rất khó để xác định sự dịch chuyển của một khối có thể sinh ra từ khoá chính xác. Do đó, chúng ta cần một phương pháp tốt hơn chứ không phải là sử dụng tấn công vét cạn

Có một phương pháp đơn giản dựa trên tần số của các chữ cái bằng tiếng Anh. Vì mỗi khối được mã hóa bởi cùng một chữ cái, tần số của nó không giống như một đặc trưng văn bản tiếng anh. Dịch chuyển một khối sẽ làm thay đổi tần số của nó. Trong số 26 dịch chuyển có thể, một trong số đó có thể mang lại bản rõ ban đầu, cái mà có tần số rất gần với tần số của tiếng Anh. Do đó, chúng ta có thể so sánh tần số của mỗi lần dịch chuyển với tần số của tiếng Anh, và sự dịch chuyển nào đó tạo ra một tần số gần với tần số bằng tiếng Anh có khả năng là sự dịch chuyển chính xác. Trong thống kê có nhiều phương pháp để đo lường sự phù hợp,một trong số đó là phương pháp **χ2**.

Cho một tập hợp các giá trị quan sát được f1, f2,..., fn và một tập giá trị đã biết (mong đợi) tương ứng ​​F1, F2,..., Fn, khi đó **χ2** được tính như sau:



Trong công thức này, F là các giá trị trong bảng tần số tiếng Anh, cái mà đã biết, và f là tần số thu được từ một sự dịch chuyển. Sự dịch chuyển của một khối tạo ra giá trị **χ2** nhỏ nhất là một trong số các tần số gần với tần số của tiếng Anh.

Xét tập khối thứ 2 là **WWBQCUOBSW** được trình bày ở trên, ta tính được số lượng tần số fi, Fi và giá trị **χ2** như sau:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Chữ cái** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** |
| **Count** | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **fi** | 0 | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Fi** | 0.82 | 0.14 | 0.028 | 0.038 | 0.131 | 0.029 | 0.020 | 0.053 | 0.064 | 0.001 | 0.004 | 0.034 | 0.025 |
| **Chữ cái** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** | **U** | **V** | **W** | **X** | **Y** | **Z** |
| **Count** | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 3 | 0 | 0 | 0 |
| **fi** | 0 | 0.1 | 0 | 0.1 | 0 | 0.1 | 0 | 0.1 | 0 | 0.3 | 0 | 0 | 0 |
| **Fi** | 0.071 | 0.080 | 0.020 | 0.001 | 0.068 | 0.061 | 0.105 | 0.025 | 0.009 | 0.015 | 0.002 | 0.020 | 0.001 |
| **χ2** | **17.083** | | | | | | | | | | | | |

Giả sử tập khối **WWBQCUOBSW** dịch chuyển sang trái một vị trí khi đó ta có tập **VVAPBTNARV**, tương tự như trên ta cũng tính được số lượng tần số fi, Fi và giá trị **χ2** như sau:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Chữ cái** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** |
| **Count** | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **fi** | 0.2 | 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Fi** | 0.082 | 0.014 | 0.028 | 0.038 | 0.131 | 0.029 | 0.020 | 0.053 | 0.064 | 0.001 | 0.004 | 0.034 | 0.025 |
| **Chữ cái** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** | **U** | **V** | **W** | **X** | **Y** | **Z** |
| **Count** | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 |
| **fi** | 0.1 | 0 | 0.1 | 0 | 0.1 | 0 | 0.1 | 0 | 0.3 | 0 | 0 | 0 | 0 |
| **Fi** | 0.071 | 0.080 | 0.020 | 0.001 | 0.068 | 0.061 | 0.105 | 0.025 | 0.009 | 0.015 | 0.002 | 0.020 | 0.001 |
| **χ2** | **10.8557** | | | | | | | | | | | | |

Bảng dưới đây hiển thị ra 26 giá trị **χ2** của mỗi khối với giá trị nhỏ nhất được in đậm. Giá trị χ2 nhỏ nhất của khối 1 là **1.9532** tương ứng với chữ cái B (cụ thể là dịch chuyển sang bên trái một vị trí).Giá trị **χ2** nhỏ nhất của khối 2 là **2.1695** tương ứng với chữ cái O (cụ thể là dịch chuyển sang bên trái 14 vị trí). Giá trị χ2 nhỏ nhất của khối 3 là **2.3933** tương ứng với chữ cái Y (cụ thể dịch chuyển sang bên trái 24 vị trí). Nói một cách khác, các khối thứ nhất, thứ hai và thứ ba được mã hóa lần lượt bằng cái chữ cái **B**, **O** và **Y** tương ứng.Trong ví dụ này, chúng ta đã may mắn và tìm thấy chính xác từ khóa **BOY**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dịch chuyển**  **vị trí** | **Chữ cái tương ứng** | **Giá trị χ2**  **của khối 1** | **Giá trị χ2**  **của khối 2** | **Giá trị χ2**  **của khối 3** |
| 0 | **A** | 12.6808 | 17.0130 | 33.4114 |
| 1 | **B** | **1.9532** | 10.8557 | 47.2982 |
| 2 | **C** | 16.6228 | 61.7972 | 3.3558 |
| 3 | **D** | 10.2763 | 15.4671 | 9.8983 |
| 4 | **E** | 24.9700 | 35.7427 | 4.4140 |
| 5 | **F** | 16.1760 | 17.4307 | 19.7483 |
| 6 | **G** | 29.5341 | 82.8543 | 22.8300 |
| 7 | **H** | 2.5481 | 14.5767 | 66.6135 |
| 8 | **I** | 6.3800 | 4.3482 | 41.4530 |
| 9 | **J** | 20.3966 | 9.5387 | 26.0354 |
| 10 | **K** | 8.4236 | 6.4101 | 62.5271 |
| 11 | **L** | 14.2454 | 43.6223 | 8.4614 |
| 12 | **M** | 9.8439 | 31.8807 | 26.1736 |
| 13 | **N** | 15.2270 | 70.7267 | 3.6981 |
| 14 | **O** | 11.8107 | **2.1695** | 14.6269 |
| 15 | **P** | 19.8472 | 16.2274 | 11.9170 |
| 16 | **Q** | 22.7962 | 6.1626 | 51.0042 |
| 17 | **R** | 8.1086 | 31.2416 | 10.6299 |
| 18 | **S** | 19.9131 | 34.4761 | 56.6600 |
| 19 | **T** | 4.6458 | 29.8624 | 36.4451 |
| 20 | **U** | 18.6617 | 4.8624 | 36.3898 |
| 21 | **V** | 3.3357 | 27.0150 | 14.0996 |
| 22 | **W** | 21.9697 | 3.7015 | 21.4566 |
| 23 | **X** | 18.5799 | 118.3588 | 33.7453 |
| 24 | **Y** | 19.8023 | 14.2303 | **2.3933** |
| 25 | **Z** | 19.8783 | 55.4882 | 19.8128 |

Tuy nhiên, sự dịch chuyển tương ứng với giá trị **χ2** nhỏ nhất có thể không phải là sự lựa chọn đúng. Nói chung, chúng ta cần phải kiểm tra một số dịch chuyển tương ứng với một số giá trị **χ2** nhỏ nhất.Ví dụ dưới đây sẽ chứng minh cho điều ta nói ở trên.

Giả sử ta ước lượng được đồ dài của từ khóa là 4, biết được bản mã là:

**YITZU GRFFE TZZOC GSITS XUEAH EIKUT P**

và bản rõ là

**MICHIGAN TECHNOLOGICAL UNIVERSITY**

Cũng giống như ví dụ trước ta tính toán giá trị **χ2** của mỗi lần dịch chuyển tương ứng với các tập khối.

Bảng dưới đây là kết quả tính toán được giá trị **χ2** của mỗi lần dịch chuyển của mỗi khối. Để tiết kiêm không gian chúng ta chỉ lấy 4 giá trị **χ2** nhỏ nhất của mỗi tập khối và giá trị nhỏ nhất của mỗi khối được in đậm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dịch chuyển vị trí** | **Chữ cái tương ứng** | **Giá trị χ2 của khối 1** | **Giá trị χ2 của khối 2** | **Giá trị χ2 của khối 3** | **Giá trị χ2 của khối 4** |
| 0 | **A** |  | **2.2259** | 2.2292 |  |
| 1 | **B** |  | 2.9978 |  |  |
| 2 | **C** |  | 5.6245 | 3.6112 |  |
| 5 | **F** | 4.1750 |  |  |  |
| 6 | **G** |  | 5.2742 |  |  |
| 12 | **M** | 3.9131 |  |  | 3.4856 |
| 14 | **O** |  |  |  | 4.3258 |
| 15 | **P** |  |  | **1.9889** |  |
| 17 | **R** |  |  | 5.9857 |  |
| 18 | **S** | 4.6828 |  |  | **2.0008** |
| 20 | **U** | **2.3036** |  |  |  |
| 25 | **Z** |  |  |  | 3.6293 |

Dựa vào bảng này ta xác định được từ khóa là **UAPS** (do các chữ cái này tương ứng với giá trị **χ2** nhỏ nhất của mỗi khối) và kết quả giải mã là:

**EIEHAGCNLEEHFONOYIEADUPINETSATA**

Điều này chắc chắn là không đúng bởi vì nếu chúng ta sắp xếp bản rõ,bản mã và văn bản được giải mã với nhau như dưới đây thì chúng ta sẽ thấy có vấn đề.

**MICHIGAN TECHNOLOGICAL UNIVERSITY  
YITZUGRF FETZZOCGSITSX UEAHEIKUTP  
EIEHAGCN LEEHFONOYIEAD UPINETSATA**

Rõ ràng là sự dịch chuyển thứ 2 vị trí tương ứng với chữ cái A và sự dịch chuyển 4 vị trí tương ứng với chữ cái S là chính xác, bởi vì các chữ cái ở các vị trí tương ứng của bản rõ và bản mã giống hệt nhau. Nhưng mà sự dịch chuyển 1 vị trí tương ứng với chữ cái U và sự dịch chuyển ba vị trí tương ứng với chữ cái P là không chính xác.Do đó, để xác định được từ khóa ta phải xét đến các giá trị **χ2** nhỏ nhất ở bảng trên, khi đó từ khóa chưa biết có thể như sau

|  |  |  |  |
| --- | --- | --- | --- |
| F | A | A | S |
| M | C |
| S | P |
| U | R |

Có tất cả 16 vị trí kết hợp có thể (các từ khóa có thể)

**FAAS MAAS SAAS UAAS  
FACS MACS SACS UACS  
FAPS MAPS SAPS UAPS  
FARS MARS SARS UARS**

Kết quả của việc giải mã bản mã với 16 từ khóa có thể

|  |  |  |  |
| --- | --- | --- | --- |
| FAAS | TITHPGRN AETHUOCONTITAS UEICEISPTP | FACS | TIRHPGPN AERHUOAONIRAS UCICEGSPTN |
| FAPS | TIEHPGCN AEEHUONONIEAS UPICETSPTA | FARS | TICHPGAN AECHUOLONICAS UNICERSPTY |
| MAAS | MITHIGRN TETHNOCOGITAL UEIVEISITP | MACS | MIRHIGPN TERHNOAOGIRAL UCIVEGSITN |
| MAPS | MIEHIGCN | **MARS** | **MICHIGAN TECHNOLOGICAL UNIVERSITY** |
| SAAS | GITHCGRN NETHHOCOAITAF UEIPEISCTP | SACS | GIRHCGPN |
| UAAS | EITHAGRN LETHFOCOYITAD UEINEISATP | UACS | EIRHAGPN LERHFOAOYIRAD UCINEGSATN |
| SAPS | GIEHCGCN NEEHHONOAIEAF UPIPETSCTA | SARS | GICHCGAN NECHHOLAICAF UNIPERSCTY |
| UAPS | EIEHAGCN LEEHFONOYIEAD UPINETSATA | UARS | EICHAGAN LECHFOLOYICAD UNIVERSATY |

Khi đó từ khóa chính xác là **MARS**.Thông thường để khôi phục lại từ khóa với phương pháp vét cạn thì cần mất tới 264 =456,976 phép tính nhưng đối với phương pháp này thì chỉ cần 16 phép tính.Nói chung phương pháp này có tỷ lệ thành công cao nếu bản mã dài và từ khóa ngắn.

* 1. **Phân tích tần số**

Một số mật mã ban đầu chỉ sử dụng duy nhất một từ khóa. Những mật mã này còn được gọi là mật mã thay thế đơn giản hoặc mật mã đơn ký tự. Nói chung, có hai hằng số nguyên a và b, một chữ cái x trong bản bản rõ được mã hóa thành chữ cái (ax + b) mod 26 trong bản mã. Nếu a bằng 1 thì đây là mật mã Caesar. Trong thực tế, nếu chúng ta chọn một từ khóa có độ dài 1 trong mật mã Vigenere thì nó trở nên mật mã Caesar. Phương pháp **χ2** là một cách hiệu quả để giải mã một bản mã được mã hóa bằng cách sử dụng phương pháp thay thế đơn giản, bởi vì chỉ có một chữ cái trong từ khóa và do đó chỉ có một khối. Chon nên, chúng ta chỉ cần chọn độ dài từ khóa là 1 và sử dụng phương pháp **χ2** để tìm ra sự phù hợp nhất.

Phương pháp **χ2** là một biện pháp có tính chất phù hợp tốt, cái mà làm việc với hai chuỗi giá trị. Trong trường hợp của chúng ta có hai chuỗi giá trị là tần số của chữ cái tiếng anh và tần số chữ cái của một lần dịch chuyển cụ thể đối với chỉ một khối. Sự dịch chuyển của một khối cái mà sinh ra giá trị **χ2** nhỏ nhất là có khả năng được mã hóa bằng chữ cái tương ứng với sự dịch chuyển đó. Cho nên, chuyển dịch khối 26 lần và tìm chữ cái có giá trị χ2 nhỏ nhất có thể sẽ tìm ra từ khóa có duy nhất một chữ cái.

Xét bản mã sau đây được mã hóa bằng mật mã Caesar:

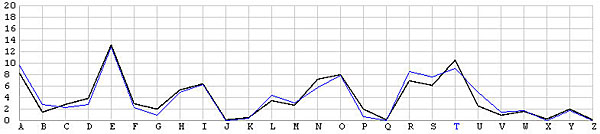
**YKBXG WLKHF TGLVH NGMKR FXGEX GWFXR HNKXT KLBVH FXMHU NKRVT XLTKG HMMHI KTBLX ABFMA XXOBE MATMF XGWHE BOXLT YMXKM AXFMA XZHHW BLHYM BGMXK KXWPB MAMAX BKUHG XLLHE XMBMU XPBMA VTXLT KMAXG HUEXU KNMNL ATMAM HEWRH NVTXL TKPTL TFUBM BHNLB YBMPX KXLHB MPTLT ZKBXO HNLYT NEMTG WZKBX OHNLE RATMA VTXLT KTGLP XKWBM AXKXN GWXKE XTOXH YUKNM NLTGW MAXKX LMYHK UKNMN LBLTG AHGHN KTUEX FTGLH TKXMA XRTEE TEEAH GHNKT UEXFX GVHFX BMHLI XTDBG VTXLT KLYNG XKTE**

Dưới đây là bảng chứa giá trị **χ2** của mỗi mỗi lần dịch chuyển với các chữ cái tương ứng.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** |
| 11.7 | 11.4 | 12.9 | 3.79 | 16.1 | 5.56 | 4.08 | 3.13 | 19.1 |
| **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** |
| 9.90 | 5.04 | 13.3 | 14.2 | 18.8 | 21.2 | 7.18 | 6.39 | 9.75 |
| **S** | **T** | **U** | **V** | **W** | **X** | **Y** | **Z** |  |
| 4.99 | **0.0852** | 17.9 | 8.64 | 15.0 | 1.87 | 26.5 | 2.72 |  |

Do giá trị **χ2** tương ứng với chữ cái **T** là nhỏ nhất nên bản mã ở trên có thể được mã hóa bởi **T**

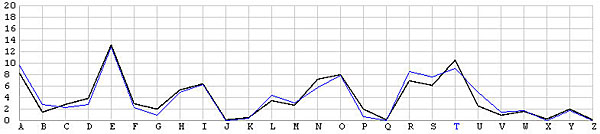
Hình ảnh dưới dây có hai chứa hai đồ thị tần số. Trục nằm ngang chứa 26 chữ cái tiếng anh và trục nằm dọc hiển thị giá trị tần số tương ứng với cái chữ cái này. Đường màu đen là tần số chữ cái tiếng anh điển hình và đường màu xanh hiển thị ra tần số của mỗi chữ cái thu được từ lần dịch chuyển T.



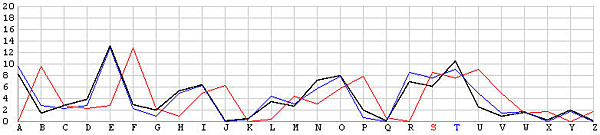
Rõ ràng là cả hai đồ thị tần số trên đều khớp với nhau rất tốt. Do đó, chúng ta chắc chắn rằng bản mã được mã hóa bằng cách sử dụng từ khóa T.

Chúng ta hãy nhìn vào các đồ thị tần số của một số thay đổi không chính xác. Những hình ảnh sau đây hiện thị sự dịch chuyển từ T trở về L. Trong sự dich chuyển T có một đỉnh cao ở chữ E. Khi sự dịch chuyển thay đổi từ T sang L, đỉnh này di chuyển sang phải từ E tới N. Trong thực tế, khi chúng ta di chuyển từ sự dịch chuyển T sang bên trái một vị trí một lần thì đỉnh dịch chuyển từ E sang bên phải một vị trí tại một lần. Dựa trên sự quan sát này, một phân tích tần số, nói nôm na là chỉ vẽ biểu đồ tần số của một sự thay đổi cụ thể và trượt biểu đồ tần số này sang trái hoặc phải cho đến khi nó phù hợp với đồ thị tần số tiếng Anh cho càng gần càng tốt. Sự phù hợp tốt nhất này có thể đem lại từ khóa chưa biết.

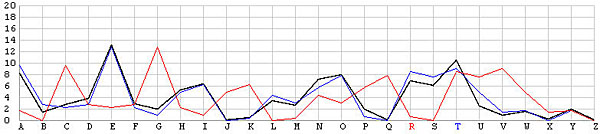
T



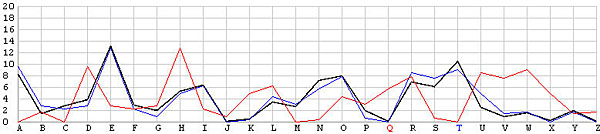
S



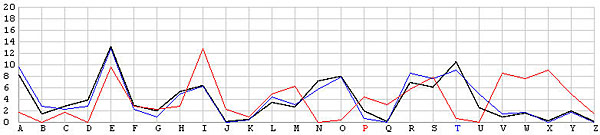
R



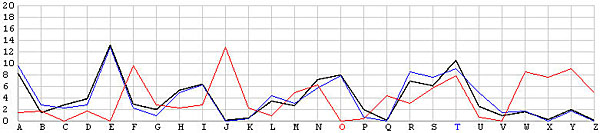
Q



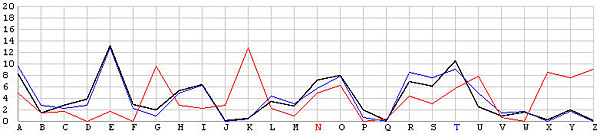
P



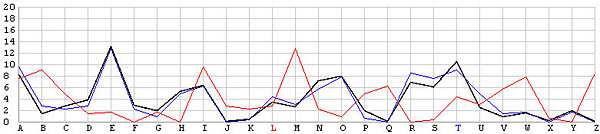
O



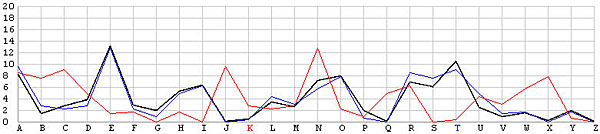
N



M



L



# **PHẦN IV. KẾT LUẬN CHUNG**

**TÀI LIỆU THAM KHẢO**