BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG
-------o0o-------

ISO 9001:2008

TÌM HIỂU LÝ THUYẾT M- DÃY VÀ ỨNG DỤNG CỦA NÓ TRONG MẬT MÃ HỌC

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ Thông tin

Sinh viên thực hiện:	Nguyễn Thị Hồng
Giáo viên hướng dẫn:	TS.Hồ Văn Canh
Mã số sinh viên:	1351010038

HẢI PHÒNG - 2013

MỤC LỤC

LỜI NÓI ĐẦU

Như ta đều biết ngày nay sự phát triển nhanh chóng của khoa học công nghệ mà nổi bật nhất là sự phát triển của công nghệ thông tin và mạng máy tính đang trở thành một nhu cầu không thể thiếu trong cuộc sống hàng ngày của mỗi con người. Nhưng nói đến mạng máy tính là nói đến sự chia sẻ tài nguyên của những người dùng. Vậy làm thế nào để mỗi cá nhân tổ chức, quốc gia có những thông tin nhạy cảm không bị rơi vào tay những đối tượng không mong muốn? Thế giới đã có nhiều phương pháp bảo vệ những thông tin quan trọng, trong đó việc bảo vệ thông tin nhạy cảm bằng kỹ thuật mật mã được sử dụng rộng rãi hơn cả. Hiện nay việc bảo vệ thông tin được mã hóa bởi khóa thuật toán có ý nghĩa quan trọng trong ngoại giao, quốc phòng an ninh. Nhưng việc mã hóa bởi khóa thuật toán luôn tạo ra dãy giả ngẫu nhiên tuần hoàn có chu kỳ là điều không ai mong muốn. Vấn đề đặt ra là làm sao tạo ra dãy giả ngẫu nhiên với chu kỳ cực đại mà người ta gọi là m-dãy. Trong khuôn khổ đồ án tốt nghiệp em chọn đề tài: ”Tìm hiểu lý thuyết m dãy và ứng dụng của nó trong mật mã học”, làm nội dung báo cáo của mình. Do khả năng và thời gian không cho phép, em chỉ nêu ra một ví dụ ứng dụng có tính chất minh họa.

Nội dung đồ án của em gồm có các chương chính sau:

Chương 1: Mật mã và an toàn thông tin

Chương 2: Lý thuyết m-dãy

Chương 3: Ứng dụng lý thuyết m-dãy vào việc mã hóa bởi hệ mã Vigenere

CHƯƠNG I: TỔNG QUAN AN TOÀN BẢO MẬT THÔNG TIN VÀ MẬT MÃ

1.1. Nội dung của an toàn và bảo mật thông tin

Khi nhu cầu trao đổi thông tin dữ liệu ngày càng lớn và đa dạng, các tiến bộ về điện tử - viễn thông và công nghệ thông tin không ngừng được phát triển ứng dụng để nâng cao chất lượng và lưu lượng truyền tin thì các quan niệm ý tưởng và biện pháp bảo vệ thông tin dữ liệu cũng được đổi mới. Bảo vệ an toàn thông tin dữ liệu là một chủ đề rộng, có liên quan đến nhiều lĩnh vực và trong thực tế có thể có rất nhiều phương pháp được thực hiện để bảo vệ an toàn thông tin dữ liệu. Các phương pháp bảo vệ an toàn thông tin dữ liệu có thể được quy tụ vào ba nhóm sau:

- Bảo vệ an toàn thông tin bằng các biện pháp hành chính.
- Bảo vệ an toàn thông tin bằng các biện pháp kỹ thuật(phần cứng).
- Bảo vệ an toàn thông tin bằng các biện pháp thuật toán(phần mềm).

Ba nhóm trên có thể được ứng dụng riêng rẽ hoặc phối kết hợp. Môi trường khó bảo vệ an toàn thông tin nhất và cũng là môi trường đối phương dễ xâm nhập nhất đó là môi trường mạng và truyền tin. Biện pháp hiệu quả nhất và kinh tế nhất hiện nay trên mạng truyền tin và mạng máy tính là biện pháp thuật toán.

An toàn thông tin bao gồm các nội dung sau:

- Tính bí mật: tính kín đáo riêng tư của thông tin.
- Tính xác thực của thông tin, bao gồm xác thực đối tác(bài toán nhận danh), xác thực thông tin trao đổi.
- Tính trách nhiệm: đảm bảo người gửi thông tin không thể thoái thác trách nhiệm về thông tin mà mình đã gửi.

Để đảm bảo an toàn thông tin dữ liệu trên đường truyền tin và trên mạng máy tính có hiệu quả thì điều trước tiên là phải lường trước hoặc dự đoán trước các khả năng không an toàn, khả năng xâm phạm, các sự cố rủi ro có thể xảy ra đối với thông tin dữ liệu được lưu trữ và trao đổi trên đường truyền tin cũng như trên mạng. Xác định càng chính xác các nguy cơ nói trên thì càng quyết định được tốt các giải pháp để giảm thiểu các thiệt hại.

Có hai loại hành vi xâm phạm thông tin dữ liệu đó là: vi phạm chủ động và vi phạm thụ động. Vi phạm thụ động chỉ nhằm mục đích cuối cùng là nắm bắt được thông tin(đánh cắp thông tin). Việc làm đó có khi không biết được nội dung cụ thể

nhưng có thể dò ra được người gửi, người nhận nhờ thông tin điều khiển giao thức chứa trong phần đầu các gói tin. Kẻ xâm nhập có thể kiểm tra được số lượng, độ dài và tần số trao đổi. Vì vậy vi phạm thụ động không làm sai lệch hoặc hủy hoại nội dung thông tin dữ liệu được trao đổi. Vi phạm thụ động thường khó phát hiện nhưng có thể có những biện pháp ngăn chặn hiệu quả. Vi phạm chủ động là dạng vi phạm có thể làm thay đổi nội dung, xóa bỏ, làm trễ, sắp xếp lại thứ tự hoặc làm lặp lại gói tin tại thời điểm đó hoặc sau đó một thời gian. Vi phạm chủ động có thể thêm vào một số thông tin ngoại lai để làm sai lệch nội dung thông tin trao đổi. Vi phạm chủ động dễ phát hiện nhưng để ngăn chặn hiệu quả thì khó khăn hơn nhiều.

Một thực tế là không có một biện pháp bảo vệ an toàn thông tin dữ liệu nào là an toàn tuyệt đối. Một hệ thống dù được bảo vệ chắc chắn đến đâu cũng không thể đảm bảo là an toàn tuyệt đối.

1.1.1. Các chiến lược an toàn hệ thống

- ✓ **Giới hạn quyền hạn tối thiểu(Last Privilege)**

Đây là chiến lược cơ bản nhất theo nguyên tắc này bất kỳ một đối tượng nào cùng chỉ có những quyền hạn nhất định đối với tài nguyên mạng, khi thâm nhập vào mạng đối tượng đó chỉ được sử dụng một số tài nguyên nhất định.

- ✓ **Bảo vệ theo chiều sâu(Defence In Depth)**

Nguyên tắc này nhắc nhở chúng ta: Không nên dựa vào một chế độ an toàn nào dù cho chúng rất mạnh, mà nên tạo nhiều cơ chế an toàn để tương hỗ lẫn nhau.

- ✓ **Nút thắt(Choke Point)**

Tạo ra một "cửa khẩu" hẹp, và chỉ cho phép thông tin đi vào hệ thống của mình bằng con đường duy nhất chính là "cửu khẩu" này. Phải tổ chức một cơ cấu kiểm soát và điều khiển thông tin đi qua cửa này.

- ✓ **Điểm nối yếu nhất(Weakest Link)**

Chiến lược này dựa trên nguyên tắc: "Một dây xích chỉ chắc tại mắt duy nhất, một bức tường chỉ cứng tại điểm yếu nhất"

Kẻ phá hoại thường tìm những chỗ yếu nhất của hệ thống để tấn công, do đó ta cần phải gia cố các yếu điểm của hệ thống. Thông thường chúng ta chỉ quan tâm đến kẻ tấn công trên mạng hơn là kẻ tiếp cận hệ thống, do đó an toàn vật lý được coi là yếu điểm nhất trong hệ thống của chúng ta.

- ✓ **Tính toàn cục:**

Các hệ thống an toàn đòi hỏi phải có tính toàn cục của các hệ thống cục bộ. Nếu có một kẻ nào đó nào đó có thể bẻ gãy một cơ chế an toàn thì chúng có thể thành công bằng cách tấn công hệ thống tự do của ai đó và sau đó tấn công hệ thống từ nội bộ bên trong.

- ✓ **Tính đa dạng bảo vệ**

Cần phải sử dụng nhiều biện pháp bảo vệ khác nhau cho hệ thống khác nhau, nếu không có kẻ tấn công vào được một hệ thống thì chúng cũng dễ dàng tấn công vào các hệ thống khác.

1.1.2. Các mức bảo vệ trên mạng:

Vì không thể có một giải pháp an toàn tuyệt đối nên người ta thường phải sử dụng đồng thời nhiều mức bảo vệ khác nhau tạo thành nhiều hàng rào chắn đối với các hoạt động xâm phạm. Việc bảo vệ thông tin trên mạng chủ yếu là bảo vệ thông tin cất giữ trong máy tính, đặc biệt là các server trên mạng. Bởi thế ngoài một số biện pháp nhằm chống thất thoát thông tin trên đường truyền mọi cố gắng tập trung vào việc xây dựng các mức rào chắn từ ngoài vào trong cho các hệ thống kết nối vào mạng. Thông thường bao gồm các mức bảo vệ sau:

- ✓ **Quyền truy nhập**

Lớp bảo vệ trong cùng là quyền truy nhập nhằm kiểm soát các tài nguyên của mạng và quyền hạn trên tài nguyên đó. Dĩ nhiên là kiểm soát được các cấu trúc dữ liệu càng chi tiết càng tốt. Hiện tại việc kiểm soát thường ở mức tệp.

- ✓ **Đăng ký tên / mật khẩu**

Thực ra đây cũng là kiểm soát quyền truy nhập, nhưng không phải truy nhập ở mức thông tin mà ở mức hệ thống. Đây là phương pháp bảo vệ phổ biến nhất vì nó đơn giản ít phí tổn và cũng rất hiệu quả. Mỗi người sử dụng muốn được tham gia vào mạng để sử dụng tài nguyên đều phải có đăng kí tên và mật khẩu trước. Người quản trị mạng có trách nhiệm quản lý, kiểm soát mọi hoạt động của mạng và không gian(nghĩa là người sử dụng chỉ được truy nhập trong một khoảng thời gian nào đó tại một vị trí nhất định nào đó).

Về lý thuyết nếu mọi người đều giữ kín mật khẩu và tên đăng ký của mình thì sẽ không xảy ra các truy nhập trái phép. Song điều đó khó đảm bảo trong thực tế vì nhiều nguyên nhân rất đời thường làm giảm hiệu quả của lớp bảo vệ này. Có thể khắc phục bằng cách người quản trị mạng chịu trách nhiệm đặt mật khẩu hoặc thay đổi mật khẩu hoặc thay đổi mật khẩu theo thời gian.

✓ **Mã hóa dữ liệu**

Để bảo mật thông tin trên đường truyền người ta sử dụng các phương pháp mã hóa. Dữ liệu bị biến đổi từ dạng nhận thức được sang dạng không nhận thức được theo một thuật toán nào đó và sẽ được biến đổi ngược lại ở trạm nhận(giải mã). Đây là lớp bảo vệ thông tin rất quan trọng.

✓ **Bảo vệ vật lí**

Ngăn cản các truy nhập vật lý vào hệ thống. Thường dùng các biện pháp truyền thống như ngăn cấm tuyệt đối người không phận sự vào phòng đặt máy mạng, dùng ổ khóa trên máy tính hoặc các máy trạm không có ổ mềm.

✓ **Tường lửa**

Ngăn chặn thâm nhập trái phép và lọc bỏ các gói tin không muốn gửi hoặc nhận vì các lý do nào đó để bảo vệ một máy tính hoặc cả mạng nội bộ(intranet).

✓ **Quản trị mạng**

Trong thời đại phát triển của công nghệ thông tin, mạng máy tính quyết định toàn bộ hoạt động của một cơ quan, hay một công ty xí nghiệp. Vì vậy việc bảo đảm cho hệ thống máy tính hoạt động một cách an toàn, không xảy ra sự cố là một công việc cấp thiết hàng đầu. Công tác quản trị mạng máy tính phải được thực hiện một cách khoa học đảm bảo các yêu cầu sau:

- Toàn bộ hệ thống hoạt động bình thường trong giờ làm việc.
- Có hệ thống dự phòng khi có sự cố về phần cứng hoặc phần mềm xảy ra.
- Backup dữ liệu quan trọng treo định kỳ.
- Bảo dưỡng mạng định kì.
- Bảo mật dữ liệu, phân quyền truy cập, tổ chức nhóm làm việc trên mạng.

1.1.3. An toàn thông tin bằng mật mã

Mật mã là ngành khoa học chuyên nghiên cứu các phương pháp truyền tin bí mật. Mật mã bao gồm: Lập mã và phá mã. Lập mã bao gồm hai quá trình: mã hóa và giải mã.

Để bảo vệ thông tin trên đường truyền người ta thường biến đổi nó từ dạng nhận thức được sang dạng không nhận thức được trước khi truyền đi trên mạng, quá trình này được gọi mã hóa thông tin(encryption), ở trạm nhận phải thực hiện quá trình ngược lại, tức là biến đổi thông tin từ dạng không nhận thức được(dữ liệu đã được mã hóa) về dạng nhận thức được(dạng gốc), quá trình này được gọi là giải mã. Đây là lớp bảo vệ thông tin rất quan trọng và được sử dụng rộng rãi trong môi trường mạng.

Để bảo vệ thông tin bằng mật mã người ta thường tiếp cận theo hai hướng:

- Theo đường truyền(Link_Oriented_Security)
- Từ nút đến nút(End_ to_End)

Theo cách thứ nhất thông tin được mã hóa để bảo vệ trên đường truyền giữa hai nút mà không quan tâm đến nguồn và đích của thông tin đó. Ở đây ta lưu ý rằng thông tin chỉ được bảo vệ trên đường truyền, tức là ở mỗi nút đều có quá trình giải mã sau đó mã hóa để truyền đi tiếp, do đó các nút cần phải được bảo vệ tốt.

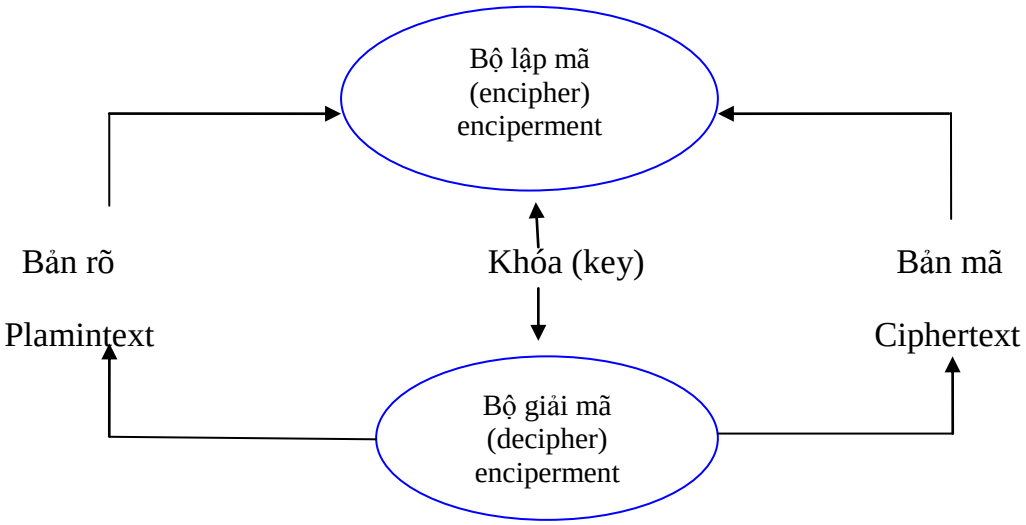
Ngược lại theo cách thứ hai thông tin trên mạng được bảo vệ trên toàn đường truyền từ nguồn đến đích. Thông tin sẽ được mã hóa ngay sau khi mới tạo ra và chỉ được giải mã khi về đến đích. Cách này mắc phải nhược điểm là chỉ có dữ liệu của người dùng thì mới có thể mã hóa được còn dữ liệu điều khiển thì giữ nguyên để có thể xử lý tại các nút.

1.2. LÝ THUYẾT MẬT MÃ

1.2.1 Giới thiệu

Lý thuyết mật mã là khoa học nghiên cứu cách viết bí mật, trong đó các bản rõ (paint text, clear text) được biến đổi thành bản mật mã(cipher text và sryptogrm). Quá trình biến đổi đó gọi là sự mã hóa(enciperment, encryption). Quá trình ngược lại, biến đổi từ bản mã thành bản rõ được gọi là sự giải mã(decipherment hay decryption).

Cả hai quá trình trên được điều khiển bởi một (hay nhiều) khóa mật mã:



Có 2 kiểu mật mã cơ sở: chuyển vị(transposition) và thay thế (sustituion)

a. Mã hóa chuyển vị sắp xếp lại các bít hay các ký tự trong dữ liệu.

Thí dụ trong "mật mã răng cưa" bản rõ được viết theo hình thức răng cưa, bản mã nhận được bằng cách lấy từng hàng.

Đ				Ọ				N				H				H			
	Ạ		H		C		Â		L		P		Ả		P		Ò		G
		I				D				Ậ				I				N	

ĐẠI HỌC DÂN LẬP HẢI PHÒNG ⟶ ĐỌNHH Ạ H C AALP ARP OFG ID ẬIN

Khóa của mật mã được cho bởi độ sâu của đường răng cưa, cụ thể trong thí dụ trên giá trị khóa là 3.

a. Mật mã kiểu thay thế cho các bit, các kí tự hay khóa các ký tự hàng các đối tượng thay thế.

Một loại mât mã đơn giản chuyển vị mỗi chứ cái trong bảng chữ cái tiếng Anh theo chiều tiến k vị trí(qua Z thì vòng về A); k là khóa mật mã. Mật mã này thường được gọi là mã Cacsar vì Julia cacsar đã dùng với K=3.

Thí dụ:

PASCAL

↓

SDVSFO

Mật mã(code) là một loại mật mã thay thế đặc biệt dùng một cuốn sách mã (code book) làm khóa. Các từ hay câu của bản rõ được đưa vào danh sách mã cùng với các thay thế của bản mã được chỉ rõ :

Từ	mã	C500
C	0907	↓
500	2002	
UNIVESITY	3636	0907 2002 3636

Thuật ngữ mã đôi khi được dùng để chỉ một kiểu viết mã(cipher: một phương pháp viết bí mật) bất kỳ.

Trong các ứng dụng tin học, chuyển vị thường được tổ hợp với thay thế. Chẳng hạn tiêu chuẩn mã hóa dữ liệu(Data encryption standard (DES), mã hóa các khối 64 bit bằng cách dùng một tổ hợp của dịch chuyển và thay thế.

Thám mã(cryptanalysis) là khoa học nghiên cứu về các phương pháp phá mật mã(brecking cipher).

Một mật mã là phá được(breakable) nếu có thể xác định được bản rõ hay khóa từ bản mã, hay xác định được khóa từ cặp bản rõ –bản mã.

Có 3 phương pháp tấn công cơ bản:

- Chỉ với bản mã(cipphertext only attack) : có thể biết phương pháp mã hóa ngôn ngữ của bản rõ, chủ đề của bản rõ là một bản rõ và một số từ có thể có.
- Với bản rõ đã biết(know plaintext attack): biết một số cặp bản rõ, bản mã.

Thí dụ thông báo được mã hóa của người dùng gửi tới máy tính với chuẩn LOGIN; các chương trình đã được mã hóa dễ bị tấn công vì sự xuất hiện của từ khóa begin, end , var, procedure, if, then , for….

- Với bản rõ được chọn(chosen plain text atack) (theo nghĩa người thám mã có thể được (nhận được) bản mã ứng với bãn rõ đã được chọn lọc). Đây là trường hợp thuận lợi nhất cho người thám mã. Các hệ số cơ sở dữ liệu có thể dễ bị tấn công bởi loại máy này vì người dùng có thể xem(insert) các phần tử vào cơ sở dữ liệu và sau đó quan sát các thay đổi trong bản mã được lưu giữ.

Một phương pháp mật mã là an toàn không điều kiện(unonditionally secure) nếu như dù có cách được bao nhiêu bản mã thì cũng không đủ thông tin trong bản mã để xác định được bản rõ một cách duy nhất.

Trừ một ngoại lệ, mọi phương pháp viết mật mã đều phá vỡ được nếu như các tài nguyên vô hạn. Do đó, ta quan tâm đến mật mã không thể phá vỡ được về mặt tính toán(computationnally infeasible to break). Một phương pháp mật mã an toàn về mặt tính toán(computationnally secure or song) nếu nó không thể bị phá vỡ bởi sự phân tích có hệ thống các tài nguyên sẵn có.

Cuối cùng ta có công thức tóm tắt sau:

Crytolory = crytogaphy + crytanalysis

Đối tượng cơ bản của mật mã là tạo khả năng liên lạc trên một kênh không mật của 2 người sử dụng sao cho đối phương không hiểu được thông tin truyền đi, kênh này có thể là một đường dây điện thoại hoặc một mạng máy tính . Thông tin Việt Anh gửi cho Sơn(bản rõ) có thể là một văn bản tiếng Anh, các dữ liệu bằng số hoặc bất cứ tài liệu nào có cấu trúc tùy ý. Việt Anh sẽ mã hóa bản rõ bằng một khóa đã được xác định trước và gửi bản mã kết quả trên kênh. Dũng có bản mã trên kênh

song song bản mã thu trộm được trên kênh song song không thể xác định được nội dung bản rõ, nhưng Sơn(người đã biết khóa mã) có thể giải mã và thu được bản rõ .

Dưới đây là mô tả hình thức hóa nội dung bằng khái niệm toán học :

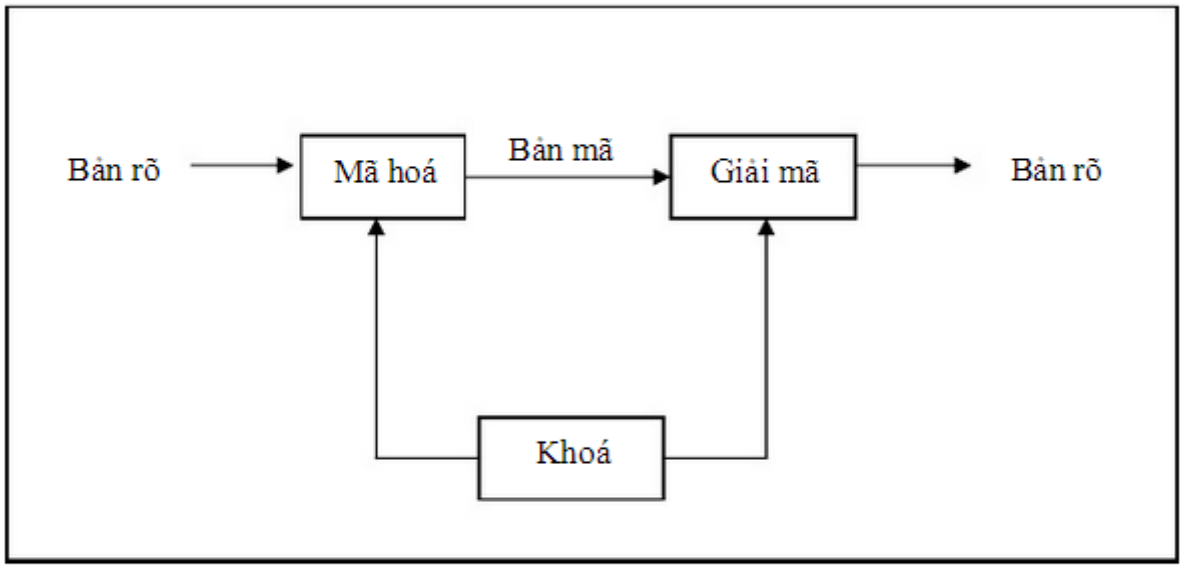
1.2.2. Định nghĩa :

Một hệ mật là một bộ 5(P,C,K,E,D) thỏa mã các điều kiện sau :

1) P là một tập hữu hạn các bản rõ(Plain Text),nó được gọi là không gian bản rõ.
2) C là một tập hữu hạn các bản mã(Crypto), nó còn gọi là không gian các bản mã.
3) K là một tập hữu hạn các khóa còn gọi là không gian khóa.
4) Đối với mỗi k∈ K có một quy mã e_k : P→C và một quy tắc giải mã tương ứng d_k ∈D : M→ R .Mỗi e_k : P→C và d_k : : C→ P là những hàm mà:

$$d_k(e_k(x))=x \text{ với mọi bản rõ } x\in P.$$

Khi bên A muốn gửi một tài liệu(bản rõ x) đến cho bên B, bên A sẽ mã hóa bản rõ đó bằng e_k: P →C và gửi đi trên đường truyền qua cho bên B, bên B nhận được sau đó sẽ giải mã bằng d_k : C->P để thu lại bản rõ ban đầu. Bên A và B sẽ áp dụng thủ tục sau dùng hệ mật khóa riêng. Trước tiên họ chọn một khóa ngẫu nhiên k∈ K . Điều này được thực hiện khi họ ở cùng một chỗ và không bị bên C theo dõi hoặc họ có một kênh mật trong trường hợp họ ở xa nhau. Sau đó giả sử bên A muốn gửi một thông báo cho B trên một kênh không mật và ta xem thông báo này là một chuỗi:
x = $x_1, x_2, \dots, x_n$ với số nguyên b ≥ 1 nào đó. Ở đây mỗi kí hiệu của mỗi bản rõ xi ∈ P , 1 ≤ i ≤ n . Mỗi xi sẽ được mã hóa bằng quy tắc mã e_k với khóa K xác định trước đó. Bởi vậy bên A sẽ tính 1 ≤ i ≤ n và chuỗi bản mã nhận được y = $y_1, y_2, \dots, y$ sẽ được gửi trên kênh. Khi bên B nhận được y = $y_1, y_2, \dots, y$ anh ta sẽ giải mã bằng hàm giải mã d và thu được bản rõ gốc x = $x_1, x_2, \dots, x_n$.



Mã hoá với khoá mã và khoá giải giống nhau

1.2.3. Phân loại hệ mật mã

Có nhiều cách để phân loại hệ mật mã. Dựa vào cách truyền khóa có thể phân các hệ mật mã thành hai loại:

- Hệ mật mã đối xứng(hay còn gọi là mật mã khóa bí mật): là những hệ mật mã dùng chung một khóa cả trong quá trình mã hóa dữ liệu và giải mã dữ liệu. Do đó khóa phải được giữ bí mật tuyệt đối.

- Hệ mật mã bất đối xứng(hay còn gọi là mật mã khóa công khai): Hay còn gọi là hệ mật mã công khai, các hệ mật mã này dùng một khóa để mã hóa sau đó dùng một khóa khác để giải mã, nghĩa là để mã hóa và giải mã là khác nhau. Các khóa này tạo nên từng cặp chuyển đổi ngược nhau và không có khóa nào có thể suy ra từ khóa kia. Khóa dùng để mã hóa có thể công khai nhưng khóa dùng để giải mã phải giữ bí mật.

Ngoài ra nếu dựa vào thời gian đưa ra hệ mật mã ta còn có thể phân làm hai loại: Mật mã cổ điển và mật mã hiện đại.

Còn nếu dựa vào cách thức tiến hành mã thì hệ mật mã còn được chia làm hai loại: Mã dòng(tiến hành mã từng khối dữ liệu, mỗi khối lại dựa vào các khóa khác nhau, các khóa này được sinh ra từ hàm sinh khóa, được gọi là dòng khóa) và mã khối(tiến hành mã từng khối dữ liệu với khóa như nhau).

1.2.4. Tiêu chuẩn đánh giá hệ mật mã

Để đánh giá một hệ mật mã người ta thường đánh giá thông qua các tính chất sau:

- Độ an toàn: Một hệ mật mã được đưa vào sử dụng đầu tiên cần phải có độ an toàn cao. Ưu điểm của mật mã là có thể đánh giá được độ an toàn thông qua độ an toàn tính toán mà không cần cài đặt.

Một hệ mật mã được gọi là tốt thì nó cần phải đảm bảo các tiêu chuẩn sau:

- Chúng phải có phương pháp bảo vệ mà chỉ dựa trên sự bí mật của các khóa, công khai thuật toán.

Khi cho khóa công khai e_K và bản rõ P thì chúng ta dễ dàng tính được e_K(P)=C. Ngược lại khi cho d_k và bản mã C thì dễ dàng tính được $d_k(M) = P$. Khi không biết d_K thì không có khả năng để tìm được M từ C.

- Bản mã C không được có đặc điểm gây chú ý, nghi ngờ.
- Tốc độ mã và giải mã: Khi đánh giá hệ mật mã chúng ta phải chú ý đến tốc độ mã và giải mã. Hệ mật mã tốt thì thời gian mã và giải mã nhanh.
- Phân phối khóa: Một hệ mật mã phụ thuộc vào khóa, khóa này được truyền công khai hay truyền khóa bí mật. Phân phối khóa bí mật thì chi phí sẽ cao hơn so cới các hệ mật có khóa công khai. Vì vậy đây cũng là một tiêu chí khi lựa chọn hệ mật mã.

1.3. Các hệ mã cổ điển

1.3.1. Hệ mã hóa CAESAR

Hệ mã hoá CAESAR là một hệ mã hóa thay thế đơn làm việc trên bảng chữ cái tiếng Anh 26 ký tự(A, B, …, Z).

Trong hệ Caesar và các hệ tương tự còn lại ta sử dụng các số tự nhiên thay cho các ký tự - đánh số các ký tự trong bảng chữ cái theo thứ tự:

A	B	C	D	E	F	….	X	Y	Z
0	1	2	3	4	5	….	23	24	25

Các phép toán số học thực hiện theo modul 26. Có nghĩa là 26 đồng nhất với 0, 27 đồng nhất với 1, ….

Hệ Caesar sử dụng thuật toán mã hóa trong đó mỗi ký tự được thay thế bởi một ký tự khác được xác định bằng cách dịch ký tự cần mã hóa sang phải k bước theo modulo 26:

$E_k(\alpha) = (\alpha + k) \bmod 26$. Trong đó α là một ký tự, $0 \leq k \leq 26$

Thuật toán giải mã tương ứng D_k là lùi lại k bước trong bảng chữ cái theo modul 26:

$D_k(\alpha) = (\alpha - k) \bmod 26$

Không gian khóa của hệ Caesar bao gồm 26 số.

Ví dụ :

Mã Hóa:

Plaint: "TRICH"

Bản rõ số: 19 17 8 2 7

Khóa: k=3

Bản mã số: 22 20 11 5 10

Bản mã chữ là : "WULFK"

Giải mã: Lùi lại với k=3 ta thu được bản rõ: "TRICH"

Nhận xét: hệ mã hóa Caesar là hệ mã hóa cũ và không an toàn vì không gian khóa của nó rất nhỏ, do đó có thể thám mã theo phương pháp vét cạn. Khóa giải mã có thể tính ngay được từ khóa mã hóa. Do chỉ có 26 khóa nên ta có thể thử lần lượt các khóa cho đến khi tìm được khóa đúng.

1.3.2. Mã hóa Ceasar mở rộng

Đây là một dạng mở rộng của mã Ceasar, giả sử các ký tự trong bảng chữ cái (với số lượng chữ cái là n) được đánh số thứ tự từ 0,….(n-i). Khi đó :
Mã hóa: E i=ai+b với USCLN(a,n) =1
Giải mã : D:i=(i-b).a^{-1} (mod n)

1.3.3. Mật mã thay đổi ngẫu nhiên

Phương pháp này được thực hiện đơn giản bằng cách thay thế ngẫu nhiên một ký tự bằng một ký tự khác trong bảng chữ cái.
Mã hóa: $E_k = y_i = x_i + k_i(\bmod 26)$
Giải mã: $D_k = x_i = y_i - k_i(\bmod 26)$

1.3.4. Mật mã sử dụng bảng chữ cái

Trong thực tế để thiết lập bảng chữ cái được hoán vị các ký tự, ta sử dụng một vài từ bí mật bao gồm các ký tự không lặp lại, và viết chúng vào một hình chữ nhật để sử dụng như một từ gợi nhớ .

Giả sử từ bí mật(secret word) là "Star word ", sẽ được đổi thành từ gợi nhớ "Starw".

Ta viết :

S T A R W

B C D E F

G H I J K

L M N O P

Q U V X Y

Z

Các cột được đọc thứ tự từ trên xuống dưới sẽ cho ta một sự chuyển đổi các ký tự từ plaintext sang ciphertext:

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

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

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

Ví dụ :

I K N O W T H A T I K N O W T H I N G

H U I N F O C S O H U I N F I N O C H I T

Phương pháp này cũng như phương pháp trên, rất dễ bị lộ mật mã(phá khóa)

Chú ý:

Để thay thế các ciphers các bảng chữ cái gồm ký tự, ta có khóa n! khóa có thể có, tức là số cách có thể xắp sếp lại n ký tự trong bản chữ cái. Nếu tất cả các khóa là như nhau thì tổng text cần thiết để phá khóa là(Shannon unicity dis tance) :N =H(K)/D=$\log_2$n!/D trong đó D là độ dư thừa của ngôn ngữ(3,2 bít/ký tự đối với tiếng Anh), suy ra:

N= $\log_2(\frac{26}{2})$ /3.2=27.6

Do đó, 27 hoặc 28 ký tự là đủ để phá khóa các mã này bằng việc' phân tích tần số lặp'.

Với các ciphers trên cơ sở đa thức được sử dụng cho phép chuyển đổi số học modular thì tổng số text đòi hỏi là :

N = $\log_2 26$ /3.2= 1.5

Do đó, 2 ký tự để bẻ gãy các mã này bằng "phân tích tần số lặp".

1.3.5. Mật mã hoán vị

Một kỹ thuật được sử đụng trên các bảng chữ cái là hoán vị các ký tự.

Plaintext → sắp sếp lại các ký tự → cipherstext

Ví dụ :

Ta viết plaintext CRYPTTANALYST như một ma trận 3*4:

1 2 3 4

C R Y P

T A N A

LY S T

Và đọc từng cột theo thứ tự 2,4,1,3 ta nhận được :

R A Y P A T C T C Y N S

Kỹ thuật này còn có thể được sử dụng cho các mảng n chiều .

1.3.6. Mã hóa thay thế cũng phát âm

Các ký tự xuất hiện nhiều lần có thể ánh xạ nhiều thành phần ký tự trong cipherstext để tránh bị phá khóa bởi " Phân tích tần số lặp ". Số ký tự mật mã với một ký tự cho trước được xác định bằng tần số lặp của nó trong ngôn ngữ xét.

Ví dụ giả sử bảng chữ cái được ánh xạ sang các số từ 1 đến 99 với :

Ánh xạ : E thành 17,19,23,47,64

A thành 8,20,25,49

R thành 1,29,65

T thành 16 ,31,85

Còn các ký tự khác thứ I sẽ ánh xạ thành 3*i.

Khi đó , plaintext sau :

MANY A SLIP TWIXT THE CUP ANH THE LIP sẽ trở thành:

M A N Y A S L I P T W I X T T H E C P A N D T H E L với các tương ứng: 20 08 39 72 20 54 33 24 15 16 66 24 69 31 85 21 17 06 60 45 25 39 09 16 21 47 33

1.3.7. Mã hóa thay thế sử dụng nhiều bảng chữ cái

Trong khi các mã khóa thay thế cũng phát âm che dấu phân bố tần số bằng cách sử dụng homophomic thì mã hóa thay thế sử dụng nhiều bảng chữ cái lại che dấu nó bởi nhiều phép thay thế trên nhiều bảng chữ cái.

Allbeti(năm 1568) đã dùng hai discs mà chúng được quay đi sau khi sử dụng. Sau một chu kỳ d nó sẽ tạo d bảng chữ cái cho cipher là C1,C2…..Cd .

Do đó, với một plaintext alphabert A:

$\int i$:A (+) Ci

Và message :M =$m_1 m_2$…..md$md_{+1} md_{+2}$……m_{2d}

Sẽ trở thành :

Ek(M)=f_1(mm_1))=f_2(m_2)… =f_d (md_1))f_1(md_{+1})f_2(md_{+2})…. f_d(m_{2d})

Với d=1 chúng ta sẽ trở lại trường hợp monoalphabetic ciphers. Bây giờ chúng sẽ nghiên cứu một số phương pháp tại monoalphabetic ciphers.

Vigenere cipher

Khóa được đặc tả bằng một chuỗi các ký tự :

K=k_1…..k_d

Trong đó ki(i=1,…d) cho biết giá trị phải dịch(shift) trong bảng chữ cái thứ i tức là:

Fi(a) = a+ki(mod n)

Ví dụ: mã hóa message INDIVIDUAL CHARATER với khóa HOST, ta có:

M= INDI VIDU ALCH TER

K=HOST HOST HOST HOS

Ek(m)=PBVB CWVN HZUA HFSV ASJ

Beauford cipher.

Tương tự như trên, nhưng ta sử dụng:

Fi(a) = (ki-a) (mod n)

Chú ý: trong trường hợp này ta có thể sử dụng hàm sau đây để nhận ciphertext.

f_i^{-1} = (ki-c)(mod n)

Do đó, cipher này thực sự đảo(invese) các ký tự của bảng chử cái và sau đó dịch chúng về bên phải đi(ki+1) vi trí. Ta có thể thấy bằng cách viết lại f_i như sau :

f_i(a) = [(n-1)-a+(k_i+1)](mod n)

ví dụ:

Giả sử k_i = D thì $f_i(a)$ = (d-a)(mod 26), sau đó để ý rằng D là ký tự số 3 của bảng chữ cái và thấy rằng (3-0)=3,(3-1)=2,(3-2)=1 ,……chúng ta sẽ có:

Plaintext : A B C D E F G H I J K L M….

Ciphertext :D C B A Z Y Ư V U T S R…

Variant beauford cipher.

Ta sử dụng công thức $f_i(a)$ = $(a-k_i)(\text{mod } n)$.

Vì $(a-k_i)$ = $a + (n-k_i)(\text{mod } n)$ nên Variant beauford cipher vì nếu một cái được sử dụng để mã hóa thì cái kia dùng để giải mã.

1.3.8. Mật mã Hill

Hệ mật mã thay thế gọi là mật mã Hill, mật mã này so với Lester S.Hill đưa ra năm 1929.

Giả sử m là số nguyên dương, đặt R = M = $(Z_{26})^m$, ý tưởng ở đây là lấy m tổ hợp tuyến tính của m ký tự trong một phân tử của bản rõ để tạo ra m ký tự của một phân tử của bản mã.

Ví dụ, nếu m =2 ta có thể viết một phần tử của bản rõ là x = (x1,x20 và một phần tử của bản mã là y = (y1, y2).Ở đây y1 và y2 đều là một tổ hợp tuyến tính của x1 và x2, chẳng hạn lấy :

$$y_1 = 11x_1 + 3x_2$$

$$y_2 = 8x_1 + 7x_2$$

Có thể viết gọn theo ký hiệu ma trận như sau:

$$(y_1 y_2) = (x_1 x_2)\begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix}$$

Nói chung có thể lấy ma trận K kích thước m×m làm khóa. Những phần tử ở hàng i cột j của K là $(k_{i,j})$, với x=$(x_1, x_2 \ldots\ldots x_m)$∈P và

k_i ∈ K, ta tính y = $e_K(x) = (y_{1,}y_2,\ldots y_m)$ như sau :

$$= (y_{1,}y_2,\ldots y_m)(x_1, x_2, \ldots, x_m)\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,m} \\ k_{2,1} & k_{2,2} & k_{2,m} \\ k_{m,1} & k_{m,2} & k_{m,m} \end{bmatrix}$$

Nói cách khác y = xK.

Ta nói rằng bản mã nhận được từ bản rõ nhờ phép biến đổi tuyến tính. Ta xét xem phải thực hiện giải mã như thế nào, tức phải làm thế nào để tính x từ y. Chúng ta lấy ma trận nghịch đảo K^{-1} để giả mã, bản mã được giải mã bằng công thức yK^{-1}.

Dưới đây là mô tả chính xác mật mã Hill trên $(Z_{26})^m$

Cho m là một số nguyên dương cố định. Cho P=C=$(Z_{26})^m$ và cho K = (các ma trận khả nghịch cấp m×m trên z_{26}) với một khóa k∈ K ta xác định :

$e_k(x)=x(K)$

$d_k(y)=yK^{-1}$

Và tất cả các phép toán thực hiện trên Z_{26}.

1.3.9. Mã Vigenere

Hệ mã Vigenere lấy tên của Blaise de Vigenere sống vào thế kỷ XVI

Sử dụng phép tương ứng A ⇔ 0, B⇔1,…..,Z⇔25 mô tả ở trên, ta có thể gắn cho mỗi khóa K với một chuỗi kí tự có độ dài m được gọi là từ khóa. Mật mã Vigenere sẽ mã hóa đồng thời m kí tự. Mỗi phần tử của bản rõ tương đương với m ký tự.

Xét một ví dụ:

Giả sử m=6 và từ khóa là CIPHER. Từ khóa này tương ứng với dãy số K = (2,5,7,9,8). Giả sử bản rõ là xâu:

Nguyenthithanhhuyen

Cho m là một số nguyên dương cố định nào đó. Định nghĩa P=C=K=$(z_{26})^m$.

Với khóa K=$(k_1, k_2, \dots, k_m)$ ta xác định:

$$e_k(x_1, x_2, \dots, x_m) = (x_1 + k_1, x_2 + k_2, \dots, x_m + k_m)$$

Và $$d_k(y_1, y_2, \dots, y_m) = (y_1 - k_1, y_2 - k_2, \dots \dots, y_m - k_m)$$

Trong đó tất cả các phép toán được thực hiện trong Z_{26}

Mật mã Vigenere

Ta sẽ biến đổi các phần tử của bản rõ thành các thặng dư theo modulo 26, viết chúng thành các nhóm 6 rồi cộng với từ khóa theo modulo 26 như sau:

13	6	20	24	4	13	19	7	8	19
2	5	7	9	8	2	5	7	9	8
15	11	1	7	12	25	24	14	17	1
7	0	13	7	7	20	24	4	13	
2	5	7	9	8	2	5	7	9	
9	5	20	16	15	22	3	11	22	

Dãy ký tự tương ứng của xâu bản mã sẽ là: P L B H M Z Y O R B J F U Q P C D L W

Để giải mã ta có thể dùng cùng từ khóa nhưng thay cho cộng, ta trừ cho nó theo modulo26. Ta thấy rằng các từ khóa có thể với số độ dài m trong mật mã Vigenere là 26^m bởi vậy, thậm chí với các giá trị m khá nhỏ, phương pháp tìm kiếm vét cạn cũng yêu cầu thời gian khá lớn.

Trong hệ mật Vigenere có từ khóa độ dài m, mỗi ký tự có thể được ánh xạ vào trong m ký tự có thể có(giả sử rằng từ khóa chứa m ký tự phân biệt). Một hệ mật mã như vậy được gọi là hệ mật thay thế đa biểu.

1.4. Các hệ mã dòng

Trong các hệ mật mã nghiên cứu trên, các phần tử liên tiếp của bản rõ đều được mã hóa bằng cùng một khóa K. Tức xâu bản mã nhận được có dạng:

$$M = y_1y_2\ldots\ldots y1y2\ldots..e_k(x_1)\ e_k(x_2)\ldots..$$

Các hệ mật mã thuộc dạng này thường được gọi là các mã khối. Một quan điểm sử dụng khác là mật mã dòng. Ý tưởng cơ bản ở đây là tạo ra một dòng khóa z = z_1z_2 ... ..và dùng nó để mã hóa một xâu bản rõ x = x_1x_2 ....theo quy tắc:

$$M = y_1y_2\ldots\ldots y1y2\ldots..e_{z1}(x_1)\ e_{z2}(x_2)\ldots..$$

Mã dòng hoạt động như sau : Giả sử k∈K là khóa và x_1x_2 ....là xâu bản rõ. Hàm fi được dùng để tạo zi(zi là phần tử thứ I của dòng thứ 1 của dòng khóa) trong đó fi là một hàm của khóa k và i-1 là kí tự đầu tiên của bản rõ :

$$Zi = fi(k, x_1x_2\ldots.x_{i-1)}$$

Phần tử thứ I của dòng khóa được dùng mã xi rạo ra yi = $e_{z1}(x_1)$. Bởi vậy, để mã hóa xâu bản rõ x_1x_2.....Ta phải tính liên tiếp $z_1y_1z_2y_2$.....Việc giải mã xâu bản mã y_1y_2... có thể thực hiện bằng cách tính liên tiếp : $z_1y_1z_2y_2$....

- ***Định nghĩa :***

Mật mã dòng là một bộ(P,C,K,L,E,F,D)thỏa mãn các điều kiện sau:

P là tập các hữu hạn các bản rõ có thể.

C là tập hữu hạn các bản mã có thể.

K là tập hữu hãn các khóa có thể.

L là hữu hạn bộ chữ của dòng khóa.

F=(f1f2….) là bộ tạo dòng khóa .Với i≥1,

$f_1: K \times P^{i-1} \longrightarrow L$

Với, mỗi z∈L có một quy tắc mã e_z∈ E và một qui tắc giãi mã tương ứng dz∈D , sao cho e_z:P $\longrightarrow$ C và d_z:C $\longrightarrow$ P là 1 hàm thỏa mãn:

$d_x(e_x((x) = x$ với mọi bản rõ x∈R.

Một phương pháp tạo dòng khóa hồi quy tuyến tính theo quan hệ đệ quy tuyến tính cấp m mà sẽ bị mã thám như trình bày tại 1.3 là:

$z_{i+m} = \sum_{j=0}^{m-1} CjZi - j$ mod2

Trong đó $c_1 c_2 \ldots . c_{m-1} \in Z_2$ là các hằng số cho trước.

Nhận xét :

Phép đệ qui này được nói là có bậc m vì mỗi số hạng phụ thuộc vào m số hạng đứng trước. Phép đệ qui này là tuyến tính bởi vì Z_{i+m} là 1 hàm tuyến tính của các số hạng đứng trước. Chú ý ta có thể lấy là C_o=1 mà không làm mất tính tổng quát. Trong trường hợp ngược lại phép đệ qui sẽ có bậc m-1.

Ở đây khóa K gồm 2m giá trị $k_1 k_2 \ldots\ldots . k_m, c_0 c_1 \ldots\ldots c_{m-1}$. Nếu ($k_1 k_2 \ldots\ldots . k_m$) = (0….0) thì dòng khóa sẽ chưa các số 0. Dĩ nhiên phải tránh điều này vì khi đó bản mã sẽ đồng nhất với bản rõ. Tuy nhiên:

Nếu chọn thích hợp các hằng số $c_0 c_1 \ldots\ldots c_{m-1}$ thì một vector khởi đầu khác ($k_1 k_2 \ldots\ldots . k_m$) sẽ tạo nên một dòng khóa có chu kì 2^m-1. Bởi vậy, một khóa ngắn sẽ tạo nên một dòng khóa có chu kì rất lớn. Đây là một tính chất đáng lưu tâm tạo khóa giả ngẫu nhiên.

Ví dụ minh họa:

Giả sử m = 4 và dòng khóa được tạo bởi quy tắc:

$z_{i+4} = z_i + z_{i+1} mod 2$

Nếu dòng khóa bắt đầu một véc tơ bất kỳ khác với véc tơ(0,0,0,0) thì ta thu được dòng khóa có chu kỳ 15.

Một véc tơ khởi đầu khác không bất kỳ khác sẽ tạo một hoán vị vòng(cyclic) của cùng dòng khóa.

Một hướng đáng quan tâm của phương pháp tạo dòng khóa hiệu quả bằng phần cứng là sử dụng bộ ghi hồi tiếp tuyến(hay LFSR). Ta dùng một bộ ghi ghi dịch có m tần, Vector($k_1 k_2 \ldots\ldots . k_m$)sẽ dùng để khởi tạo(đặt các giá trị ban đầu)cho thanh ghi dịch. Ở mỗi đơn bị thời gian, các phép toán sau sẽ được thực hiện đồng thời.

k_1 được tính ra dùng làm bít tiếp theo của dòng khóa.

$k_1 k_2 \ldots\ldots k_m$ sẽ dịch một lần về phía trái.

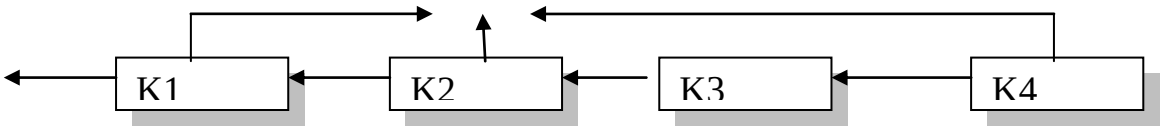
Giá trị mới k_m được tính bằng.

$\sum_{j=0}^{m-1} Cj\ k_{j+1}$ (Đây là hồi tiếp tuyến tính).

Ta nhận thấy rằng thao tác hồi tiếp tuyến tính sẽ được tiến hành bằng cách lấy tín hiệu ra từ một số tầng nhất định của thanh ghi(được xác định bởi các hằng Cj có giá trị "1") và tính tổng theo module 2(là phép hoặc loại trừ). Hình sau mô tả LFSR dùng để tạo dòng khóa được tạo theo quy tắc :

$z_{i+4}=(z_i+z_2) \bmod 2$

Thanh ghi dịch phản hồi (LFSR)



Một ví dụ về mã dòng không đồng bộ là mã khóa tự sinh như sau(Mật mã này do Vigenere đề xuất).

Mật mã khóa tự sinh

Cho P=C=K=L=Z_{26}

Cho $z_1 = K$ và $\boldsymbol{z_{i-1}(i \geq 2)}$

Với 0≤z≤25 ta xác định

$e_x(x) = x + z \bmod 26$

$d_z(y) = y - z \bmod 26$

(x,y∈ Z_{26})

Thám mã hệ mã dòng xây dựng trên LFSR.

Ta nhớ lại rằng, bản mã là tổng thể của modulo 2 của bản rõ và dòng khóa tức$y_i = x_i + z_i \bmod 2$. Dòng khóa được tạo từ$(z_1, z_2, \ldots z_m)$ theo quan hệ đệ quy tuyến tính:

$$z_{m+1} = \sum_{j=0}^{m-1} c_j z_{j+1} \bmod 2$$

Trong đó $c_0, \ldots . c_m \in Z_2$(và $c_0 = 1$)

Vì tất cả các phép toán này là tuyến tính nên có thể hy vọng rằng, hệ mật mã này có thể bị phá theo phương pháp tấn công với bản rõ đã biết như trường hợp mật mã Hill. Giả sử rằng, Minh có một xâu bản rõ $x_1x_2 \dots x_n$ và xâu bản mã tương ứng $y_1y_2 \dots y_n$. Sau đó tính các bít dòng khóa $z_i = x_i + y_i \text{mod} 2, 1 \le i \le n$. Ta cũng giả thiết rằng Minh cũng đã biết giá trị của m. Khi đó Oscar chỉ cần tính $c_0, \dots c_{m-1}$ để có thể tái tạo lại toàn bộ dòng khóa. Nói cách khác Oscar cần phải có khả năng xác định các giá trị m ẩn số.

Với i≥1 bất kì ta có:

$$z_{m+1} = \sum_{j=0}^{m-1} c_j z_{i+j} \text{mod} 2$$

Là một phương trình tuyến tính n ẩn. Nếu n≥2n thì có m phương trình tuyến tính m ẩn có thể giải được.

Hệ m phương trình tuyến tính có thể viết dưới dạng ma trận như sau:

$$(z_{m+1}, z_{m+2}, \dots, z_{2m}) = (c_0, c_1, \dots, c_{m-1}) \begin{bmatrix} z_1 & z_2 \dots \dots . z_m \\ z_2 & z_3 \dots \dots . z_{m+1} \\ \dots & \dots \dots \dots \dots \dots \\ z_m & z_{m+1} \dots . . z_{2m-1} \end{bmatrix}$$

Nếu ma trận hệ số có nghịch đảo(theo modulo 2) thì ta nhận được nghiệm:

$$(C_0, C_1 \dots . C_{m-1}) = (z_{m+1}, z_{m+2}, \dots z_{2m}) \begin{bmatrix} Z_1 & Z_2 \dots . . & Z_m \\ Z_2 & Z_3 \dots . . & Z_{m+1} \\ \dots . . & \dots \dots . & \dots . . \\ Z_m & Z_{m+1} \dots . . & Z_{2m-1} \end{bmatrix}^{-1}$$

Trên thực tế, ma trận sẽ có nghịch đảo nếu bậc của phép đệ quy được dùng để tạo dòng khóa là m. Minh họa điều này qua một ví dụ.

Ví dụ:

Giả sử Minh thu được xâu bản mã :

101101011110010

Tương ứng với xâu bản rõ :

110100100001010

Ta cũng giả sử rằng, Minh biết dòng khóa được tạo từ một thanh ghi dịch phản hồi(LFSR) có 5 tầng. Khi đó, anh ta sẽ giải phương trình ma trận sau(nhận được 10 bít đầu tiên của dòng khóa):

$$(0,1,0,0,0) = (c_0, c_1, c_2, c_3, c_4)\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Có thể kiểm tra được rằng:

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Từ đó ta có:

$$(c_0, c_1, c_2, c_3, c_4) = (0,1,0,0,0)\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$=(1,0,0,1,0)$$

Như vậy phép đệ quy được dùng để tạo dòng khóa là :

$$z_{i+5} = z_i + z_{i+3} \text{mod} 2$$

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT M-DÃY

Trong phần này đề cập các thông tin có ích trong việc nhìn nhận về dãy giả ngẫu nhiên và ứng dụng trong bảo mật.

M dãy thường được dùng để thiết kế các bộ tạo dãy giả ngẫu nhiên cũng như trong các kĩ thuật đồng bộ, nhẩy tần… Nhiều tài liệu đã khảo sát tính tổng thể của chúng. Mặt khác khóa thuật toán dùng trong mật mã có chu kì rất lớn nhưng trong thực tế sử dụng mỗi lần mã hóa ta dùng đoạn khóa nào đó có độ dài tương ứng với bản rõ đã cho. Liệu các đoạn khóa con trong chu kì của dãy khóa có thực sự thể hiện tính địa phương tốt. Dưới đây giới thiệu một số tính chất địa phương của M dãy và tập trung khảo sát tính chất tương quan từng phần và các bộ tạo chuỗi nhị phân tuyến tính.

2.1.Bộ tạo chuỗi nhị phân tuyến tính

2.1.1. Giới thiệu

Bộ tạo thanh ghi dịch thường được sử dụng sinh ra chuỗi nhị phân đáp ứng các mục đích khác nhau. Thiết bị này nhỏ, gọn và ít chi phí vận hành và trả về các chuỗi khác nhau rất hiệu quả. Ví dụ một thanh ghi r trạng thái có $(2)^{2^{r2}}$ khả năng kết hợp các nối xuôi, ngược. Phần lớn các bộ tạo không tuyến tính, khó phân tích thường thể hiện các thuộc tính không mong muốn như là chuỗi đầu ra có chu kì ngắn. Do đó thanh ghi dịch tuyến tính thường được sử dụng hơn,với thanh ghi r trạng thái thì chu kỳ lớn nhất đạt được là 2^{r-1}. Với các hàm sinh đơn giản chỉ cần 2r bít đầu ra có thể xác định được hệ thống sinh và giá trị khởi tạo của thanh ghi r trạng thái mà không gặp khó khăn gì.

- Có thể nhận thấy rằng, một chuỗi nhị phân có chu kỳ đưa ra có thể được tạo bởi một thanh ghi dịch tuyến tính nào đó. Thành phần thanh ghi dịch nào có số trạng thái ít nhất gọi là tương đương tuyến tính của bộ tạo được sử dụng thực sự tạo chuỗi có chu kỳ nhất định. Có một số cách xác định tương đương tuyến tính, giải thuật được Massey đưa ra như là một tư tưởng chính. Chúng ta có thể hiểu độ phức tạp của chuỗi(hay bộ tạo) cũng như là độ dài của tương đương tuyến tính.
- Khả năng có thể dự đoán trước của bộ tạo tuyến tính đã được thúc đẩy các nhà thiết kế sử dụng các phép toán không tuyến tính trong bộ tạo đầu ra để

tăng hơn nữa độ phức tạp trong khi vẫn duy trì được các quan hệ có độ dài thực sự ngắn.

- Dưới đây đưa ra một số phương pháp phân tích dựa trên giả thiết trường Galois đảm bảo dự đoán kết quả phức tạp của bộ tạo từ các thao tác không tuyến tính. Ngoài ra còn cung cấp khái niệm cơ bản cho đồng bộ các thiết bị với các đặc tính mong muốn và sẵn sàng mở rộng cho các lớp bài toán lớn hơn của bộ tạo phức tạp.

2.1.2. Thể hiện trường Galois

Xem xét một chuỗi nhị phân $\{a_n\}$ trong đó a_n là thành phần thứ n của chuỗi, n=0,1,2… Nếu chuỗi được tạo bởi thanh ghi dịch tuyến tính r trạng thái với các giá trị khởi tạo ban đầu $a_0\ a_1 \dots a_{r-1}$ và phép đệ quy tuyến tính xác định:

$$a_n + \sum_{i=1}^{r} c_r\, a_{n-1} = 0, n \geq r \qquad (1)$$

trong đó $a_n$và c_i thuộc GF(2)tức nhận 2 giá trị 0 và 1.

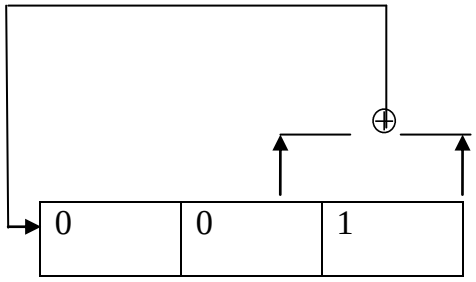
Phép đệ quy tuyến tính này có thể hiện dưới dạng một biểu thức tuyến tính khác:

$$(E^r + \sum_{i=1}^{r} c_i E^{r-1}) a_{n=0}\ , n \geq 0 \qquad (2)$$

trong đó E là phép toán dịch thực hiện trên a_n trả về trên a_{n+1}; v…..v…., $E\,a_n = a_{n+1}$Với 2 ta được đa thức đặc trưng sau:

$$x^r + \sum_{i=1}^{r} c_i\, x^{r-1} = 0 \qquad (3)$$

Chúng ta hãy cùng xem xét một thanh ghi dịch tuyến tính 3 trạng thái được minh họa như hình sau:



Thanh ghi dịch tuyến tính đơn giản

Bộ tạo được xác định bởi phép đệ quy tuyến tính sau:

$\alpha_n + \alpha_{n-2} + \alpha_{n-3}$,n>=3 và các giá trị khởi tạo ban đầu $\alpha_0 = 1, \alpha_1 = 0, \alpha_2 = 0$ chuỗi thỏa mãn đa thức đặc trưng và tối giản sau :

$$x^3 + x + 1 = 0 \quad (5)$$

Gọi α là nghiệm của (5) ta có:

$$\alpha^3 + \alpha + 1 = 0 \quad (6)$$

Từ (6)chúng ta có thể tạo ra các mũ của α

$\alpha = \alpha$

$\alpha^2 = \alpha^2$

$\alpha^3 = \alpha + 1$

$\alpha^4 = \alpha^2 + \alpha$

$\alpha^5 = \alpha^2 + \alpha + 1$

$\alpha^6 = \alpha^2 + 1$

$\alpha^7 = 1$

Cùng với 0 tạo nên 8 thành phần của GF(2^3);

Nhận thấy bất kỳ đa thức f(x) trên GF(2) thì $[f(x)]^2 = f(x^2)$

Vì vậy ta có dạng:

$$a_n = A1x^n + A2\alpha^{2n} + A3(\alpha^2 + \alpha)^n$$

Từ giá trị ban đầu a0, a1, a2 chúng ta có thể xác định A1=A2=13=1, vậy:

$$a_n = \alpha^n + \alpha^{2n} + (\alpha^2 + \alpha)^n$$

Do nên chuỗi có chu kỳ 7 và kết quả đầu ra là {1,0,0,1,0,1,1}

Từ trên có thể đưa ra một giả thiết sau:

Nếu P(x) là một đa thức tối giản bậc d trên trường GF(2) thì trường mở rộng GF(2^d) được tạo bởi việc liên kết nghiệm α của đa thức p(x) tới GF(2) chứa d nghiệm khác biệt của p(x) và những nghiệm này được đưa dưới dạng:

α^{2k} với k=0,1,2…d-1.

Tóm lại, LFSR có thể được thực hiện bằng nghiệm của đa thức đặc trưng trên trường Galois, qua đó cho ta xây dựng các hệ thống không tuyến tính có thể kiểm soát được để tăng độ phức tạp của chuỗi kết quả đầu ra và cung cấp khả năng đồng bộ của các bộ tạo phức tạp.

2.2 KHÁI NIỆM M-DÃY

2.2.1. Định nghĩa:

Định nghĩa m-dãy (tiếng Anh là Maximum Sequence): Cho p(x) là một đa thức nguyên thủy cấp n trên trường GF(2). Khi đó, bộ hai thành phần:(n,p(x)) được gọi là bộ ghi dịch có độ dài cực đại. Dãy bít ngẫu nhiên do bộ ghi dịch đó tạo ra được gọi là m- dãy với chu kỳ của nó là $N = 2^n - 1$.
Ví dụ: Cho p(x) = x^4+ x+1.. Do đó theo định nghĩa bộ hai (4 , x^4+x+1) là bộ ghi dịch có độ dài cực đại và chu kỳ của dày đầu ra đó là: 2^4-1 = 16-1 =15. Còn một đa thức p(x) cấp n được gọi là đa thức nguyên thủy (Primitive Polynomial) nếu nó không thể được phân tích thành tích các đa thức cố cấp nhỏ hơn n (khác 1). Bậc cao nhất của một đa thức được gọi là cấp của nó.

Với các thanh ghi dịch phản hồi tuyến tính (LFSR) có cấp L, ta có định lý sau:

2.2.2.Định lý 1:

Một thiết bị sinh tuyến tính của một bộ nhớ với kích thước cho trước sẽ sinh ra một dãy các phần tử trong trường GF(q) có chu kỳ lớn nhất nếu và chỉ nếu đa thức đặc trưng tương ứng của nó là đa thức nguyên thủy trên GF(q). Ta không chứng minh định lý này.

Hai dãy $\{b_n\}$ và $\{b_n'\}$, mỗi dãy có chu kỳ N được gọi là tương đương tuần hoàn nếu có tồn tại một số nguyên không âm τ sao cho $b_{n+\tau} = b_n'$ $\forall n = 1,2,\ldots$ Hai dãy không có tính chất đó được gọi là không tương đương tuần hoàn hoặc gọi là phân biệt tuần hoàn (Cyclically distinctness). Từ đó các m-dãy là không tương đương tuần hoàn nếu và chỉ nếu số các m-dãy không tương đương tuần hoàn trên GF(q) với bao tuyến tính L bằng số $N_p(L)$ các đa thức nguyên thủy cấp L.

Để xác định số $N_p(L)$, ta đưa vào khái niệm sau đây gọi là hàm phi – Ơle được ký hiệu là Φ(n). Ta có định nghĩa sau đây:

Cho n là một số tự nhiên tùy ý, Ta ký hiệu (x,n) là ước số chung lớn nhất của 2 số x và n. Khi đó hàm phi – Ơle Φ(n) là số tất cả x: $1 \leq x < n$ mà nguyên tố với n, tức là:

$$\Phi(n) = \#\{1 \leq x < n : (x,n) = 1\}.$$

Hàm Φ(n) có nhiều ý nghĩa trong khoa học kỹ thuật và thực tiễn. Sau đây là một vài tính chất cơ bản của hàm Φ(n):

Tính chất 1: nếu n là số nguyên tố thì $\Phi(n) = n-1$

Tính chất 2: nếu n = p.q là tích của 2 số nguyên tố p và q thì $\Phi(n) = \Phi(p.q) = (p-1).(q-1)$.

Tính chất 3: nếu n = pk với p là số nguyên tố, còn k là số nguyên dương tùy ý thì:

$$\Phi(n) = \Phi(p^k) = p^{k-1}(p-1).$$

Tính chất 4: nếu n = M.Lvới M và L là 2 số nguyên dương tùy ý và nguyên tố cùng nhau thì:

$$\Phi(M.L) = \Phi(M)\Phi(L).$$

Tính chất 5: nếu $n = p_1^{l_1}.p_2^{l_2}\ldots.p_k^{l_k}$ với $p_1, p_2, \ldots..p_k$, là những số nguyên tố khác nhau từng đôi một, còn $l_1, l_2, \ldots l_k$ là những số nguyên không âm, khi đó:

$$\Phi(n) = \prod_{i=1}^{k} p_i^{l_i-1}(p_i - 1)$$

Bây giờ ta có định lý sau:

2.2.3. Định lý 2:

Số $N_p(L)$ tất cả các đa thức nguyên thủy có cấp L trên trường GF(q) được cho bởi hệ thức:

$$N_p(L) = \frac{\phi(q^L - 1)}{L}$$

Như vậy, số $N_p(L)$ của m-dãy không tương đương tuần hoàn với L mở rộng tuyến tính được cho trong bảng sau (q = 2)

L	$N_p(L)$	L	$N_p(L)$
2	1	16	2048
3	2	17	7710
4	2	18	7776
5	6	19	27594
6	6	20	24000
7	18	21	84672
8	16	22	120032
9	48	23	356962
10	60	24	276480
11	176	25	1296000
12	144	26	1719900
13	630	27	4202496
14	756	28	4741632
15	1800	29	18407808

Bảng này nhằm minh họa số các $N_p(L)$ đối với m-dãy nhị phân tương ứng L= 2 ÷ 29 và q = 2.

2.3. Tính chất địa phương của M dãy

Xét m dãy $(a_i) = (a_1 a_2 \dots \dots \dots a_{n-1})$, có chu kì N = $2^r - 1, a_i \in GF(2)$, với đa thức đặc trưng tương ứng f(x) bậc r.

Các tính chất tổng thể

Mệnh đề 1:

M dãy có tính chất toàn cục như sau.

Chu kì cực đại: N = $2^r - 1$

Có phân bố gần đều cho M tupelo với 0 <M<=r

Có độ phức tạp tuyến tính là r, tức là 2r bits lien tiếp sẽ xác định duy nhất f(x)

Hàm tự tương quan có 2 mức:

$$C(\tau)=\begin{cases}1, \tau = k-N \\ -\frac{1}{N}, \tau \quad k-N, k \in Z\end{cases}$$

Gỉa sử đặt:

$$M_m=((a_{1,}a_{i+1,}\dots..a_{i+m-1}));i=0\dots N-1;M>0)$$

Từ các tính chất tổng thể trên ta thấy:

Độ phức tạp $\gamma(m)$ <=r,với m$\in M_m$bất kỳ.

Nếu 0<M<=r,M_m có phân bos gần đều cho M tupple.

Nếu M>r,M_m chứ đúng N bộ M tuple khác nhau.

Nếu M=N,hàm tương quan trên M_m sẽ chỉ nhận 2 giá trị .

Với M<N,hàm này chưa được khảo sát

- Sau đây chúng ta sẽ chính xác hóa khái niệm và khảo sát phân bố giá trị tương quan trên M_m với r<M<N

2.4.Tương quan địa phương của M dẫy

2.4.1.Khái niệm

Trở lại xét M_m= (($a_{i,}, a_{i+1}, \dots . a_{i+m-1}$), i=0..N-1),M>0

Rõ ràng {M_m} = 2^r-1 = N

Giả sử m1,m2$\in M_m$=>m*=m1+m2$\in M_m$(tính chất minh của m dẫy) m1#m2

Hàm tương quan giữa m1,m2 là: C(m1,m2) = $\frac{M-2wt(m^*)}{M}$, ở đây wt(m*)là trọng số Heeming của vector m*.

Khi m1,m2 lấy ngẫu nhiên đều trên Mm, thì m* cũng xuất hiện ngẫu nhiên đều trên Mm, vì vậy ta có thể xét biến ngẫu nhiên wt(m*)khi m* lấy ngẫu nhiên đều trên Mm. Rõ ràng phân bố của biến ngẫu nhiên này cho ta hình ảnh đặc trưng tính chất tương quan địa phương trên các đoạn con M bit của M dẫy đã cho.

Với 0<M<=r, phân bố này có thể tìm được từ phân bố của M tuple. Tuy nhiên với r<M<N việc tìm hàm phân bố của wt(m*)rất khó khăn. Do vậy phần sau

chỉ giới thiệu phương pháp tìm một số mô men đầu tiên đặc trưng chủ yếu cho hình ảnh hình học của phân bố này.

2.4.2 Các mô men của phân bổ giá trị tương quan

Đặt wi = wt($a_1, a_{i+1}.....a_{i+m-1}$) = $\sum_{n=0}^{M-1} ai + n$

Gọi Aw là số các M tupple trong Mm có wi = w, nếu M>r thì w∈[1,M].Nếu gọi W^p, W_c^p tương ứng là các mômen bậc p và mô men trung tâm bậc p của biến ngẫu nhiên wt(m*), ta có:

$$W_c^p = \frac{1}{N}\sum_{i=0}^{N-1} w i_i^p = \frac{1}{N}\sum_{w=1}^{M} w^p \text{Aw}$$

$$W_c^p = \frac{1}{N}\sum \ (w_i - w^1 \ ))$$

Công thức tính moomen trung tâm bậc 1,2,3,4 cụ thể như sau:

Mệnh đề 2: Với gải thiết đã nêu(với M>r) ta có:

w_c^1=0.

w_2^c=$\frac{M}{4}$(1-$\frac{M-1}{N}$).

$w_3^c = \frac{-M^3}{8N} + \frac{3(N+1)}{4N} B_3$(M).

$w_c^4 = \frac{3M^2-2M}{16} - \frac{M(M-1)(M-2)}{16N} + \frac{3(N+1)}{2N} B_{4(M).}$

trong đó:$B_3(M)$ là các bội ba (i,j,k) với 0<= i,j,k<=M-1 sao cho:

$a_{n+1} \oplus a_{n+j} = a_{n+k}, \forall n = \overline{0, N-1}$

$B_4(M)$ là các bộ 4(i_1, i_2, i_3, i_4) mà $0 \le i_1 \le i_2 \le i_3 \le i_4 \le M-1$ sao cho:

$a_{n+i1} \oplus a_{n+i2} \oplus a_{n+i3} = a_{n+i4}, \forall n = \overline{0, N-1}$

Nhận xét về tính tương quan địa phương của M dẫy.

Bảng sau đây cho ta một so sánh phân bố giá trị tương quan của m dẫy với dẫy ngẫu nhiên đúng, với giả thiết r<M<<N.

Moment rung tâm bậc n	Dẫy ngẫu nhiên	M dẫy
1	0	0
2	M/4	M/4
3	0	$\frac{3}{4}B_3(M)$
4	$\frac{3M^2-2M}{16}$	$\frac{3M^2-2M}{16}+\frac{3}{2}B_4(M)$

Từ kết quả trên ta thấy khác tính chất tương quan toàn phần, tính chất tương quan địa phương của m dẫy là hoàn toàn phân biệt bơi m dẫy khác nhau(dù có cùng chu kì), thể hiện ở chỗ m dẫy sẽ có bộ ($w_c^3, w_{c_1^1}^4 \ldots$)đặc trưng cuả riêng nó. Chẳng hạn với:

f(x)=1 + $x^5 + x^{23}$

có w_c^3= -1338

g(x)=1+x

$x^3 + x^5 + x^6 + x^7 + x^8 + x^{12} + x^{15} + x^{16} + x^{18} + x^{20} + x^{21} + x^{22} + x^{23}$

có w_c^3= -1245

với r = 23,M = 500

Ta cũng thấy rằng m dẫy có thể xem là có tương quan ngẫu nhiên địa phương phạm vi M nếu có $B_3 = B_4 = 0$. Và hiển nhiên trong số các m dẫy cùng chu kỳ, tuy cùng tính chất tổng thể nhưng m dẫy nào có $B_3 = B_4 = 0$ sẽ có dáng điệu ngẫu nhiên địa phương hơn.

Qua khảo sát trên đây ta thấy mỗi m dãy đều có những thông tin rất riêng biệt về chúng, đó là những tính chất địa phương và đặc biệt là tính chất tương quan địa phương. Những thông tin này chắc chắn sẽ rất có ích cho việc lựa chọn, thiết kế và sử dụng m dãy độc lập cũng như trong các mạch tổ hợp tạo khóa nhất là cơ sở tạo khóa giả ngẫu nhiên và sẽ được ứng dụng trong bảo mật thông tin như sẽ trình bày như sau.

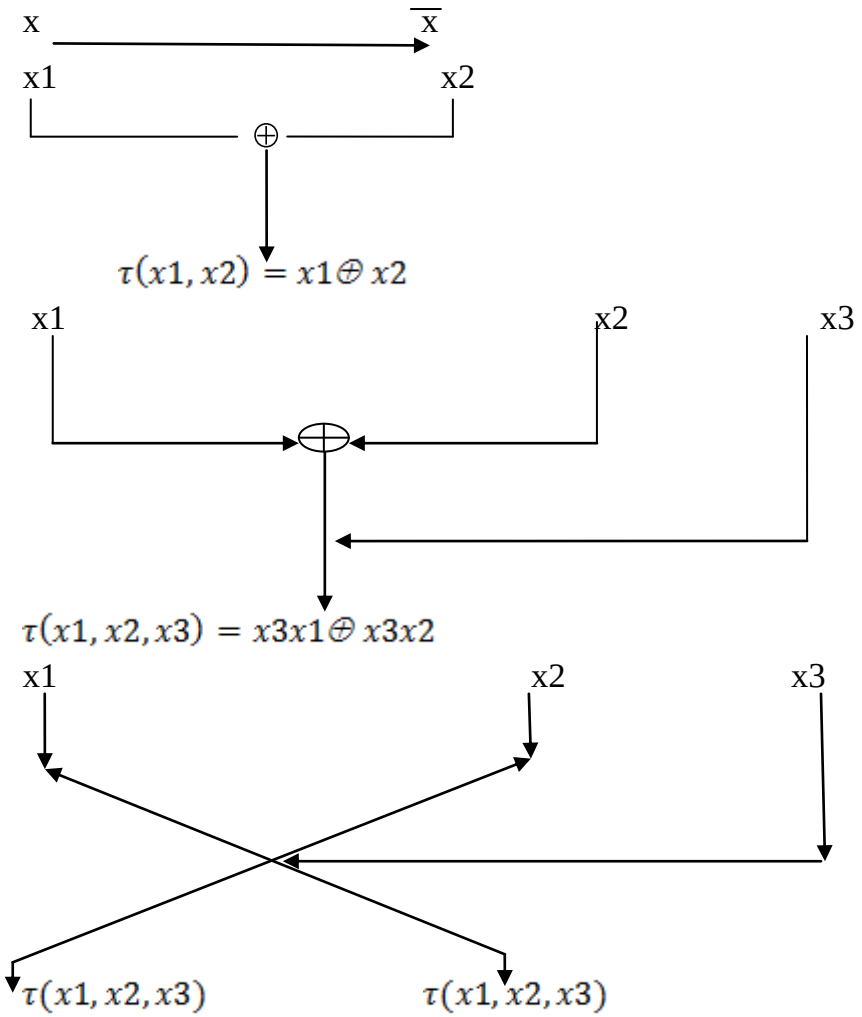
Nguồn nhiễu ngẫu nhiên được tạo bởi một quá trình vật lý nào đó, ở các máy mã thế hệ điệ tử trên; thường sử dụng quá trình biến động gẫu nhiên của nhiệt nhiễu ở các linh kiện điện tử như diod, transistor…..

Khóa ngoài thường là một vector nhị phân nào đó, có độ dài hữu hạn, trạng thái này là ngẫu nhiên. Khóa ngoài có thể đưa vào dưới dạng băng đục lỗ như E3(băng này được sản xuất từ máy sản xuất khóa); hoặc là trạng thái ngẫu nhiên ban đầu của thanh ghi dịch như máy mã thoại E10.

Khối phức tạp hóa thể hiện chức năng của một hàm f nào đó, có tác dụng kết hợp và biến đổi giữa nguồn nhiễu(X), khóa ngoài K và các thanh ghi dịch:

τ=f(K,X….)

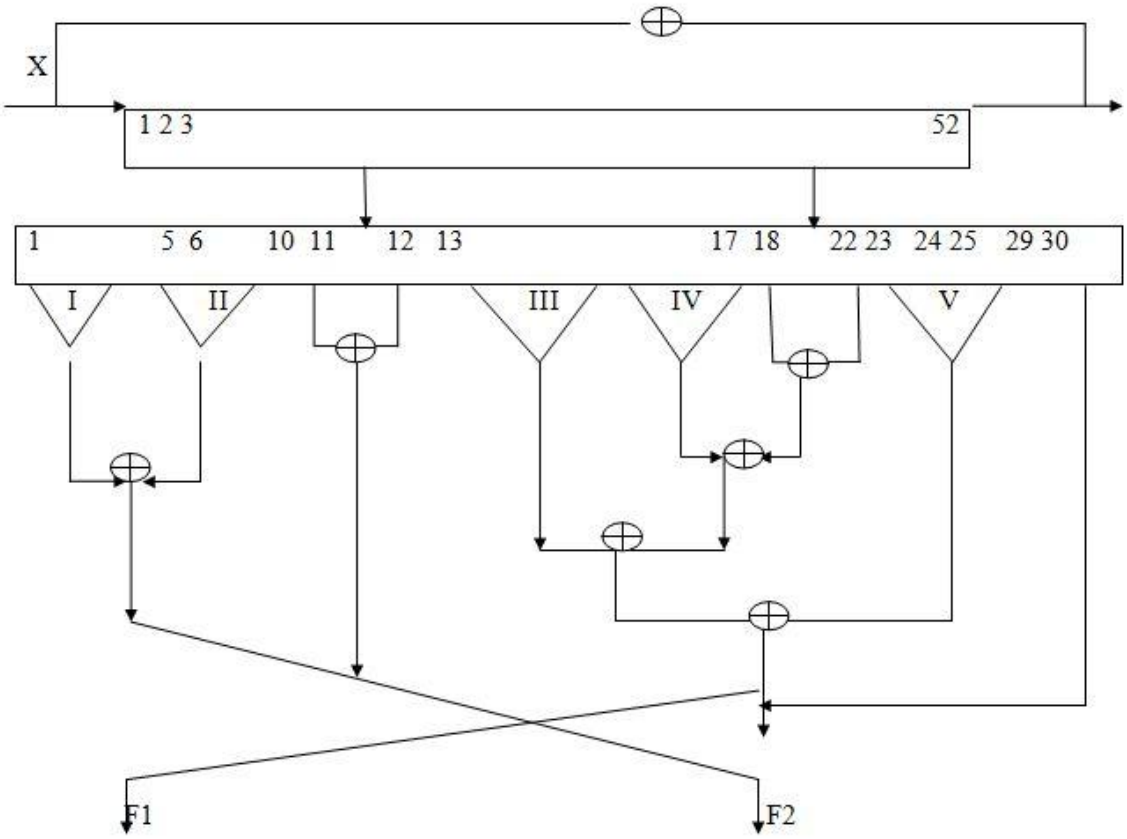
Hàm phức tạp hóa thể hiện theo các hàm Boolean cơ bản và kết hợp với nhau được cứng hóa theo các mạch logic sau:



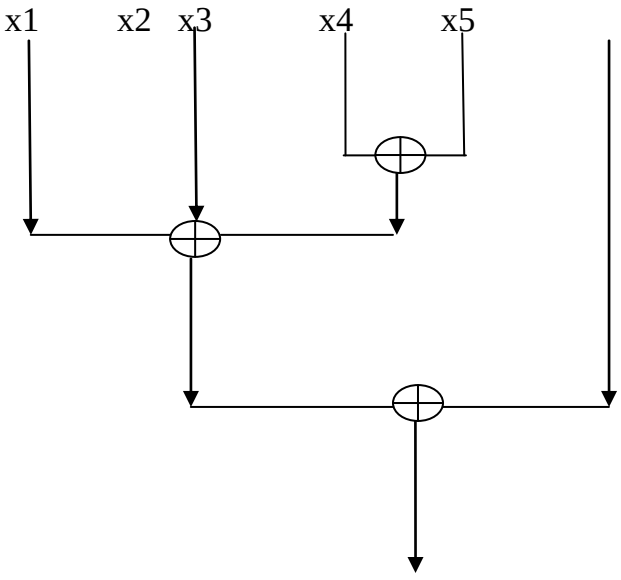
$$\begin{cases} \tau 1 = \overline{x3}x1 \oplus x3\overline{x2} \\ \tau 2 = \overline{x3}x2 \oplus x3\overline{x1} \end{cases}$$

2.5.Thuật toán tạo khóa của máy mã thoại E10

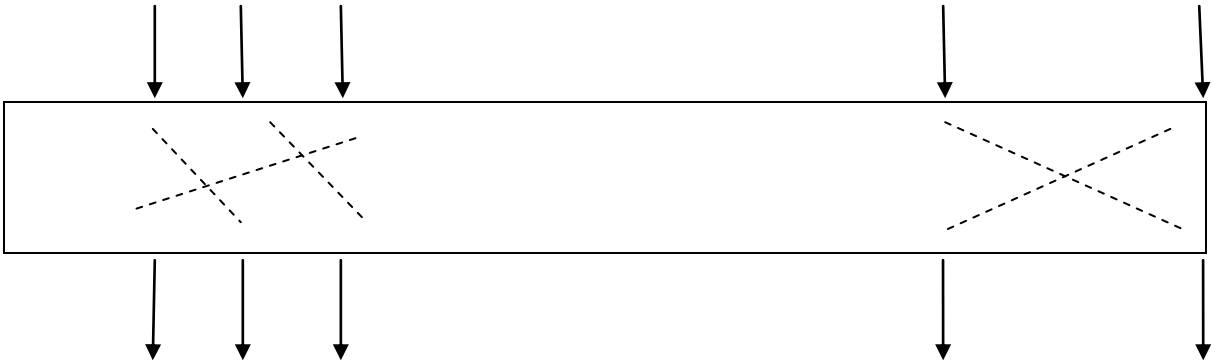
Máy mã thoại E10 tạo khóa làm việc theo sơ đồ sau:



Khối i = $\overline{I,V}$ có cấu trúc:



Thanh ghi dịch 52 bits để tạo ra dãy gamma ban đầu, có trạng thái đầu tiên là từ nguồn ngẫu x. Bộ đảo mạch có 30 đầu vào và 30 đầu ra, 30 đầu vào lấy từ 30 bits của một trạng thái ghi dịch, cụ thể theo hình vẽ sau:



Cách nối từ mỗi đầu vào và 30 đầu ra tương ứng 1:1 bất kỳ, mỗi cách nối 30 đầu vào với 30 đầu ra không cố định thay đổi theo thời gian định trước(Bộ đảo mạch có tác dụng tạo khóa ngoài).

Khối phức tạp hóa thực hiện nhờ sự kết hợp logic giữa các hàm cơ bản ở trên. Do đó có thể mô phỏng thuật toán tạo khóa như sau:

Không giảm tính tổng quát, có thể đánh số lại trạng thái thanh ghi dịch là:x1,x2…x52.

Khi đó 30 bits đầu vào của bộ đảo mạch giả sử là :x1,x2,x3…x30.

Theo sơ đồ mật mã ta có:

dk1=x3 ⊕ x4

dk2=x8 ⊕ x9

y1=$\overline{dk1}$x1⊕ dk1$\overline{x2}$

y2=$\overline{dk2}$x6 ⊕ dk2$\overline{x7}$

t1=x5 ⊕ y1

t2=x10 ⊕ y2

dk4=x20 ⊕ x21

dk5=x27 ⊕ x28

y4=$\overline{dk4x}$18 ⊕ dk$\overline{4x19}$

y5=$\overline{dk5}$x25 ⊕ dk5$\overline{x26}$

t4=x22 ⊕ y4

t5=x29 ⊕ y5

τ1=x23 ⊕ x24

dk3=x15 ⊕ x16

t’4=$\overline{\tau 1t4}$ ⊕ $\tau 1\overline{t4}$

y3=$\overline{dk3}$x13 ⊕ dk3$\overline{x14}$

t”4=t3⊕ t’4

t3=x17 ⊕ y3

$\tau_2 = x11 \oplus x12$

f1 = $\overline{t1}t2 \oplus t1\overline{t2}$

f2 = $\overline{x30}t''4 \oplus x30\,\overline{t5}$

Cuối cùng bộ giao thực hiện:

$$\begin{cases} F1 = \overline{\tau 2} f1 \oplus \tau 2 \overline{f2} \\ F2 = \overline{\tau 2} f2 \oplus \tau 2 \overline{f1} \end{cases}$$

Như vậy,cứ mỗi nhịp tạo ra được 2 bits khóa. Từ phân tích trên ta rút ra các kết luận sau đây:

Chu kỳ của khóa là $2^{52} - 1$, vì bộ ghi dịch tuyến tính với đa thức sinh là nguyên thủy trên trường GF(2)sẽ cho ta m dẫy có chu kỳ cực đại.

M dẫy trên, trong kỹ thuật máy mã thoại E10 ngoài tác dụng tạo ra chu kì còn có tác dụng làm chức năng đồng bộ giữa máy phát và máy thu.

Dung lượng khóa ngoài hiển nhiên là 30!(Bộ đảo mạch làm chức năng hoán vị 30 bits ngẫu nhiên của thanh ghi dịch)

Phân tích độ an toàn mật mã của thuật này, chu kỳ dẫy sinh khóa không lớn so với độ lớn tối thiểu của khóa từ các dẫy($>2^{100}$), song khóa không tạo theo mô hình trên có độ bảo mật cao bởi lẽ, độ phức tạp của dẫy khóa phụ thuộc vào thuật toán phức tạp hóa f và quan trọng là f phải được giữ hoàn toàn bí mật.

Cấu trúc tạo khóa của máy mã trên theo 1 nguyên lý khá chặt chẽ, có thể ứng dụng trên nguyên tắc này để mô phỏng và xây dựng các phần mềm bảo mật trên máy tính PC có độ bảo mật theo ý muốn. Việc chương trình hóa trên máy tính khá dễ dàng thuật toán trên bởi lẽ người lập trình tăng chu kỳ hoặc dung lượng khóa ngoài bằng chương trình phần mềm khá dễ hơn nhiều so với công nghệ phần cứng.

2.6. Tính chất thống kê toán của m- dãy

Các đa thức nguyên thuỷ cung cấp cho ta sự kết nối giữa các LFSR để tạo ra các m-dãy với chu kỳ cực đại có thể được. Tuy nhiên, một câu hỏi được đặt ra mà chúng ta cần quan tâm là liệu trong khoảng chu kỳ của m-dãy, dãy đó có độc lập và phân phối xác suất đều hay không? Các tính chất thống kê của m-dãy sau đây trả lời cho câu hỏi vừa nêu:

Tính chất 1: Xét dãy $\{b_n\}$ của m-dãy. Ta ký hiệu N_b là số lần xuất hiện ký tự (signal) b trong dãy $\{b_n\}$. N = 1,2,...,q^L-1 và N_0 là số lần xuất hiện ký tự 0 trong dãy $\{b_n\}$. Khi đó:

$$N_b = N_0 + 1 \quad \forall b \neq 0$$

Tính chất 2: Gọi $\{b_n\}$ là một m-dãy gồm các phần tử trong trường GF(q) với bao (span) tuyến tính L, khi đó tần số trùng lặp bộ J (J = 1,2,3,…) N[J] trong dãy $\{b_n\}$ được cho bởi công thức sau:

$$N_a[J]= \begin{cases} q^{L-1} \text{ nếu } a \neq 0 \ 1 \leq j \leq L \\ Q^{L-1} - 1 \text{ nếu } a=0 \ 1 \leq j \leq L \\ 0 \text{ hoặc } 1 \text{ nếu } L < j \end{cases}$$

Trong đó **a** = $(a_1, \ldots, a_j,)$.

Kết quả này có nghĩa là xác suất xuất hiện bộ J tín hiệu (ký tự) trong các phép thử độc lập, mỗi phép thử có q kết cục đồng khả năng thì bằng:

$$\frac{N_0}{N} = \frac{q^{L-j} - 1}{q^L - 1} \qquad (1)$$

Trong đó N là độ lớn của chu kỳ của m-dãy.

Rõ ràng rằng $\frac{N_0}{N}$ tiến tới q^{-j} khi L tăng.

Tính chất 3: Thông thường một thiết bị tạo m-dãy trên trường GF(q) được ứng dụng để sinh ra các dãy giả ngẫu nhiên gồm các ký tự 0 và $q^j - 1$ nếu coi bộ J thứ n ($b_n, b_{n+1}, \ldots\ldots, b_{n+j-1}$) trong m-dãy {bn} là một khai triển theo cơ số q của số nguyên thứ n trong dãy số nguyên do LFSR tạo ra, tức là:

$$I_n = \sum_{i=c}^{j-1} b_{n+i} q^i$$

Để đo chất lượng ngẫu nhiên của dãy $\{i_n\}$, ta dùng khoảng cách Hamming được định nghĩa như sau:

Giả sử x = $(x_1, x_2, \ldots, x_n)$, y = $(y_1, y_2, \ldots, y_n)$ là 2 vectơ cùng độ dài n, khi đó khoảng cách Hamming giữa x và y là :

$$H(x, y) = \sum_{i=1}^{N} h(x_i, y_i),$$

trong đó :

$$H(x_i,y_i) = \begin{cases} 0 \text{ nếu } x_i = y_i \\ 1 \quad \text{nếu } x_i \neq y_i \end{cases}$$

i = 1, 2,…, n.

Ta có tính chất 4 như sau:

Tính chất 4: Giả sử $\{b_n\}$ là một m-dãy trên trường GF(q) với mở rộng tuyến tính L và giả sử là dãy được cho ở (1). Ta ký hiệu i_k là một chu kỳ của dãy $\{i_n\}$ mà bắt đầu với số nguyên thứ k. Khi đó với $t \neq 0 \bmod(q^L - 1)$ thì khoảng cách Hamming giữa i_1 và i_{1+t} được cho bởi hệ thức:

$$H(i_1, i_{1+t}) = q^L(1 - q^{-J}) \qquad (2)$$

(J là độ dài của dãy chu kỳ)

Từ tính chất 4 ta có định lý sau:

Định lý 3: Cho $\{a_n\}$ là một dãy gồm các ký hiệu từ bảng chữ cái có chu kỳ N, còn **a** ký hiệu là bộ N các ký hiệu kề nhau được lấy từ dãy$\{a_n\}$ mà khởi đầu tại ký hiệu thứ J. Giả sử

$$H_{min} = \min_{0<t<N} H(\mathbf{a}_1, \mathbf{a}_{1+t}), \text{ và} \qquad (3)$$

Ta xác định các số nguyên α, β sao cho $N = \alpha|A| + \beta \quad 0 \leq \beta < |A|$ (4)

Khi đó:

$$H_{min} \leq N - \frac{\alpha}{N-1}(N + \beta - |A|) \qquad (5)$$

Trường hợp $N = q^L - 1$ và $|A| = q^{-J}$ (6)

Thì $\beta = (1 - q^{-J})$, $\alpha = q^{L-1} - 1$ (7)

Vậy (5) cho ta kết quả

$$H_{min} \leq q^L (1 - q^{-J}) \qquad (8)$$

Từ tính chất 4 ta suy ra rằng dãy $\{i_n\}$ thỏa mãn các yêu cầu của định lý. Do đó dãy $\{b_n\}$ của m-dãy đạt yêu cầu là phân bố đều có khoảng cách cực đại của m-dãy.

CHƯƠNG 3: ỨNG DỤNG LÝ THUYẾT M - DÃY VÀO VIỆC MÃ HOÁ BỞI HỆ MÃ VIGENERE

Thuật toán:

Tạo ra m- dãy

Mã hóa:

Input: - $m_1m_2\ldots m_k$

-bản rõ: $x_1x_2\ldots x_m$

Output: bản mã $y = (m_i + x_i) \bmod 26$; $i = \overline{1,m}$

a = 0, b=1, …,z=25

Giải mã:

Input: -bản mã y

$-m_1m_2\ldots m_k$

Output: Bản rõ $x = x_1x_2\ldots x_m$

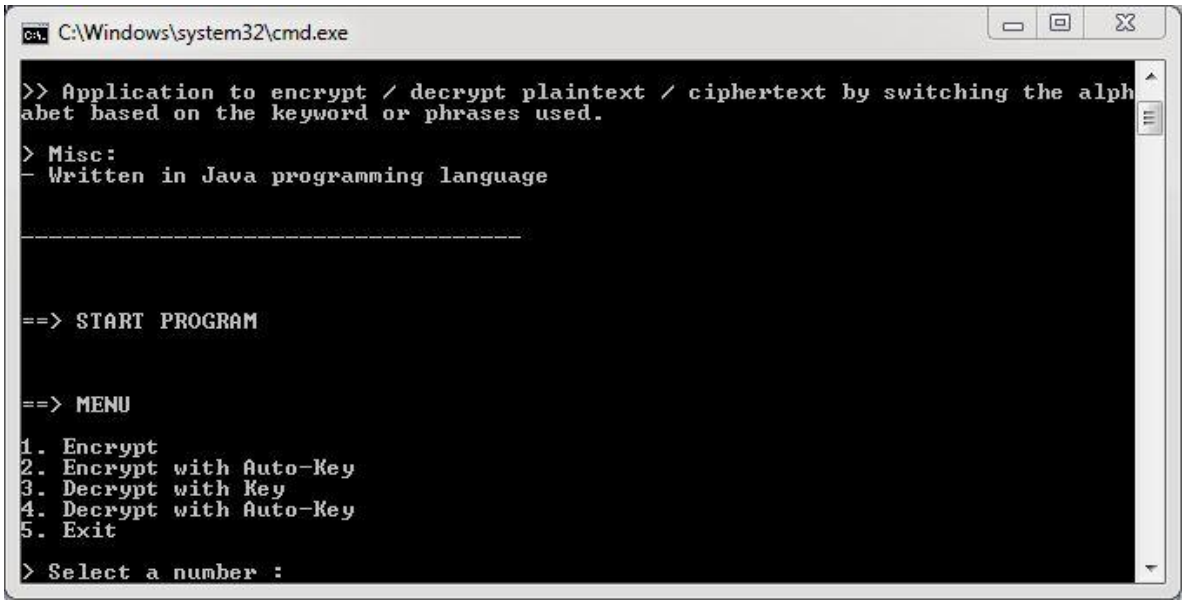
1)Đổi bản mã và khóa→các dãy số $\overline{0,25}$

2)$x_i = (y_i - m_i) \bmod 26$

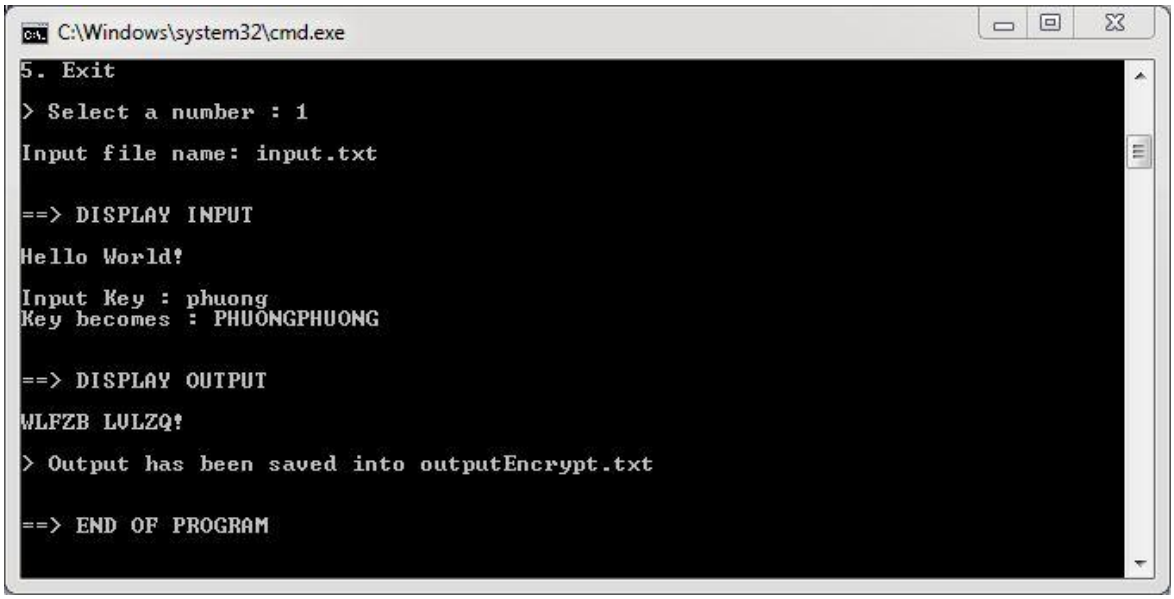
MÔ TẢ CHƯƠNG TRÌNH

Gồm có 4 phần :

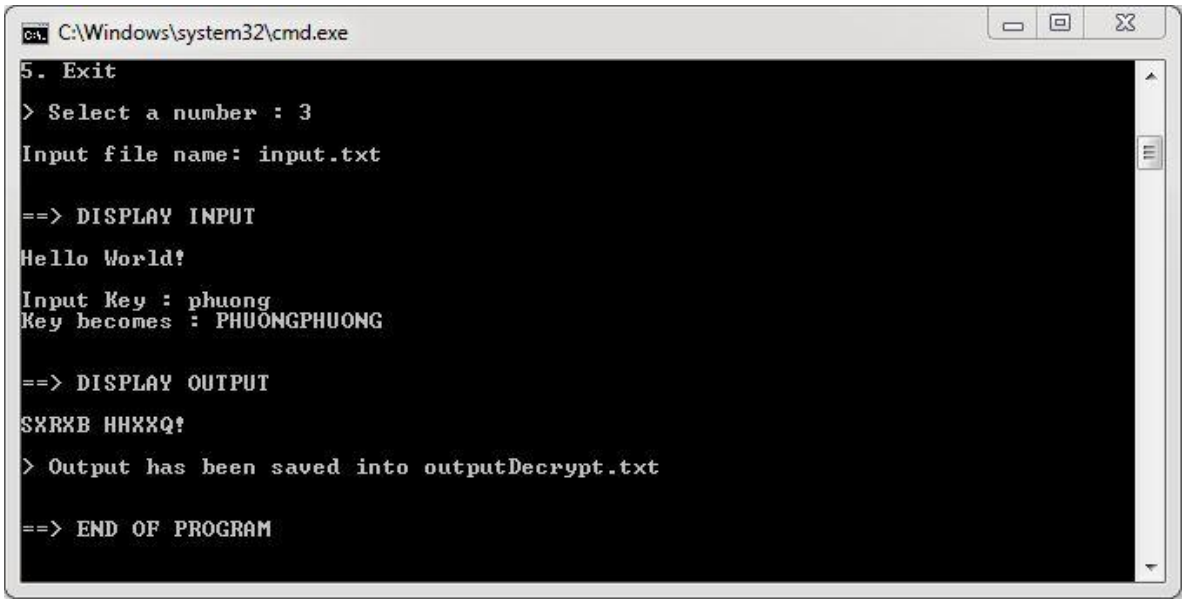
Phần 1: Menu chương trình :



Phần 2: Mã hóa một bản rõ khi có khóa cho trước:



Phần 3: Giải mã một bản rõ khi có khóa cho trước:



Phần 4: Thoát khỏi chương trình.

KẾT LUẬN

Từ kết quả có tính lý thuyết vừa nêu trên, chúng ta có thể thiết lập một bộ các thanh ghi dịch tuyến tính để tạo ra các dãy giả ngẫu nhiên có chu kỳ cực đại. Thật vậy, giả sử ta muốn tạo ra m thanh ghi dịch phản hồi tuyến tính, giả sử thanh ghi dịch thứ i có n_i ô (cell) chứa thông tin ban đầu, ở đây i = 1, 2,…, m. Để m thanh ghi dịch n_1,…, n_m tạo ra được một dãy $\{i_n\}$ có chu kỳ cực đại, theo lý thuyết trên, điều kiện ắt có và đủ là n_i, n_j phải nguyên tố cùng nhau, tức là ƯSCLN(n_i,n_j) = 1 $\forall i,j$, $i \neq j$. Trong đó có L ô chứa thông tin phản hồi. Việc mở rộng cho nội dung bài báo có thể thực hiện trên trường số phức.

Vấn đề bảo mật là một vấn đề lớn, nên trong khuôn khổ khóa luận này chỉ hy vọng đưa ra được một cách nhìn tổng quát và giải pháp cho một vấn đề của mã hóa cổ điển được sử dụng trong các thiết bị đã được cứng hóa trong các máy mã Liên xô (cũ). Nội dung chính của khóa luận là phần giới thiệu tóm tắt về cơ sở mã hóa đối xứng, cơ sở toán M- dãy cùng ứng dụng trong việc tạo khóa giả ngẫu nhiên nhằm thực hiện việc bảo mật và bảo đảm tính an toàn và có những tính năng mạnh hơn so với những thiết bị mã hóa E10. Song do lượng thời gian còn hạn chế nên đã không thể lột tả kết quả đạt được còn một số hạn chế cần phải tiếp tục hoàn thiện.

Tuy nhiên việc tiếp cận vấn đề bảo mật và giải quyết bài toán tạo khóa giả ngẫu nhiên ứng dụng trong bảo mật đã tạo cho bản thân một phương pháp nghiên cứu, vốn kiến thức hết sức quý báu. Trong tương lai nếu có điều kiện công tác và học tập thì đó là nền tảng tốt để em hoàn thiện hệ thống chương trình nhằm đem lại hiệu quả thiết thực hơn. Em tin rằng, trong điều kiện hiện nay với sự phát triển như vũ bão của Internet thì nguồn thông tin trên mạng rất cần bảo mật đảm bảo an ninh toàn thông tin.

Đối với nghiệp vụ ngành, trong phạm vi khóa luận này, em cố gắng đưa ra một giải pháp mã hóa thông tin nhằm truyền nhận thông tin bí mật trong ngành đảm bảo độ an toàn và tránh bị ngăn chặn phát hiện nội dung a=bản mã, những bản có nội dung ảnh hưởng trực tiếp đến nghiệp vụ ngành, an ninh quốc gia.

Một lần nữa em xin chân thành cảm ơn thầyTS.Hồ Văn Canh đã trực tiếp hướng dẫn cùng toàn thể thầy cô đã tận tình chỉ bảo, tạo điều kiện cho em hoàn thành khóa luận này.

CÁC TÀI LIỆU THAM KHẢO

[1] A. Lempel and H. Greenberger, " Families of sequences with optimal Hamming correlation properties", IEEE Trans. Inform. Theory, IT-15, pp,90-95. January 1974.

[2] D. E. Knut "The Art of Computer Progamming", Vol.2, chapter 3. Massachusets. Menlo Park, California. London. Mills, Ontario.

[3] H. Fukumasa, R. Kohno, and H. Imai, "Design of pseudonoise sequences with good odd and even correlation properties", Proc. IEEE Second International Symposium on Spread Spectrum Techniques, Yokohama, Japan, Nov. 29- Dec. 2, 1992, pp.139-142.

[4] Marvin K. Simon, Jim K. Omura, Robert A. Scholtz, Barry K. Levitt, " Spread Spectrum Communications Handbook", New York, London, New Delhi, Singapore, Sydney, Tokyo, 1994.

[5] M. Antweiler and L. Bomer, " complex sequences over GF(p^n) with a two-level autocorrelation function and large linear span", IEEE Trans. Inform. Theory, IT- 38, pp.120-130, January 1992.

[6] S.A. Fredricson, " Pseudo-randomness properties of binary shift register sequences", IEEE Trans. Inform. Theory, pp. 115-120, January 1995.