Nghiên cứu xây dựng thuật toán mã khối phục vụ việc bảo mật thông tin chiến lược : Luận văn ThS. Công nghệ thông tin: 1.01.1 / Nguyễn Vũ Hoàng ; Nghd. : TS. Hồ Văn Canh

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

AES	Advanced Encryption Standard.
CCITT	The Consultative Committee for International Telegraph and Telephone.
DES	Data Encryption Standard.
IDEA	International Data Encryption Algorithm.
O.T.P	One Time Pad.
NBS	National Bureau of Standards.
PGP	Pretty Good Privacy.

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

DANH MỤC CÁC BẢNG

MỞ ĐẦU

Vấn đề an ninh/an toàn mạng truyền thông nói chung, mạng máy tính nói riêng luôn luôn phát triển song hành với sự phát triển của công nghệ mạng và nhu cầu thông tin liên lạc. Ở đây ta hiểu sự an ninh/an toàn mạng bao gồm việc bảo đảm an ninh/an toàn về mặt vật lý và về toàn bộ cơ sở dữ liệu chứa thông tin trên mạng, đặc biệt là những thông tin nhạy cảm liên quan đến an ninh quốc phòng, ngoại giao, tình báo và kinh tế… Trong phạm vi khả năng của mình, em chọn đề tài: "***Nghiên cứu và xây dựng thuật toán mã khối phục vụ việc bảo mật thông tin chiến lược***" khi chúng được lưu thông trên mạng. "***Mạng***" ở đây không nhất thiết phải là mạng máy tính mà có thể là mạng truyền thông bất kỳ.

Về mật mã khối, hiện đang được quan tâm ở nhiều nước như Mỹ, Nga, Anh, Trung Quốc,… và đã cho ra đời nhiều thuật toán mã khối được công khai hoá trên thế giới như mã DES, 3DES, IDEA, AES, Gost,… Tuy nhiên, theo quan điểm của Nhà nước ta, việc bảo mật thông tin là một vấn đề rất nhạy cảm. Do đó, nếu dùng các hệ mật mã ở nước ngoài để mã hoá những thông tin chiến lược mà chưa được đánh giá thật đầy đủ về hệ mật đó thì rất nguy hiểm. Chính vì vậy, hiện nay Ban Cơ yếu Chính phủ đang thực hiện một đề tài cấp Nhà nước là xây dựng chuẩn mã dữ liệu của Việt Nam. Xuất phát từ nhu cầu thực tiễn hiện nay về bảo mật thông tin, em đã chọn đề tài: "***Nghiên cứu và xây dựng thuật toán mã khối phục vụ việc bảo mật thông tin chiến lược***", làm đối tượng nghiên cứu phục vụ cho luận văn của mình.

Bố cục của luận văn bao gồm 5 chương:

Chương 1. Tổng quan về mật mã học

Chương 2. Thực trạng an toàn thông tin trên các mạng máy tính

Chương 3. Các nguy cơ mất an toàn đối với dữ liệu được mã hoá

Chương 4. Mật mã khối và ý nghĩa của nó.

Chương 5. Đề xuất một hệ mã khối mới.

Trong đó, Chương 5 là trọng tâm của luận văn này. Ở chương này, một thuật toán tạo dãy bít giả ngẫu nhiên được đề xuất và chúng được sử dụng để mã hoá (dịch) các văn bản text la tinh. Thuật toán mã/dịch này đã được thể hiện trên máy tính bằng ngôn ngữ lập trình C#.

Do khả năng còn hạn chế, đặc biệt là khả năng toán học cho nên mặc dù em đã có nhiều cố gắng nhằm hoàn thành tốt nhất nhiệm vụ của mình nhưng không khỏi còn có nhiều thiếu sót. Em rất mong được sự chỉ bảo, đóng góp của các thầy cô giáo để luận văn này được hoàn thiện ở mức độ cao hơn.

Em xin chân thành cảm ơn./.

CHƯƠNG 1. TỔNG QUAN VỀ MẬT MÃ HỌC

Với một hệ thống thông tin mở, sử dụng công nghệ đa phương tiện như trong môi trường mạng máy tính thì về mặt lý thuyết không thể có vấn đề bảo mật/an toàn thông tin 100%, vấn đề quan trọng là chúng ta phải biết bảo vệ thế nào cho hiệu quả đối với hệ thống thông tin của mình. Bởi vậy, quy trình bảo mật thông tin cho hệ thống mạng máy tính phải tổng hoà các giải pháp đồng bộ về kỹ thuật và nghiệp vụ mật mã, đào tạo và giác ngộ về ý thức bảo mật, pháp luật và tổ chức.. Tuỳ thuộc vào yêu cầu, tính chất, mức độ quan trọng của dữ liệu và thông tin trao đổi trên mạng máy tính mà lựa chọn một, hoặc một tổ hợp giải pháp sao cho vừa hợp lý về mức chi phí, phù hợp với khả năng khai thác sử dụng, lại vừa đạt yêu cầu bảo mật đề ra. Trong hệ thống các giải pháp bảo mật/an toàn cho mạng máy tính, ***giải pháp sử dụng kỹ thuật mã hoá*** là không thể thiếu đối với hệ thống thông tin nhạy cảm.

1.1. Mật mã học

Mật mã học là một lĩnh vực liên quan với các kỹ thuật ngôn ngữ và toán học để đảm bảo an toàn thông tin, cụ thể là trong thông tin liên lạc. Về phương diện lịch sử, mật mã học gắn liền với quá trình mã hoá; điều này có nghĩa là nó gắn với các cách thức để chuyển đổi thông tin từ dạng này sang dạng khác nhưng ở đây là từ dạng thông thường có thể nhận thức được thành dạng không thể nhận thức được, làm cho thông tin trở thành dạng không thể đọc được nếu như không có các kiến thức bí mật. Quá trình mã hóa được sử dụng chủ yếu để đảm bảo tính bí mật của các thông tin quan trọng, chẳng hạn trong công tác tình báo, quân sự hay ngoại giao cũng như các bí mật về kinh tế, thương mại. Trong những năm gần đây, lĩnh vực hoạt động của mật mã đã được mở rộng: mật mã hiện đại cung cấp cơ chế cho nhiều hoạt động hơn là chỉ duy nhất việc giữ bí mật và có một loạt các ứng dụng như: chứng thực, chữ ký số, bầu cử điện tử hay tiền điện tử. Ngoài ra, những người không có nhu cầu thiết yếu đặc biệt về tính bí mật cũng sử dụng kỹ thuật mật mã, thông thường được thiết kế và tạo lập sẵn trong các cơ sở hạ tầng của công nghệ tính toán và liên lạc viễn thông.

Mật mã học là một khoa học nghiên cứu về mật mã. Nó là một chuyên ngành toán học. Mật mã là công cụ được sử dụng trong an ninh máy tính và mạng.

1.2. Thuật ngữ

Việc nghiên cứu tìm các phương thức để phá vỡ việc sử dụng mật mã được gọi là phân tích mật mã, hay *mã thám*. Mã hóa và phân tích mật mã được nhóm lại cùng nhau dưới tên gọi chung **mật mã học**, nó bao gồm toàn bộ các chủ đề liên quan đến mật mã. Trong thực tế, thuật ngữ *mật mã* thông thường được sử dụng gắn liền gắn liền với ngành Cơ yếu.

Mã hóa là quá trình chuyển đổi các thông tin thông thường (*văn bản thường*) thành dạng không đọc trực tiếp được, là *văn bản mã*. *Giải mã*, là quá trình ngược lại, phục hồi lại văn bản thường từ văn bản mã. *Mật mã* là thuật ngữ để chỉ việc mã hóa, giải mã và thám mã. Hoạt động chính xác của mật mã thông thường được kiểm soát bởi khóa — một đoạn thông tin bí mật nào đó cho phép tùy biến cách thức tạo ra văn bản mã. Các *giao thức mã hóa* chỉ rõ các chi tiết về việc mật mã (và các nền tảng mã hóa khác) được sử dụng như thế nào để thu được các nhiệm vụ cụ thể. Một bộ các giao thức, mật mã, quản lý khoá, các hành động quy định trước bởi người sử dụng thi hành cùng nhau như một hệ thống tạo ra *hệ thống mật mã*.

1.3. Phân tích mật mã

Mục tiêu của phân tích mật mã (phá mã) là tìm cách nào đó để xác định được khoá mã hoặc đọc được bản rõ ứng với bản mã thu được. Phân tích mật mã có thể được thực hiện bởi những kẻ tấn công ác ý, nhằm làm hỏng hệ thống; hoặc bởi những người thiết kế ra hệ thống (hoặc những người khác) với ý định đánh giá độ an toàn của hệ thống. Nói chung, việc phân tích mật mã đã được thực hiện bởi một tổ chức do Nhà nước qui định nhằm tìm kiếm các thông tin nhạy cảm trong các bản rõ nhận được, phục vụ yêu cầu tình báo của một quốc gia đối với quốc gia khác. Tổ chức này thường được trang bị các thiết bị vượt trội (có tính chuyên dụng) cùng với các nhà toán học giỏi nhằm nghiên cứu xây dựng các phương pháp đặc biệt phục vụ việc tìm khoá mã của đối phương, cùng thông tin rõ cần thiết. Những kết quả phân tích mật mã không bao giờ được công bố ít nhất cho đến hàng chục năm sau. Đó là nguyên tắc của bất cứ một cơ quan thám mã (phân tích mật mã) của bất cứ một quốc gia nào trên thế giới.

1.4. Lịch sử

Gậy mật mã của người Hy Lạp là một trong những dụng cụ đầu tiên trong ngành mật mã.

Mật mã có lịch sử lâu dài và đầy màu sắc. Nói chung, những dạng sớm nhất của cách viết bí mật (ngày nay gọi chung là *mã hóa cổ điển*) chỉ cần có bút và giấy. Hai phạm trù chính của mật mã cổ điển là mã chuyển vị, trong đó người ta sắp xếp lại trật tự các chữ cái của thông điệp, và mã thay thế, trong đó người ta thay thế có hệ thống các chữ cái hay các nhóm chữ cái bằng các chữ cái hay các nhóm chữ cái khác.

Các thiết bị và các kỹ thuật khác nhau đã được sử dụng để mã hóa. Một trong những thiết bị sớm nhất có lẽ là gậy mật mã (tiếng Hy Lạp: σκυτάλη). Trong nửa đầu thế kỷ 20, một số thiết bị cơ khí đã được phát minh để thực hiện mã hóa, bao gồm rotor machines — nổi tiếng nhất là máy Enigma được người Đức sử dụng trong Đại chiến thế giới 2. Mật mã thực hiện bằng các máy móc này đã tăng độ phức tạp lên đáng kể đối với công việc phân tích mã.

Các nghiên cứu rộng rãi có tính học thuật về mã hóa hiện đại là tương đối gần đây — nó chỉ được bắt đầu trong cộng đồng mở kể từ những năm thập niên 1970 với các chi tiết kỹ thuật của DES (viết tắt trong tiếng Anh của *Data Encryption Standard* tức *Tiêu chuẩn mã hóa Dữ liệu*) và sự phát minh ra RSA. Kể từ đó, mật mã đã trở thành công cụ được sử dụng rộng rãi trong liên lạc và bảo mật máy tính.

1.5. An ninh thông tin

Mật mã được sử dụng phổ biến để đảm bảo an toàn cho thông tin liên lạc. Các thuộc tính được yêu cầu là:

1. Tính bí mật: Chỉ có người nhận đã xác thực có thể lấy ra được nội dung của thông tin chứa đựng trong dạng đã mã hóa của nó. Nói khác đi, nó không thể cho phép thu lượm được bất kỳ thông tin đáng kể nào về nội dung của thông điệp.

2. Nguyên vẹn: Người nhận cần có khả năng xác định được thông tin có bị thay đổi trong quá trình truyền thông hay không.

3. Tính xác thực: Người nhận cần có khả năng xác định người gửi và kiểm tra xem người gửi đó có thực sự gửi thông tin đi hay không.

4. Không thể từ chối: Người gửi không thể từ chối việc gửi thông tin đi.

5. Chống lặp lại: Không cho phép gửi thông tin nhiều lần đến người nhận mà người gửi không hề hay biết.

1.6. Khóa đối xứng

Khóa mã đối xứng hoặc là sử dụng cùng một khóa cho việc mã hóa và giải mã hoặc là khóa (thứ hai) sử dụng để giải mã có thể dễ dàng tính được từ khóa (thứ nhất) đã dùng để mã hóa. Các thuật ngữ khác bao gồm mã hóa *khóa cá nhân*, mã hóa *một khóa*.

1.7. Khóa công khai

Khóa mã đối xứng có một số trở ngại không thuận tiện — hai người muốn trao đổi các thông tin bí mật cần phải chia sẻ khóa bí mật. Khóa cần phải được trao đổi theo một cách thức an toàn, mà không phải bằng các phương thức thông thường vẫn dùng để liên lạc. Điều này thông thường là bất tiện, và mã hóa khóa công khai (hay khóa bất đối xứng) được đưa ra như là một giải pháp thay thế. Trong mã hóa khóa công khai có hai khóa được sử dụng, là *khóa công khai* và *khóa cá nhân*, trong đó khóa công khai dùng để mã hóa còn khóa cá nhân dùng để giải mã. Rất khó để có thể thu được khóa cá nhân từ khóa công khai. Điều này có nghĩa là một người nào đó có thể tự do gửi khóa công khai của họ ra bên ngoài theo các kênh không an toàn mà vẫn chắc chắn rằng chỉ khi họ công khai nó thì mới có thể giải mã các thông điệp đã mã hóa bằng khóa đó.

1.8. Các vấn đề khác

Độ an toàn của các hệ thống mật mã thực tế vẫn chưa được chứng minh, đối với cả các hệ thống khóa bất đối xứng. Đối với mật mã khóa đối xứng, việc đánh giá độ an toàn mật mã đã được Shannon trình bày dựa trên lý thuyết độ bất định và lý thuyết xác xuất thống kê toán học. Vì thế mật mã loại này có thể có độ an toàn chứng minh được khi chống lại một tập hợp hữu hạn các cách thức tấn công. Đối với các hệ thống bất đối xứng, nói chung dựa trên độ khó của các vấn đề toán học có liên quan, nhưng chúng cũng không phải là độ an toàn có thể chứng minh.

CHƯƠNG 2. THỰC TRẠNG AN TOÀN THÔNG TIN TRÊN CÁC MẠNG MÁY TÍNH

2.1. Thực trạng an toàn thông tin hiện nay.

Mạng sử dụng các thuật toán mã hoá theo nhiều cách: trong quá trình xác nhận người sử dụng (user), để bảo mật dữ liệu, đảm bảo tính toàn vẹn của thông tin… Trên thực tế, chúng ta sử dụng mã hoá rất nhiều mà đôi khi không hề biết tới sự hiện diện của chúng.

Mật mã càng phát triển, những nhận thức sai lầm xung quanh nó xuất hiện ngày càng nhiều. Đôi khi, sự nhận thức sai lệch cản trở chúng ta tiếp cận những lợi ích thiết thực mà mật mã đem lại; sử dụng sai công cụ để thực thi một công việc bảo mật nào đó và đặt doanh nghiệp cuả mình trước những nguy cơ do các dữ liệu nhạy cảm và quan trọng không được bảo vệ ở mức cần thiết.

Sai lầm thứ nhất: Bảo mật song hành với sự toàn vẹn dữ liệu.

Sai lầm thứ 2: Checksum bảo vệ dữ liệu trong khi vận chuyển

Sai lầm thứ 3: SSL chỉ đảm bảo dữ liệu giữa trình duyệt và một website

2.2. Các mức bảo vệ an toàn mạng

Vì không thể có một giải pháp an toàn tuyệt đối, người ta thường phải sử dụng nhiều mức bảo vệ khác nhau tạo thành nhiều lớp "rào chắn" đối với các hoạt động xâm phạm. Việc bảo vệ thông tin trên mạng chủ yếu là bảo vệ các thông tin lưu giữ trên máy tính, đặc biệt là trong các server của mạng. Bởi thế, ngoài một số biện pháp nhằm chông thất thoát thông tin trên đường truyền, mọi cố gắng nhằm tập trung vào việc xây dựng các mức "rào chắn" từ ngoài vào trong cho các hệ thống . Hình dưới mô tả các lớp "rào chắn" thông dụng hiện nay để bảo vệ thông tin tại các trạm của mạng.

- Lớp bảo vệ trong cùng là quyền truy nhập (access rights) nhằm kiểm soát các tài nguyên của mạng (ở đây là thông tin) và quyền hạn trên tài nguyên đó (có thể được thực hiện những thao tác gì). Dĩ nhiên là kiểm soát được các cấu trúc dữ liệu càng chi tiết càng tốt. Hiện tại việc kiểm soát thường là ở mức tệp (file).

Hình 2.2. Các lớp rào chắn bảo vệ thông tin trên mạng

- Lớp bảo vệ tiếp theo là đăng nhập/mật khẩu (login/password). Thực ra đây cũng là kiểm soát quyền truy nhập, nhưng không phải truy cập ở mức thông tin mà ở mức hệ thống (tức truy cập vào mạng). Đây là phương pháp bảo vệ phổ biến nhất vì nó đơn giản, ít phí tổn và cũng rất có hiệu quả. Mỗi người sử dụng (kể cả người được quyền giám sát quản mạng - supervisor) muốn được vào mạng để sử dụng các tài nguyên của mạng đều phải có tên đăng nhập và mật khẩu trước đó. Người giám sát quản mạng có trách nhiệm quản lý, kiểm soát mọi hoạt động của mạng và xác định quyền truy nhập của những người sử dụng khác tuỳ theo thời gian và không gian (nghĩa là một người sử dụng có thể chỉ được phép vào mạng ở những thời điểm và từ những vị trí nhất định). Tuy nhiên, trong thực tế việc đánh cắp mật khẩu thường xuyên xảy ra. Thật vậy, ta biết rằng các mật

khẩu thường được lưu ở dạng rõ trong cơ sở dữ liệu và do đó các hackers có thể tìm cách lấy mật khẩu một các dễ dàng. Hơn nữa, chính những mật khẩu được các nhà quản trị nắm rất rõ và do đó việc mất mật khẩu là không thể tránh khỏi. Để khắc phục thực trạng này, thường người ta lưu các giá trị của mật khẩu qua hàm Hash mà không lưu trực tiếp mật khẩu ở dạng rõ trong cơ sở dữ liệu và các mật khẩu được thay đổi theo định kỳ thời gian.

- Để bảo mật thông tin truyền trên mạng, người ta sử dụng các phương pháp mật mã. Đây là một trong những lớp bảo vệ thông tin quan trọng nhất, dữ liệu được biến đổi từ dạng nhận thức được sang dạng không nhận thức được theo một giải thuật nào đó (mã hoá) và sẽ đực biến đổi ngược lại ở trạm nhận (giải mã). Đây là một lớp bảo vệ thông tin rất quan trọng và được sử dụng rộng rãi trong môi trường mạng.

- Bảo vệ vật lý (physical protection) nhằm ngăn cản các truy nhập vật lý bất hợp pháp và 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 khoá trên máy tính (ngắt nguồn điện đến màn hình và bàn phím nhưng vẫn giữ liên lạc trực tuyến giữa máy tính với mạng, hoặc cài cơ chế báo động khi có truy nhập vào hệ thống hoặc dùng các máy trạm không có ổ đĩa mềm…)

- Để bảo vệ từ xa một máy tính hoặc cho cả một mạng nội bộ (intranet), người ta thường dùng các hệ thống đặc biệt là tường lửa (firewall). Chức năng của các tường lửa là ngăn chặn các thâm nhập trái phép (theo danh sách truy nhập xác định trước) và thậm chí có thể "lọc" bỏ các gói tin mà ta không muốn gửi đi hoặc nhận vào vì lý do nào đó. Phương thức bảo vệ này được sử dụng nhiều trong môi trường liên mạng kết nối Internet.

2.3. Bảo vệ thông tin bằng mật mã:

Qúa trình chuyển đổi thông tin gốc sang dạng mật mã đực gọi là mã hoá (encrytion hay enciphering). Có hai cách tiếp cận để bảo vệ thông tin bằng mật mã đó là: tiếp cận theo đường truyền (link-oriented security) và tiếp cận từ mút-đến-mút (end-to-end). Trong cách thứ nhất, thông tin được mã hoá để bảo vệ trên đường truyền giữa hai nút không quan tâm đến nguồn và đích của thông tin đó. Ưu điểm của cách tiếp cận này là có thể bí mật được luồng thông tin giữa nguồn và đích và có thể ngăn chặn được toàn bộ các vi phạm nhằm phân tích lưu thông trên mạng. Nhược điểm của nó là vì thông tin chỉ được mã hoá trên đường truyền nên đòi hỏi các nút phải được bảo vệ tốt. Ngược lại, trong cách thứ hai, thông tin được bảo vệ trên toàn đường đi từ nguồn tới đích. Thông tin được mã hoá ngay khi mới được tạo ra và chỉ được giải mã khi đến đích. Ưu điểm chính của tiếp cận này là một người sử dụng hoặc máy chủ (host) có thể dùng nó mà không ảnh hưởng gì đến người sử dụng hoặc máy chủ khác. Nhược điểm của phương pháp này là chỉ có dữ liệu người sử dụng được mã hoá, còn thông tin điều khiển thì phải giữ nguyên để có thể xử lý tại các nút.

2.4. Tấn công vào các dữ liệu được mã hoá

Trước khi vai trò của các máy tính được lên ngôi thì việc mã hoá và giải mã các bản tin mật chủ yếu dựa vào tài năng khôn khéo của con người và nó có thể là một bí quyết nào đó. Với nhiều loại mã cổ điển thì việc phân tích một bản tin đã mã hoá đều có thể thực hiện được bởi cách này hay cách khác. Sự ra đời của các máy tính đã giúp cho công việc mã hoá và giải mã tiến một bước khá dài. Máy tính có thể thực hiện các phép tính phức tạp mà bằng các phương pháp khác phải mất hàng năm, hàng chục năm, hàng trăm năm mới thực hiện được.

Ngày nay, việc đánh giá độ mật của mật mã thường được giả thiết rằng, thuật toán mã đã biết và vấn đề còn lại là giải mã thông báo đã được mã hoá bằng cách khám phá ra khoá mã. Công việc sẽ rất khó khăn cho việc phân tích mã bởi vì duy nhất chỉ dựa vào thông báo đã được mã hoá, không có thông tin gì về bản thông báo rõ. Nếu như không có một sự dư thừa nào trong bản thông báo, thì việc khám phá khoá mã gặp nhiều khó khăn, Nếu như biết được một phần nào đó của bản tin (thông báo) nguyên thuỷ thì bài toán có thể giải được, ví dụ trong phần tiêu đề của bản tin (thông báo) có phần tiêu đề được viết theo chuẩn. Các khoá có thể được thử

lần lượt cho đến khi phần tiêu đề được thừa nhận xuất hiện trong bản tin được giải mã. Nếu như phần tiêu đề theo chuẩn đó khá dài thì việc nhận dạng khoá càng chính xác.

2.5. Các mối đe doạ mất an toàn

2.5.1. Các xâm nhập thụ động

Các mối đe doạ đối với toàn bộ hệ thống được bảo mật thuộc một hệ thống truyền tin hoặc một hệ thống lưu trữ dữ liệu như đã biết có thể phân làm hai loại: các đe doạ dạng tích cực và các đe doạ dạng thụ động. Một xâm nhập thụ động đối với một hệ thống là các mưu toan vượt qua hàng rào bảo vệ để thu những thông tin dữ liệu truyền trên kênh truyền hoặc lưu giữ trong bộ nhớ mà không làm sai lạc các nội dung thông tin dữ liệu. Loại xâm nhập này đối với các hệ thống truyền tin xuất hiện từ thời phát minh ra điện báo

Một sự rẽ nhánh thụ động không cần phải lọt qua các giao thức kiểm tra dòng dữ liệu. Người ta có thể kiểm tra phát hiện một sự rẽ nhánh thụ động trên đường truyền vật lý bằng cách đo chính xác các đặc tính kỹ thuật của đường truyền. Nhưng việc đo lường các thông số đó không thể thực hiện được trong trường hợp đường truyền kết nối vô tuyến. Để bảo vệ chống lại các xâm nhập thụ động trong các trường hợp này chỉ có cách dùng phương pháp mã hoá để bảo vệ dữ liệu truyền là tốt nhất.

2.5.2. Các xâm nhập tích cực:

Các xâm nhập tích cực là mối nguy hiểm tiềm tàng. Ở đây kẻ xâm nhập tìm cách làm sai lạc các dữ liệu truyền hoặc các dữ liệu lưu trữ và hy vọng rằng, chủ sở hữu hoặc người sử dụng hợp pháp không nhận biết. Việc làm sai lệch các dữ liệu được lưu giữ có thể gây ra các sử dụng sai lạc của các đường truy nhập. Ở đây sẽ không đề cập đến các thủ tục truy nhập, mặc dù đó là một chủ đề quan trọng, mà chỉ đề cập đến vấn đề bảo vệ các dữ liệu chống lại các xâm nhập tích cực làm sai lạc dữ liệu. Để làm được điều đó thì phải có một cấu trúc mã sao cho tất cả các việc làm sai lạc các cấu trúc đó không thể thực hiện được nếu không phân tích được mã. Một khi mà kẻ xâm nhập sử dụng phương tiện truy nhập bất hợp pháp vào một cơ sở dữ liệu thì việc phá hoại các dữ liệu hoặc ăn cắp thông tin không thể tránh khỏi. Đó là chủ đề lớn thuộc lĩnh vực bảo vệ dữ liệu lưu giữ.

Một sự xâm nhập tích cực trên đường truyền tin hầu như cũng khó ngăn chặn giống như các xâm nhập thụ động. Tuy vậy việc phát hiện chúng thì dễ hơn nhiều, không cần phải mã hoá. Việc làm sai lêch các dữ liệu lưu trữ cần phải có thời gian và độ chính xác của thời gian truyền các dữ liệu cũng có thể giúp phát giác các xâm nhập.

Ở các đường truyền tin vô tuyến thì rất khó phát giác việc xâm nhập, và đây thực sự là một cuộc đấu trí. Nếu như một đường truyền vật lý được giám sát chặt chẽ và thường xuyên thì lúc đó hầu như không một xâm nhập nào không được phát hiện.

Việc thực hiện một xâm nhập rẽ nhánh tích cực là một công việc không đơn giản. Việc xâm nhập tích cực phải ngắt giao thức đó và đưa vào một "giao thức giả". Như vậy các thông tin thực sẽ từ nguồn về kẻ xâm nhập và các thông tin giả từ kẻ xâm nhập về đầu cuối mà dầu cuối không nhận biết được. Với một số biến đổi, việc xâm nhập tích cực như trên có thể thiết lập được với kênh truyền tin. Nếu mạng truyền tin là mạng diện rộng, hoặc nối mạng Internet thì việc thiết lập xâm nhập tích cực trên càng có điều kiện, bởi vì các giao thức truyền tin đã được công bố.

CHƯƠNG 3. BẢO VỆ THÔNG TIN BẰNG MẬT MÃ HOÁ

3.1. Vai trò của mật mã:

Việc mã hoá thông tin đóng vai trò quan trọng trong việc đảm bảo sự an toàn cho dữ liệu lưu giữ và truyền tải trên hệ thống. Đó là:

- *Mật mã được dùng để che dấu thông tin riêng ở nơi mà thông tin bị phơi bày* trong các thành phần của hệ thống, như là các kênh truyền vật lý có thể bị tấn công bởi nghe trộm và giả mạo thông báo. Ứng dụng này của mật mã tương ứng với ứng dụng truyền thống của nó trong quân sự và trong các hoạt động tình báo.

- *Mật mã được sử dụng trong việc xác thực các thông báo*. Một đối tượng giải mã thành công một thông báo khi dùng khoá nghịch đảo đặc biệt thì có thể thừa nhận thông báo là xác thực (đáng tin cậy) nếu nó chứa một giá trị mong đợi. Nó hoàn toàn không giống sự phát sinh từ việc giải mã thông báo nhận được bằng một khoá khác nào đó và do vậy người nhận có thể suy luận ra rằng người gửi thông báo có khoá mã tương ứng. Vậy nếu các khoá được giữ bí mật (hoặc nếu một khoá của cặp khoá được giữ bí mật trong trường hợp lược đồ mã hoá khoá công khai) thì việc giải mã thành công sẽ xác nhận thông báo đã giải mã đến từ một người gửi cụ thể (là chủ nhân của thông báo).

- *Mật mã được dùng để thi hành một cơ chế nhận biết gọi là chữ ký số.* Chữ ký số mô phỏng vai trò của các chữ ký truyền thống thông thường, kiểm tra rồi cho đối tượng thứ ba biết rằng thông báo là một bản sao còn nguyên chưa bị thay đổi của một thông báo do một đối tượng cụ thể nào đó gửi. Khả năng cung cấp các chữ ký số phụ thuộc vào điều mà người gửi đầu tiên có thể làm được nhưng những người khác không thể làm được.

3.2. Quy trình mã hoá dữ liệu

Để mã hoá thông tin, chúng ta chuyển đổi thông tin đó thành dạng mà bất cứ người nào cũng không thể hiểu được ngoại trừ đúng người nhận có trong tay "khoá" để chuyển đổi thông tin đã mã hoá thành dạng ban đầu. Trong phần này chúng ta điểm lại 2 cách tiếp cận hay được sử dụng nhất để mã hoá trên máy tính.

Các kỹ thuật mã hoá trên máy tính chia thành 2 lớp chính – các phương pháp mã hoá khoá bí mật và mã hoá khoá công khai. Cả hai đều hữu ích trong việc thực hiện sự an toàn và chúng ta sẽ thấy rằng trong nhiều trường hợp cả hai phương pháp mã hoá đều được sử dụng mà không có ảnh hưởng lớn đến các phương pháp xác nhận và kiểm soát truy nhập được dùng. Chúng ta dùng thuật ngữ "văn bản" trong phần sau để ám chỉ một dãy các dữ liệu theo một kiểu nào đó; "bản rõ" (plain text) là văn bản chưa được mã hoá; "bản mã" (cipher text) là văn bản đã được mã hoá.

Khoá: gồm một số hữu hạn các bit thường được biểu thị dưới dạng các xâu ký tự, số thập phân hoặc thập lục phân. Mặc dù độ dài của khoá có thể bằng độ dài của bản rõ nhưng mầm khoá thường dùng các khoá 8 ký tự.

Phương pháp truyền thống thường dùng cùng một khoá để mã hoá và giải mã. Lúc đó khoá phải được giữ bí mật tuyệt đối. Người ta thường gọi đó là hệ thống mã hoá cổ điển, đối xứng hoặc một - khoá.

Một phương pháp khác sử dụng khoá công khai (còn gọi là phương pháp mã hoá bất đối xứng hay hai - khoá) trong đó khoá để mã hoá và khoá để giải mã là khác nhau. Các khoá này tạo thành một cặp chuyển đổi ngược nhau và không khoá nào có thể suy ra được từ khoá kia về mặt tính toán có hiệu quả. Khoá mã hoá là công khai, còn khoá giải mã thì phải được giữ bí mật.

Để tăng cường độ an toàn cho mật mã người ta chia khoá thành các nhóm khoá:

- Khoá dài hạn (Master key).
- Khoá ngắn hạn (Secondary key).
- Khoá phiên (Session key).

Nhờ đó, nếu một trong các lớp khoá đó bị lộ thì hệ mật mã vẫn đảm bảo an toàn nhờ các khoá còn lại. Các khoá này thường được gửi tới người nhận bằng các kênh khác nhau.

Xu hướng tạo khoá như vậy hiện nay đang phát triển mạnh nhưng không được công khai. Chẳng hạn các hãng sản suất máy mã trên thế giới như: Crypto-AG, GreTag (Thuỵ sỹ); Racal (Anh); BBG (Nhật Bản); Mills (Áo); Siemen (Đức)… đã và đang thực hiện cơ chế tạo "khoá bội" này.

3.3. Hàm mã hoá và khoá mã

Hàm mã hoá là phép biến đổi f từ bản rõ thành bản mã. Tức là một ánh xạ f từ không gian các bản rõ vào không gian các bản mã. Ánh xạ này phải đảm bảo tính duy nhất của bản rõ. Nghĩa là, giả sử ánh xạ f: $X \rightarrow Y$ trong đó X là không gian bản rõ, Y là không gian bản mã. Khi đó không tồn tại $X_1 \neq X_2$, $X_1, X_2 \in X$ mà $f(X_1)=f(X_2)$; tức là hàm f phải là hàm đơn trị. Tuy nhiên, điều kiện đó chưa đảm bảo bí mật cao vì về lý thuyết tập hợp X và Y là hữu hạn và hàm f không thay đổi nên theo thời gian và tốc độ tính toán của máy tính ngày càng được cải thiện thì sớm muộn các nhà phân tích mật mã có tổ chức sẽ tìm được hàm f (được thể hiện bằng thuật toán mã/dịch). Do đó cần phải có "mầm khoá – Key seed" bổ trợ. Chính mầm khoá này đóng vai trò quyết định cho độ mật của một hệ mật mã. Thực tế hàm mã f là một ánh xạ từ tích $X \times K$ vào không gian bản mã Y tức là ánh xạ:

$f: X \times K \rightarrow Y$

sao cho với mỗi $k \in K$ và với mỗi bản rõ $x \in X$, ta có : $f(x,k)=y \in Y$.

Để đảm bảo độ an toàn mật mã cao, các nhà thiết kế hệ mật phải xây dựng không gian khoá K với lực lượng đủ lớn nhằm chống lại các nhà thám mã dùng thuật toán 'vét cạn' bằng cách thử từng khoá (mầm khoá) k một. Hơn thế nữa, người ta yêu cầu sự phân bố của các phần tử khoá $k \in K$ phải là ngẫu nhiên và phân bố đều. Ngày nay, trên lĩnh vực thương mại người ta yêu cầu hàm mã hoá f là công khai và chỉ giấu kín phần tử khoá. Mục đích là để tiện lợi cho người dùng mã hoá dữ liệu như các thuật toán mã hoá : DES, 3DES, IDEA, AES, Gost,… Để nơi nhận đích thực có khoá để giải mã sau khi họ nhận được bản mã gửi đến từ nơi gửi. Tất nhiên, nơi gửi và nơi nhận phải có giao thức trao đổi khoá (hoặc phân phối khoá).

Độ mật của mật mã không phải ở thuật toán mà là ở khoá mã. Mọi thuật toán đối phương có thể có. Chính vì vậy một trách nhiệm lớn lao đối với người quản lý hệ thống trong việc tạo khoá, phân phối khoá và huỷ khoá sau khi sử dụng. Có hai giải pháp thực hiện việc trao đổi (hoặc phân phối) khoá. Đối với các hệ mật mã khoá đối xứng (một khoá) trước đây người ta thực hiện việc trao đổi (phân phối) khoá cho nhau bằng một kênh liên lạc hoàn toàn bí mật mà chỉ nơi gửi và nơi nhận mới biết được. Còn đối với các hệ mật mã khoá bất đối xứng, khoá mã và khoá giải mã là khác nhau và khó suy được khoá này từ khoá kia nên vấn đề trao đổi khoá coi như được giải quyết. Do tính bảo mật cao đã được đứng vững trong hàng chục năm và dù tốc độ mã hoá rất nhanh được thực hiện bởi máy vi tính hoặc bằng các máy mã chuyên dụng nên mật mã khoá đối xứng vẫn là nhu cầu tối cần thiết cho nhiều ứng dụng trong an ninh, quốc phòng, trong ngoại giao, tình báo và cả trong thương mại… Tuy nhiên, sự bất tiện nhất của mật mã khoá đối xứng là vấn đề trao đổi khoá. Do đó, xu hướng hiện nay là kết hợp cả hai hệ mật: Hệ mật mã khoá đối xứng được dùng để mã/dịch dữ liệu, còn hệ mật mã khoá bất đối xứng lại được sử dụng để thực hiện việc trao đổi khoá (phân phối khoá). Một trong những hệ mật như thế chúng ta có thể tìm thấy chẳng hạn hệ mật mã PGP (Pretty Good Privacy).

3.4. Tạo các dãy bít giả ngẫu nhiên và ứng dụng trong việc mã hoá dữ liệu

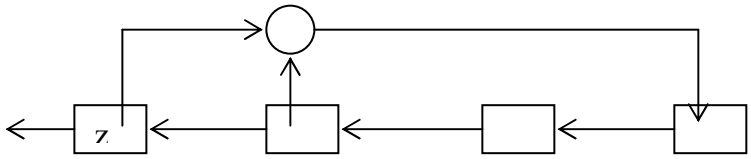
Tạo các số ngẫu nhiên, các xâu bít ngẫu nhiên là một vấn đề quan trọng trong nhiều bài toán của mật mã học. Ví dụ các khoá mật cần phải được tạo một cách ngẫu nhiên từ một không gian khoá xác định: nhiều giao thức yêu cầu phải tạo được các số ngẫu nhiên trong quá trình thực hiện. Tạo các số ngẫu nhiên bằng cách tung đồng xu hoặc bằng các quá trình vật lý đòi hỏi nhiều thời gian và chi phí. Bởi vậy trong thực tế người ta thường sử dụng các bộ tạo bít giả ngẫu nhiên. Các bộ tạo bít giả ngẫu nhiên này bắt đầu bằng một xâu bít ngắn (được gọi là mầm) và sẽ mở rộng nó thành một xâu bít "có vẻ ngẫu nhiên" dài hơn nhiều rất cần cho nhiều ứng dụng.

Việc tạo các số giả ngẫu nhiên, các xâu bít ngẫu nhiên được xếp vào loại tốt khi nó thoả mãn các điều kiện sau:

- Các số ngẫu nhiên kế tiếp có thể được tính toán không quá phức tạp, giá thành chấp nhận.
- Dãy được tạo ra xuất hiện đúng thực sự là ngẫu nhiên, có nghĩa là các số ngẫu nhiên kế tiếp là hoàn toàn độc lập, không có tương quan với các số khác trước đó.
- Chu kỳ của chúng (thời gian sau đó nó lặp lại) là đủ lớn sao cho chúng có độ dài ít nhất là bằng độ dài các bản thông báo cần mã. [1].

3.4.1. Tạo bít giả ngẫu nhiên với thanh ghi chuyển dịch phản hồi tuyến tính

Dòng khoá đó được sinh bởi thanh ghi chuyển dịch phản hồi tuyến tính sau đây:



3.4.2. Tạo bít giả ngẫu nhiên RSA.

Cơ chế tạo dãy bít giả ngẫu nhiên RSA được mô tả như sau : Chọn số nguyên $n = p.q$ là tích của hai số nguyên tố p và q có biểu diễn nhị phân với độ dài cỡ $m/2$ bít (như vậy n có biểu diễn nhị phân cỡ m bít), và một số b sao cho $\gcd(b, \phi(n)) = 1$ (gcd - ước số chung lớn nhất). Lấy $R = Z_n^*$, và với mỗi $r \in R$ xác định dãy số s_0, s_1, s_2,.... như sau:

$$\begin{cases} s_0 = r, \\ s_{i+1} = s_i^b \bmod n, \end{cases}$$

và sau đó định nghĩa $z_i = \varphi(r,i) = s_i \bmod 2$, tức z_i là bit thấp nhất trong biểu diễn nhị phân của số s_i. Dãy $K = z_1 z_2 .... z_i ....$ là dòng bít đồng bộ được tạo ra bởi mầm r.

3.4.3. Tạo bít giả ngẫu nhiên BBS (Blum-Blum-Shub):

Cơ chế tạo bít giả ngẫu nhiên BBS được mô tả như sau : Chọn $n = p.q$ là tích của hai số nguyên tố dạng $4m + 3$, tức $p \equiv 3 (\text{mod} 4)$ và $q \equiv 3 (\text{mod} 4)$. Gọi $QR(n)$ là tập các thặng dư bậc hai theo modn. Lấy $R = QR(n)$, và với mỗi $r \in R$ xác định dãy số s_0, s_1, s_2,.... như sau:

$$\begin{cases} s_0 = r, \\ s_{i+1} = s_i^2 \bmod n, \end{cases}$$

và sau đó định nghĩa $z_i = \varphi(r,i) = s_i \bmod 2$, tức z_i là bít thấp nhất trong biểu diễn nhị phân của số s_i. Dãy $K = z_1 z_2 .... z_i ....$ là dòng bít đồng bộ được tạo ra bởi mầm r.

3.4.4. Tạo bít giả ngẫu nhiên dựa vào bài toán logarit rời rạc :

Chọn p là một số nguyên tố lớn, và α là một phần tử nguyên thuỷ theo modp. Tập các mầm khoá là $R = Z_p^*$. Với mỗi mầm khoá $r \in R$ ta xác định dãy số $s_0,...,s_i ....$ bởi :

$$s_0 = r,$$

$$s_{i+1} = \alpha^{s_i} \bmod p.$$

Sau đó định nghĩa $z_i = \varphi(r,i)$ ($i = 1,2,....$) như sau: $z_i = 1$ nếu $s_i > p/2$, và $z_i = 0$ nếu $s_i < p/2$. Và $K = z_1 .... z_i .....$ là dòng khoá, tức dòng bít giả ngẫu nhiên, được tạo ra.

CHƯƠNG 4. MẬT MÃ KHỐI VÀ Ý NGHĨA CỦA NÓ

4.1. Thế nào là một hệ mã khối

Mã khối là một ánh xạ từ những khối bản rõ thường là có độ dài cố định vào những khối bản mã có cùng độ dài. Để có thể giải mã được thì các bản rõ khác nhau phải ánh xạ vào các bản mã khác nhau. Các thuật toán được đề xuất là DES, IDEA, AES,…

Mật mã khối xử lý các khối dữ liệu có độ dài cố định (ví dụ 64 bít) và độ dài đoạn tin có thể bất kỳ. Có 4 phương pháp ứng dụng mã khối thường gặp trong các hệ thống bảo mật, đó là:

- Chế độ dùng từ điển điện tử, còn gọi là chế độ ECB (Electronic CodeBook).
- Chế độ móc xích các khối đã được mã hoá, còn gọi là chế độ CBC (Cipher Block Chaining)
- Chế độ phản hồi bản tin mã hoá, còn gọi là chế độ CFB (Cipher FeedBack)
- Phương pháp phản hồi đầu ra, còn gọi là chế độ OFB (Output FeedBack).

4.2. Ý nghĩa của mật mã khối

Mã khối có một ưu điểm đó là tốc độ mã hoá rất nhanh, nó có thể đáp ứng cho việc mã dữ liệu, tiếng nói.

Mã khối là yếu tố nổi bật và quan trọng trong một số các hệ thống mật mã. Cụ thể chúng cung cấp một sự tin cậy. Giống như việc xây dựng một khối cơ bản, chúng linh hoạt hơn trong việc cho phép xây dựng các số giả ngẫu nhiên, dòng khoá, MAC và các hàm băm.

Mã khối có thể ứng dụng cho cả các thuật toán mã hoá bí mật và công khai.

4.3. Quan sát một số hệ mật mã khối đang được sử dụng rộng rãi trên thế giới.

4.3.1. DES (Data Encryption Standard)

4.3.2. PGP (Pretty Good Privacy)

4.3.3. AES (Advanced Encryption Standard)

CHƯƠNG 5. ĐỀ XUẤT MỘT HỆ MÃ KHỐI MỚI

Nghiên cứu xây dựng thuật toán mã/dịch văn bản (dữ liệu) dựa trên các thanh ghi dịch phản hồi phi tuyến tính

Thuật toán gồm các khối con sau đây:

5.1. Thuật toán tạo dãy giả ngẫu nhiên

5.1.1. Mầm khoá (key seed)

Hệ thống này gồm 2 mầm khoá được gọi là khoá dài hạn (master key) và khoá phiên (session key); lần lượt được ký hiệu là SK_1 và SK_2.

SK_1: gồm một dãy 20 chữ cái La tinh, được ký hiệu là: $b_1, b_2,\ldots,b_{20}$. $b_i \in \{a, b, c, \ldots, z\}$.

SK_2: gồm một dãy 10 chữ cái La tinh, được ký hiệu là: $d_1, d_2, \ldots, d_{10}$. $d_i \in \{a, b, c, \ldots, z\}$.

Các chữ cái b_i và d_j này được chuyển thành các vectơ nhị phân 5 thành phần theo bảng Code quốc tế $N^{\circ}2$ của CCITT và được ký hiệu lần lượt là: $\vec{b_i}$, $\vec{d_j}$, i = 1..20, j = 1..10.

Như vậy, khi ta viết b_i và d_j thì có nghĩa đó là các chữ cái thứ i và thứ j lần lượt của mầm khoá SK_1 và SK_2. Còn khi ta viết $\vec{b_i}$, $\vec{d_j}$ thì các ký hiệu này có nghĩa là các vectơ nhị phân 5 thành phần ứng với b_i và d_j.

Ví dụ: b_i = a, d_j = b thì tức là $\vec{b_i}$ = (11000), $\vec{d_j}$ = (10011). Xem bảng mã quốc tế $N^{\circ}2$ (MTK_2) của CCITT.

Bảng mã quốc tế MTK_2

Chữ cái La tinh	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Vec tơ nhị phân tương ứng (α)	1	1	0	1	1	1	0	0	0	1	1	0	0	0	0	0
	1	0	1	0	0	0	1	0	1	1	1	1	0	0	0	1
	0	0	1	0	0	1	0	1	1	0	1	0	1	1	0	1
	0	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0
	0	1	0	0	0	0	1	1	0	0	0	1	1	0	1	1
Số $g_1(\alpha)$	3	25	14	9	1	13	26	20	6	11	15	18	28	12	24	22
Số $g_2(\alpha)$	24	19	14	18	16	22	11	5	12	26	30	9	7	6	3	13

Chữ cái La tinh	Q	R	S	T	U	V	W	X	Y	Z	3	4	8	9	+	/
Vec tơ nhị phân tương ứng (α)	1	0	1	0	1	0	1	1	1	1	0	0	1	0	1	0
	1	1	0	0	1	1	1	0	0	0	0	1	1	0	1	0
	1	0	1	0	1	1	0	1	1	0	0	0	1	1	0	0
	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1	0
	1	0	0	1	0	1	1	1	1	1	0	0	1	0	1	0
Số $g_1(\alpha)$	23	10	5	16	7	30	19	29	21	17	8	2	31	4	27	0
Số $g_2(\alpha)$	29	10	20	1	28	15	25	33	21	17	2	8	31	4	27	0

$$\mathbf{g_1(\alpha)}=\sum_{j+1}^{5} 2^{j-1} * \alpha_j = \boldsymbol{\alpha}^* \in \mathbf{C_{32}} \quad \textbf{và } \mathbf{g_2(\alpha)}=\sum_{j+1}^{5} 2^{j-1} * \alpha_{6-j} = \boldsymbol{\alpha}^* \in \mathbf{C_{32}}$$

5.1.2. Cấu tạo của hệ thống các thanh ghi dịch

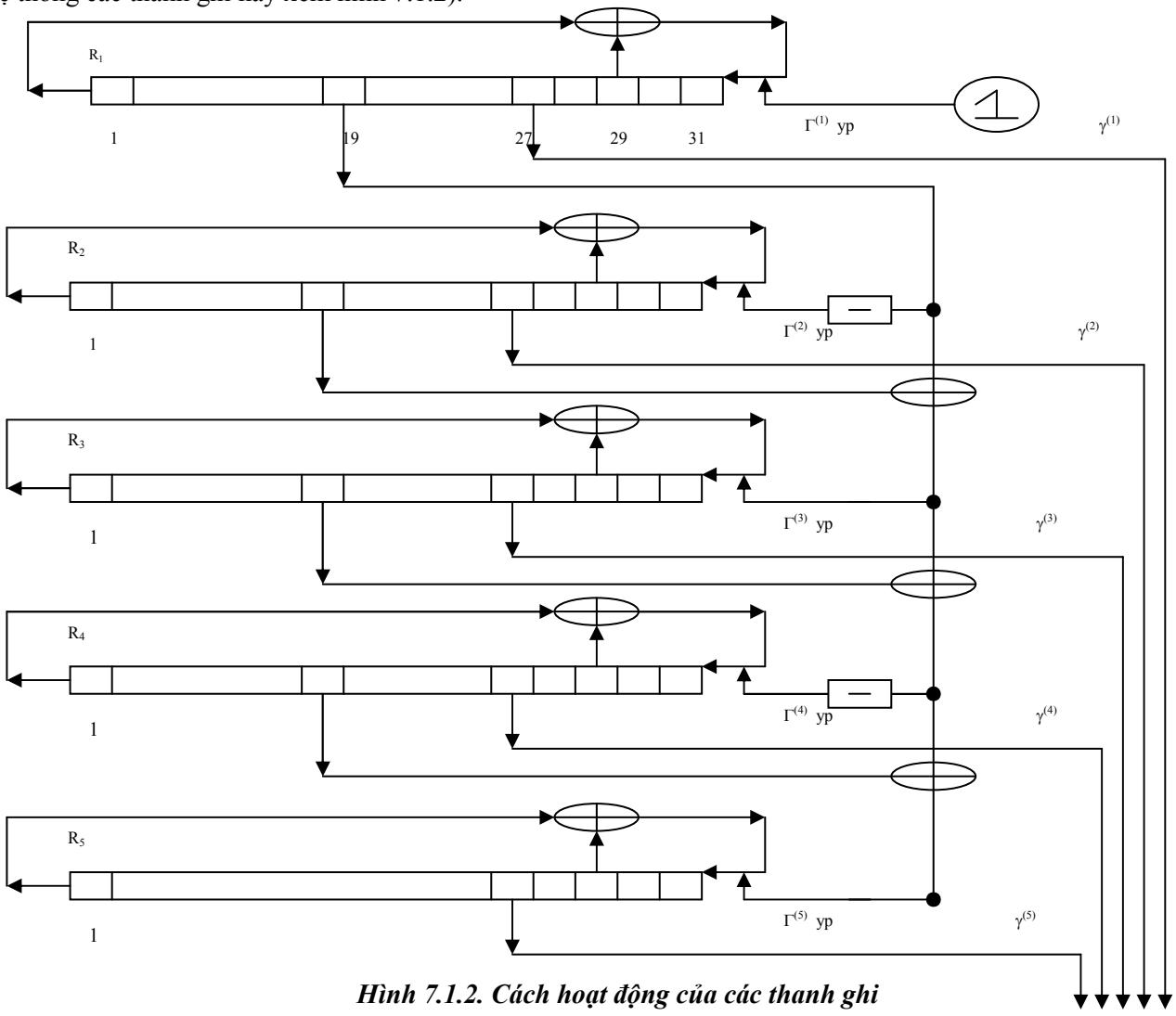
Việc xây dựng hệ thống các thanh ghi dịch gồm bao nhiêu thanh, qui tắc hoạt động của chúng ra sao và độ dài ngắn của từng thanh ghi như thế nào là tuỳ thuộc vào người thiết kế. Tuy nhiên, nếu hệ thống chỉ gồm

một thanh thôi thì dãy số giả ngẫu nhiên do nó tạo ra sẽ không tốt. Còn nếu số thanh tăng lên thì nếu lấy độ dài không tốt thì sẽ gây ra những chu kỳ con của dãy số giả ngẫu nhiên được tạo ra và như vậy chất lượng của dãy số giả ngẫu nhiên cũng tồi. Theo lý thuyết toán học, độ dài các thanh ghi phải nguyên tố với nhau từng đôi một, hoặc phải là một số nguyên tố. Thuận lợi nhất là số lượng các thanh ghi phản hồi phải bằng 5 hoặc bằng 8. (Trong quá trình thực hiện, chúng ta sẽ làm sáng tỏ điều này).

Hệ thống thanh ghi dịch phản hồi phi tuyến tính trong trường hợp cụ thể của chúng ta là gồm 5 thanh ghi được ký hiệu lần lượt là R_1, R_2,...,R_5.

Độ dài của mỗi thanh ghi gồm 31 ô, 31 ô của thanh R_j được ký hiệu là: $y_1^{(j)}$, $y_2^{(j)}$,…,$y_{31}^{(j)}$, j = 1..5.

Thông tin đầu tiên chứa trong 31 ô này được gọi là 31 trạng thái ban đầu của thanh R_j (cách hoạt động của hệ thống các thanh ghi này xem hình 7.1.2).



Hình 7.1.2. Cách hoạt động của các thanh ghi

5.*1.3. Lấp đầy ban đầu cho tất cả các ô của 5 thanh ghi dịch R_j*

Các thanh ghi dịch được lấp đầy thông tin đầu tiên nhờ 2 hệ thống mầm khoá SK_1 và SK_2 như sau:

Giả sử SK_1 gồm: $\vec{b}_1$, $\vec{b}_2$,…, $\vec{b}_{20}$

SK_2 gồm: $\vec{d}_1$, $\vec{d}_2$,…, $\vec{d}_{10}$

Ký hiệu: $\vec{b_i} = (b_i^{(1)}, b_i^{(2)},\ldots,b_i^{(5)})$; $\vec{d_i} = (d_i^{(1)}, d_i^{(2)},\ldots,d_i^{(5)})$. Với $b_i^{(j)}, d_i^{(j)} \in \{0,1\}$. $\forall i,j$.

Đặt:

$y_1^{(j)} = 1$, $\forall j=1..5$.

$y_{(i+1)}^{(j)} = b_i^{(j)}$, i=1..20, j=1..5;

$y_{i+21}^{(j)} = c_i^{(j)}$

Trong đó, $c_i^{(j)}$ là thành phần thứ j của vectơ nhị phân $\vec{c_i}$, tức là: $\vec{c_i} = (c_i^{(1)}, c_i^{(2)},..c_i^{(5)})$, i=1..10 và $\vec{c_i} = g_2^{-1}[g_2(\vec{d_i}) - g_1(\vec{b_{i+10}}) \pmod{32}]$

nghĩa là: $\vec{c_i} = g_2^{-1}(c_i)$. Trong đó: $c_i = g_2(\vec{d_i}) - g_1(\vec{b_{i+10}}) \pmod{32}$.

Còn các hàm $g_1()$, $g_2()$ được định nghĩa như sau:

$$g_1(\vec{b_i}) = \sum_{j=1}^{5} 2^{j-1} b_i^{(j)}$$

$$g_2(\vec{b_i}) = \sum_{j=1}^{5} 2^{j-1} b_i^{(6-j)}.$$

Hai hàm $g_1()$ và $g_2()$ được lập bảng ứng với 32 vectơ nhị phân của bộ mã quốc tế $N^{\circ}2$. Chú ý: Hàm g_1, g_2 là những song ánh, nghĩa là chúng là những ánh xạ 1-1.

Như vậy, sau bước này 31 ô của R_j (j=1..5) đã được lấp đầy (ban đầu) thông tin nhờ 2 hệ thống mầm khoá cho trước.

5.1.4. Cách hoạt động của 5 thanh ghi để tạo ra các vectơ nhị phân.

Giả sử, cả 5 thanh ghi dịch, đã được lấp đầy ban đầu thông tin cho tất cả các ô (mỗi thanh ghi có 31 ô), tức: $y_1^{(j)}, y_2^{(j)},..,y_{31}^{(j)}$. Ta gọi: $\vec{\gamma} = (\gamma_t^{(1)}, \gamma_t^{(2)}, .., \gamma_t^{(5)})$ là vectơ loạn số (vectơ nhị phân 5 thành phần) được R_1, R_2,..R_5 tạo ra ở bước thứ t (t=1,2,...).

Tại bước thứ nhất (t=1) ta có:

$$\vec{\gamma_1} = (\gamma_1^{(1)}, \gamma_1^{(2)}, .., \gamma_1^{(5)}) \qquad (1)$$

với $\gamma_1^{(j)} = y_{27}^{(j)}$, j=1..5.

Tuy nhiên, với t=1 thì nó chỉ được dùng để tạo nên một ký tự mã qua một ký tự rõ mà thôi. Trong lúc đó bản rõ thì khá dài. Vậy làm sao để tạo ra được $\vec{\gamma_t}$; với t=2,3,4,.... tuỳ ý.

Để xác định được dãy các vectơ loạn số $\vec{\gamma_t}$ ta cần tạo ra được các trạng thái tiếp theo của 5 thanh ghi dịch: $y_{32}^{(j)}, y_{33}^{(j)},...$;

Chúng ta làm như sau:

Đặt $y_{i+31}^{(j)} = y_i^{(j)} \oplus y_{i+28}^{(j)}$ i=1,2,3,... (2).

Nhờ hệ thức đệ quy (2) chúng ta sẽ thu được một dãy dài tuỳ ý $\{y_i^{(j)}\}$ qua 31 trạng thái ban đầu của cả 5 thanh ghi. $\{R_j\}$.

Nhờ vậy, ta xây dựng các vectơ loạn số $\vec{\gamma_t}$ bởi hệ thức sau:

$\overrightarrow{\gamma_t} = (\gamma_t^{(1)}, \gamma_t^{(2)}, .., \gamma_t^{(5)})$, trong đó $\gamma_t^{(j)} = y^{(j)}_{\varphi(j,t)+27}$, j=1.,,5, (3)

Ở đây, hàm φ(j,t) được định nghĩa theo quy tắc sau:

$$\left.\begin{array}{l}\varphi(j,t) = 0 \\ \varphi(j,t+1) = \varphi(j,t) + \Psi(j,t)+1\end{array}\right\} \begin{array}{l}\forall j=1..5, \\ \forall t=1,2,3\ldots\end{array} \quad (4)$$

Hàm Ψ(j,t) được định nghĩa theo qui tắc sau:

$$\left.\begin{array}{l}\Psi(1,t) = 1, \forall t=1,2,3\ldots \\ \Psi(2,t) = y^{(1)}_{\varphi(1,t)+19} \oplus 1 \\ \Psi(j+1,t) = \Psi(j,t) \oplus y^{(j)}_{\varphi(j,t)+19} \oplus 1, \forall j=2,3,4,\ldots\end{array}\right\} \quad (5)$$

Nhờ (3), (4), (5), chúng ta hoàn toàn tạo ra được dãy tuỳ ý các vectơ loạn số $\{\overrightarrow{\gamma_t}\}$, t=1,2,..

5.2. Phương pháp mã/dịch

Nếu chỉ tạo ra dãy giả ngẫu nhiên không thôi thì như 1 là đủ và sau đó chỉ việc dùng các tiêu chuẩn (5 tiêu chuẩn) để đánh giá chất lượng dãy giả ngẫu nhiên đó.

Nhưng ở đây để bảo mật thông tin chúng ta phải tìm cách "trộn" dãy giả ngẫu nhiên được sinh ra trong phần 7.1 với bản rõ để tạo ra các bản mã mà ngoài người có trách nhiệm ra không ai có thể đọc được các bản rõ đó vì họ không có các mầm khoá trong tay (tức là SK1 và SK2).

Trước hết, chúng ta thấy rằng: Trong bộ Code $N^{\circ}2$ của CCITT có 6 ký tự không được tham gia tạo ra các ký tự mã mà đó là các ký tự điều khiển. Cụ thể đó là 3, 4, 8, 9, +, /. Ta nói đó là các ký tự "cấm". Tập hợp các ký tự "cấm" như vậy được ký hiệu là tập hợp V. Vậy V = {3, 4, 8, 9, +, /}.

Các bộ mã nhị phân tương ứng với các ký tự bị cấm được cho trong bảng Code $N^{\circ}2$. Cụ thể:

3 ~ (00010). 4 ~ (01000). 8 ~ (11111). 9 ~ (00100). + ~ (11011). / ~ (00000).

Các giá trị của các ký tự bị cấm qua hàm $g_1()$ theo thứ tự là {8, 2, 31, 4, 27, 0} còn qua hàm $g_2()$ là {2, 8, 31, 4, 27, 0}.

Đứng trên quan điểm tập hợp thì 2 tập này trùng nhau còn đứng theo giá trị cụ thể của từng vectơ nhị phân thì chúng không giống nhau.

Để mã hoặc dịch một văn bản (bản text), chúng ta trước hết biến đổi các vectơ giả ngẫu nhiên $\overrightarrow{\gamma_t}$ là γ_t với t = {0, 1, 2, ..}. Tức là $\gamma_t = g_1(\overrightarrow{\gamma_t})$ (6)

Như vậy $\gamma_t \in C_{32} = \{0, 1, 2, .., 31\}$ ∀t=1,2,…

Đối với bản rõ (text) ta cũng làm như vậy. Cụ thể như sau. Giả sử có bản rõ $x = x_1x_2\ldots x_t\ldots$ với $x_t \in$ {a, b, c ,…, z}. t=1, 2, … qua bảng Code $N^{\circ}2$ ta chuyển thành các vecto nhị phân 5 thành phần.

Các vecto đó ta ký hiệu là $\overrightarrow{Zt}$, t = 1,2…

Như vậy $\overrightarrow{Zt}$ là vectơ nhị phân 5 thành phần đặt tương ứng 1-1 với các chữ cái x_t ở nhịp t. Bây giờ qua ánh xạ $g_2()$. Ta có:

$Z_t = g_2(\overrightarrow{Zt})$ (7)

Ta thay đổi tập hợp bị cấm V một chút như thế này:

V={0, 2, 8, 27, 31, 17} (8)

(số 17 ứng với ký tự Z trong bản rõ vì z ít xuất hiện nên ta dùng nó thay vì 4 - tức là thay vì 9). Những ký tự bị cấm này sẽ không có mặt trong bản mã hoặc bản rõ (trừ trường hợp ký tự 9 sẽ xuất hiện trong bản mã). Nếu ta sửa chữa bộ mã Code $N^{\circ}2$ đi một chút hoặc dùng bảng mã khác ta sẽ có các ký tự bị cấm khác, thậm chí không còn ký tự bị cấm nữa. Đây là một hướng mở mà ta sẽ nghiên cứu về sau.

Từ giả thiết các ký hiệu và theo lập luận ở trên ta có công thức mã/dịch như sau:

$T_t = Z_t + k_1 \gamma_t \pmod{32}$ (9).

Trong đó, Z_t , γ_t lần lượt được cho ở các công thức (6) và (7).

Ở công thức (9) còn một tham số k. Tham số này được xác định là số tự nhiên nhỏ nhất sao cho đối với nó:

$Z_t + k_1 \gamma_t \pmod{32} \notin V$ (10).

V ở đây được cho bởi (8).

Rõ ràng $T_t \in C_{32}$, $\forall$ t=1, 2, 3,… là số tương ứng với ký tự mã thứ t (với ký tự rõ Z_t và dãy loạn số γ_t). Nếu là mã còn nếu là dịch thì T_t tương ứng với ký tự rõ thứ t của bản rõ. Từ đây, các số T_t được biến đổi ngược lại qua hàm nghịch đảo của $g_2()$. Cụ thể là $g_2^{(-1)}(T_t)$, t=1, 2, 3,…

$\overrightarrow{Xt} = g_2^{(-1)}(T_t)$ v à k ý t ự tương ứng với code $\overrightarrow{Xt}$ chính là ký tự mã hoặc rõ tương ứng ở vị trí t nếu T_t không thuộc V. Còn nếu $T_t \in V$ thì ta lấy k=2 và sẽ có:

$T_t' = Z_t + 2 \gamma_t \pmod{32}$

Nếu T_t' không thuộc V thì dừng lại và đó là ký tự mã hoặc rõ ở nhịp t; còn nếu $T_t' \in V$ thì ta lại tiếp tục tăng k = 3 và tính

$T_t' = Z_t + 3 \gamma_t \pmod{32}$

Cho đến khi $T_t^{(i)}$ không thuộc V thì dừng. Trong thực tế 1<=k<=7.

5.3. Ví dụ

Giả sử ta có văn bản x= Secret personal cần được mã hoá. Để mã hoá ta tiến hành các bước sau:

Bước 1: Cho SK_1 và SK_2 và giả sử:

SK_1=JOMPC NXRIO MDHJX BVKFM.

SK_2 = CPWKV EEBSN.

Qua bảng mã quốc tế MTK_2 ta có:

- Đối với SK_1:

	J	O	M	P	C	N	X	R	I	O	M	D	H	J	X	B	V	K	F	M
1	1	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	1	1	0
2	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	0	1	1	0	0
3	0	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	1	1	1
4	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1
5	0	1	1	1	0	0	1	0	0	1	1	0	1	0	1	1	1	0	0	1

- Đối với SK_2:

	C	P	W	K	V	E	E	B	S	N
1	0	0	1	1	0	1	1	1	1	0
2	1	1	1	1	1	0	0	0	0	0
3	1	1	0	1	1	0	0	0	1	1

4	1	0	0	1	1	0	0	1	0	1
5	0	1	1	0	1	0	0	1	0	0

Tập bị cấm là: V={0, 2, 8, 27, 31, 1}.

Bước 2: Thiết lập các trạng thái ban đầu cho cả 5 thanh ghi R_1, …,R_5.

Áp dụng các công thức đã cho ở trên, ta có:

Bước 2.1. Tính các số C_i, i=1,..,10.

$C_1 = g_2(\overrightarrow{d_1})\text{-}g_1(\overrightarrow{b_{11}})=g_2(01110)\text{-}g_1(00111)(\text{mod}32)=14\text{-}28\ (\text{mod}32)=18.$

$C_2= g_2(\overrightarrow{d_2})\text{-}g_1(\overrightarrow{b_{12}})=13\text{-}9(\text{mod}32)=4.$

$C_3= g_2(\overrightarrow{d_3})\text{-}g_1(\overrightarrow{b_{13}})=25\text{-}20(\text{mod}32)=5.$

$C_4= g_2(\overrightarrow{d_4})\text{-}g_1(\overrightarrow{b_{14}})=30\text{-}11(\text{mod}32)=19.$

$C_5= g_2(\overrightarrow{d_5})\text{-}g_1(\overrightarrow{b_{15}})=15\text{-}29(\text{mod}32)=18.$

$C_6= g_2(\overrightarrow{d_6})\text{-}g_1(\overrightarrow{b_{16}})=16\text{-}25(\text{mod}32)=23.$

$C_7= g_2(\overrightarrow{d_7})\text{-}g_1(\overrightarrow{b_{17}})=16\text{-}30(\text{mod}32)=18.$

$C_8= g_2(\overrightarrow{d_8})\text{-}g_1(\overrightarrow{b_{18}})=19\text{-}15(\text{mod}32)=4.$

$C_9= g_2(\overrightarrow{d_9})\text{-}g_1(\overrightarrow{b_{19}})=20\text{-}13(\text{mod}32)=7.$

$C_{10}= g_2(\overrightarrow{d_{10}})\text{-}g_1(\overrightarrow{b_{20}})=6\text{-}28(\text{mod}32)=10.$

Bước 2.2. Từ đó, ta có (Theo bảng Code N≅ 2).

$$\overrightarrow{c_1} = g_2^{-1}(c_1) = g_2^{-1}(18) = 10010$$
$$\overrightarrow{c_2} = g_2^{-1}(c_2) = g_2^{-1}(4) = 00100$$
$$\overrightarrow{c_3} = g_2^{-1}(c_3) = g_2^{-1}(5) = 00101$$
$$\overrightarrow{c_4} = g_2^{-1}(c_4) = g_2^{-1}(19) = 10011$$
$$\overrightarrow{c_5} = g_2^{-1}(c_5) = g_2^{-1}(18) = 10010$$
$$\overrightarrow{c_6} = g_2^{-1}(c_6) = g_2^{-1}(23) = 10111$$
$$\overrightarrow{c_7} = g_2^{-1}(c_7) = g_2^{-1}(18) = 10010$$
$$\overrightarrow{c_8} = g_2^{-1}(c_8) = g_2^{-1}(4) = 00100$$
$$\overrightarrow{c_9} = g_2^{-1}(c_9) = g_2^{-1}(7) = 00111$$
$$\overrightarrow{c_{10}} = g_2^{-1}(c_{10}) = g_2^{-1}(10) = 01010.$$

Vậy trạng thái ban đầu của các thanh ghi từ R_1 đến R_5 là:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R_1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	1	1
R_2	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	0	1	1	0
R_3	1	0	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	1	1
R_4	1	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1
R_5	1	0	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	1	0	0

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
R_1	0	1	0	0	1	1	1	1	0	0	0	1	1	0	1	1	0	1
R_2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1
R_3	1	0	1	1	0	0	1	0	1	1	0	0	1	0	1	0	1	0
R_4	1	1	0	0	1	1	1	1	0	1	1	1	0	0	0	0	1	1
R_5	1	0	0	1	1	0	1	0	0	1	0	1	1	1	0	0	1	0

	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	...
R_1	0	0	1	0	0	0	0	1	1	1	1	0	0	1	1	0	1	...
R_2	0	1	0	0	1	0	0	0	0	0	1	1	0	1	1	0	1	...
R_3	1	1	1	1	0	1	0	0	0	0	1	1	1	0	1	0	1	...
R_4	1	0	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	...
R_5	1	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	1	...

Bước 3. Xây dựng dãy vectơ loạn số (vectơ giả ngẫu nhiên – psedorandom vectors).

Cho j=1,..,5 và t=1,...,15 (15 là độ dài bản rõ).

Theo các công thức (3), (4), (5) ta có:

$\varphi(j,1)=0\ \forall j=1,..,5.$

ψ(1,t)=1 ∀t=1,..,15.

Tiếp theo đối với thanh ghi R_1 (j=1) ta có:

φ(1,2)=φ(1,1)+1+ψ(1,1)=0+1+1=2.

φ(1,3)=φ(1,2)+1+ψ(1,2)=2+1+1=4.

φ(1,4)=φ(1,3)+1+ψ(1,3)=4+1+1=6.

φ(1,5)=φ(1,4)+1+ψ(1,4)=6+1+1=8.

….

Cuối cùng đối với R ta có bảng sau tương ứng với t=1,..,15.

R_1	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	ψ(1,t)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	φ(1,t)	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28

Tính toán tương tự đối với R_2, R_3, R_4 và R_5 là:

R_2	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	ψ(2,t)	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0
	φ(2,t)	0	1	3	5	6	7	9	11	12	13	15	17	18	20	22

R_3	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	ψ(3,t)	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0
	φ(3,t)	0	1	3	5	7	8	10	12	14	16	18	19	20	21	22

R_4	T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	ψ(4,t)	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0
	φ(4,t)	0	1	3	5	6	7	9	10	11	13	15	17	18	20	22

R_5	T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	ψ(5,t)	0	1	1	1	0	1	0	1	0	1	0	0	0	0	0
	φ(5,t)	0	1	3	5	7	8	10	11	13	14	16	17	18	19	20

Áp dụng công thức $\gamma_t^{(j)} = y_{\varphi(j,t)+27}^{(j)}$ j=1,..5.

Như vậy chu kỳ của hệ thống là $2^{31}.2^{31}.2^{31}.2^{31}.2^{31}-1=2^{155}-1$ chu kỳ.

Lúc này ta xác định được các vectơ loạn số $\overrightarrow{\gamma_t}$ với t=1,..,15. và giá trị tương ứng $g_1()$.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\gamma_t^{(1)}$	1	0	0	1	1	0	0	1	0	0	1	1	0	1	1
$\gamma_t^{(2)}$	0	0	0	1	1	1	0	1	0	1	0	0	0	0	1
$\gamma_t^{(3)}$	1	0	1	0	0	1	1	1	1	0	0	0	0	0	1
$\gamma_t^{(4)}$	1	1	1	1	0	0	0	1	1	0	0	0	0	1	1
$\gamma_t^{(5)}$	1	0	1	1	1	0	1	0	1	0	0	0	1	0	1
$g_1()$	29	8	28	27	19	6	20	15	28	2	1	1	16	9	31

Bước 4. Mã hoá (hoặc dịch)

Xác định giá trị $g_2()$ của bản rõ.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Rõ	S	E	C	R	E	T		P	E	R	S	O	N	A	L
$g_2()$	20	16	14	10	16	1	4	13	16	10	20	3	6	24	9

Những giá trị cấm V={0,17,8,2,31,27}.

Xét k=1:

$T_1= z_1+k\gamma_1$ mod32= 20+29 (mod32)=17 $\in$V -> Tăng k=2 thì T_1=14 $\notin$V ký tự mã sẽ là $g_2^{(-1)}(T_1) = g_2^{(-1)}(14)$ = 01110=C.

$T_2= z_2+k\gamma_2$ mod32= 16+8 (mod32)=24 $\notin$V -> Ký tự mã T_2 sẽ là $g_2^{(-1)}(T_2) = g_2^{(-1)}(24)$= 11000=A.

Tương tự, ta sẽ tính được các ký tự còn lại. Vậy bản mã sẽ là: CARHOMAUIIY9FTM.

Bước 5. Giải mã

Thực hiện tương tự như bước mã hoá.

5.4. Một số tiếu chuẩn để đánh giá mức độ an toàn của khoá mã.

5.4.1. Giới thiệu chung

Thiết bị tạo ngẫu nhiên được thiết kế nhằm đưa ra một dãy các biến ngẫu nhiên nhị phân (nhận 2 giá trị là 0 và 1) độc lập có phân bố đối xứng, nghĩa là nó được xây dựng để thực hiện các giá trị là nguồn đối xứng nhị phân (Binary Symmetric Source - viết tắt BSS).

Những thiết bị (phần cứng hoặc phần mềm) tạo bít ngẫu nhiên có nhiều ứng dụng trong mật mã học, kiểm tra các thuật toán xác suất cũng như trong nhiều lĩnh vực khác. Ứng dụng chủ yếu của chúng ta trong mật mã là việt kiểm tra các nguồn tạo khóa bí mật hoặc công khai (Public Key) hoặc kiểm tra dòng nhị phân sinh ra từ khóa dòng (Key stream) (O.T.P).

Trong áp dụng đó, việc an toàn mật mã phụ thuộc vào tính ngẫu nhiên của dãy loạn số được tạo ra bởi thuật toán nào đó (hoặc các thiết bị nào đó). Dãy loạn số ấy có độ dài có thể coi là vô hạn (hoặc hữu hạn nhưng rất lớn), nguồn tạo loạn số ấy được coi là độc lập thống kê nếu tất cả các phần tử của nguồn ấy là độc lập nhau. Tuy nhiên việc kiểm tra để kết luận nguồn ấy là độc lập thống kê là việc không thể thực hiện được trong thực tiễn.

Vì vậy vấn đề đặt ra là có thể kiểm tra một số mẫu đủ nhỏ trong đó mà kết luận được toàn bộ nguồn ấy là độc lập hay không? tức là có thể kiểm tra một tập hợp rất hữu hạn mà kết luận cho tập hợp vô hạn được không? Nhờ lý xác suất mà nổi bật là các định lý Konmogorov, ta có câu trả lời "Có thể được".

Để giải quyết vấn đề "Có thể được", người ta đã nghiên cứu và đưa ra lý thuyết chọn mẫu, lý thuyết xác suất thống kê ứng dụng, và trong thực tế, lý thuyết này rất có hiệu lực. Ở đây tôi chỉ nghiên cứu và trình bày một số tiêu chuẩn thống kê nhằm kiểm tra những mẫu được lấy từ một tập hợp có phân bố nhị phân.

5.4.2. Các tiêu chuẩn kiểm tra cơ bản..

5.4.2.1. Một số các khái niệm về lý thuyết xác suất:

5.4.2.2. 5 tiêu chuẩn kiểm tra thống kê

Cho một chuỗi nhị phân có độ dài là n: $s=s_0, s_1,..., s_{n-1}$. Chúng ta sẽ dùng 5 tiêu chuẩn kiểm tra thống kê thường dùng để kiểm tra chuỗi nhị phân s.

(i) frequency test (kiểm tra tần số)

Mục đích của kiểm tra này đó là để xác định số lần mà số 0 hoặc số 1 trong chuỗi nhị phân s xuất hiện ngẫu nhiên trong chuỗi. Lấy n_0, n_1 lần lượt ký hiệu là số lượng các số 0 và 1 xuất hiện trong s. Công thức thống kê sẽ là:

$$X_1 = \frac{(n_0 - n_1)^2}{n}$$

Nếu dãy s là ngẫu nhiên thì X_1 có phân bố xấp xỉ phân bố χ^2 (Chi - bình phương) với 1 bậc tự do (n>=10 và n càng lớn sự xấp xỉ càng tốt).

(ii) Serial test (kiểm tra bộ đôi)

Mục đích của kiểm tra này là xác định số lượng tần số xuất hiện cảu các bộ đôi 00, 01,10,11 có ngẫu nhiên trong chuỗi nhị phân s hay không. Lấy n_0, n_1 ký hiệu cho số lượng các số 0 và 1 xuất hiện trong s, và lấy n_{00}, n_{01}, n_{10}, n_{11} ký hiệu các số 00, 01, 10, 11 xuất hiện trong s. Chú ý rằng: $n_{00} + n_{01} + n_{10} + n_{11} = (n-1)$.

Công thức thống kê sẽ là:

$$X_2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1$$

Nếu dãy s là ngẫu nhiên (là "tốt") thì X_2 có phân bố xấp xỉ phân bố χ^2 với 2 bậc tự do nếu n>=21.

(iii) Poker test (Kiểm tra Poker).

Lấy m là một số nguyên dương sao cho $\left|\frac{n}{m}\right| \geq 5.\left(2^m\right)$, và lấy k= $\left|\frac{n}{m}\right|$. Chia chuỗi s thành k phần không gối nhau, mỗi phần có độ dài là m, n_i là số lần xuất hiện thứ i của các thứ tự có độ dài m, $1 \leq i \leq 2^m$. Kiểm tra poker sẽ xác định tần suất xuất hiện dãy số độ dài m trong s.

Công thức thống kê là:

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k$$

sẽ có phân bố xấp xỉ phân bố χ^2 với 2^m-1 bậc tự do nếu dãy s là ngẫu nhiên. Chú ý rằng kiểm tra poker là sự tổng quát hoá của các bước kiểm tra ở (i) và (ii).

(iv) Runs test (Kiểm tra theo loạt).

Mục đích của kiểm tra này là dựa vào số lần xuất hiện các loạt có các độ dài khác nhau trong dãy s.

Một dãy con cực đại gồm các số 0 đứng liền nhau thì được gọi là một Gap.

Một dãy con cực đại gồm các số 1 đứng liền nhau thì được gọi là một Block.

Số các Gaps (hoặc các Blocks) có độ dài i trong một chuỗi ngẫu nhiên với độ dài n là $e_i=(n-i+3)/2^{i+2}$. Đặt k bằng số nguyên i lớn nhất sao cho e_i>=5. Lấy B_i, G_i là số lượng các Blocks hoặc các Gaps có độ dài i với mỗi i, 1<=i<=k.

Công thức thống kê là:

$$X_4 = \sum_{i=1}^{k} \frac{\left(B_i - e_i\right)^2}{e_i} + \sum_{i=1}^{k} \frac{\left(G_i - e_i\right)^2}{e_i}$$

Nếu dãy s là ngẫu nhiên thì X_4 có phân bố xấp xỉ phân bố χ^2 với 2k-2 bậc tự do.

(v) Autocorrelation test (Kiểm tra sự tương quan).

Mục đích của kiểm tra này là để kiểm tra sự tương quan giữa các bít của dãy. Lấy d là một số nguyên dương cố định, $1 \leq d \leq [n/2]$. Số lượng các bít trong s không bằng d bước dịch chuyển là $A(d) = \sum_{i=0}^{n-d-1} s_i \oplus s_{i+d}$. Với $\oplus$ là phép toán XOR. Công thức thống kê sẽ là:

$$X_5 = 2\left(A(d) - \frac{n-d}{2}\right) / \sqrt{n-d}$$

Nếu s là không tương quan, tức dãy s độc lập và không có chu kỳ thì X_5 có phân bố xấp xỉ phép một phân bố N(0,1) với $n-d \geq 10$. [5]

5.4.3. Ứng dụng các tiêu chuẩn kiểm tra thống kê để đánh giá chất lượng của dãy giả ngẫu nhiên do thuật toán được đề xuất tạo ra

Xem xét dãy số giả ngẫu nhiên được tạo ra do thuật toán mã/dịch dòng ký tự "Secret personal" để đánh giá chất lượng dãy số đó theo 5 tiêu chuẩn kiểm tra trên. Ta có dãy số nhị phân giả ngẫu nhiên là:

10111 00010 00111 11011 11001 01100 00101 11110 00111 01000 10000 10000 00001 10010 11111.

Như vậy dãy số nhị phân giả ngẫu nhiên có độ dài n=75. Áp dụng các tiêu chuẩn kiểm tra cơ bản ở trên ta có:

(i). frequency test. n_0=38, n_1=37, giá trị thống kê X_1 là 0.013.

(ii). Serial test. n_{00}=23, n_{01}=15, n_{10}=15, n_{11}= 21, giá trị thống kê X_2 là 2.17.

(iii). Poker test. Chọn m=2 vì $\left\lfloor \frac{n}{m} \right\rfloor \geq 5.(2)^m$, như vậy ta có k=37, giá trị thống kê X_3 là: -15.16.

(iv). Run test. Ở đây e_1=9.625 và k=1. Chúng ta có 8 blocks có độ dài là 1 và 6 gaps có độ dài là 1. Giá trị thống kê là X_4=1.63.

Với mức α=0.05, giá trị ngưỡng cho X_1, X_2, X_3 và X_4 là: 3.8415, 5.9915, 14.0671, 9.4877 (Xem tại bảng 1 và bảng 2). Như vậy, dãy số giả ngẫu nhiên s được đánh giá là tốt.

5.4.4. Đánh giá độ an toàn về thuật toán mã/dịch

Trên đây là những tiêu chuẩn thống kê để đánh giá chất lượng dãy loạn số (tức là dãy số giả ngẫu nhiên) do các thuật toán tạo ra. Các tiêu chuẩn này đã được Tiêu chuẩn hoá quốc tế. Tuy nhiên, để đánh giá độ an toàn mật mã do thuật toán mã/dịch vừa thiết kế trên, có hai câu hỏi được đặt ra:

Câu hỏi 1: Với khoá SK_1, SK_2 cố định thì chu kỳ của dãy giả ngẫu nhiên do thuật toán tạo ra có độ dài cực đại là bao nhiêu ?

Trong trường hợp này, với SK_1 và SK_2 cố định trước, thông thường chỉ có 5 thanh ghi dịch $R_1 \div R_5$ phản hồi tuyến tính thì chu kỳ cực đại của nó là tích các độ dài của chúng trừ đi 1 nếu chúng là các số nguyên tố, nghĩa là nếu gọi $l(R_i)$ i=1,..,5. là độ dài của 5 thanh ghi dịch là $l(R_i)=P_i$ với P_i là số nguyên tố (i=1,..,5) thì chu kỳ cực đại của nó là: $\prod_{i=1}^{5} P_i - 1$. Ở đây trong thuật toán trên, chúng tôi chọn P_i=31 $\forall$i i=1,..,5 và do đó bình thường chu kỳ cực đại là $K=31^5 - 1 = 28629150$. Tuy nhiên, thuật toán trên đã được đưa vào các hàm toán học khá phức tạp và do đó làm tăng độ dài của chu kỳ là:

$K \approx 2^{31^5 - 1} - 1 = 2^{2862910}$. Theo tôi thì đó là một chu kỳ rất lớn và chúng ta không lo lặp lại chu kỳ trong việc mã một văn bản tuỳ ý - kể cả việc mã hoá một cuốn từ điển dày đến 3000 trang. Nhưng về lý thuyết thì dù độ dài chu kỳ đến bao nhiêu mà dùng mãi cũng sẽ có lúc lặp lại. Vì vậy, trong thuật toán mã/dịch của mình, thuật toán có hai hệ thống khoá là SK_1 và SK_2. SK_1 thì tương đối ổn định (có thể dùng trong một thời gian dài mới thay khoá một lần) còn SK_2 được thay đổi mỗi khi mã hoá một bản thông báo mới.

Câu hỏi 2: Trong trường hợp thuật toán bị lộ thì độ phức tạp của thuật toán trở về độ khó của việc xác định các mầm khoá (SK_1 và SK_2). Vậy không gian khoá lớn đến mức nào để có thể đối phó với kỹ thuật "vét cạn" khoá của đối phương (với giả thiết đối phương có thiết bị máy tính mạnh).

Trong trường hợp này, trên thực tế nếu thuật toán mã dịch không được "cứng hoá" nhờ một thiết bị phần cứng mã/dịch (máy mã) tức là thuật toán mã/dịch được "gắn" vào thiết bị mã/dịch chuyên dụng thì rất có thể thuật toán sẽ bị lộ. Trong trường hợp đó, độ an toàn mật mã chỉ trong đợi vào các hệ thống khoá SK_1, SK_2. Vì SK_1 gồm 20 ký tự được chọn ngẫu nhiên từ bảng 26 chữ cái la tinh nên không gian của khoá SK_1 26^{20}. Tương tự như vậy đối với SK_2 là 26^{10} → Vậy không gian khoá của SK_1 và SK_2 là: $26^{20} * 26^{10} = 26^{30} \approx 2{,}4*10^{42}$.

Đó là một con số đủ lớn để đảm bảo chống lại sự "vét cạn" của đối phương.

KẾT LUẬN

Kết luận

Hiện nay, có nhiều tri thức được phát triển mạnh mẽ về kỹ thuật bảo vệ an ninh thông tinhằm chống lại các mối đe doạ mất an toàn trên mạng máy tính và một nền tảng lý thuyết đầy đủ để suy luận về tính hiệu quả của chúng thì đang trong tiến trình phát triển. Các kỹ thuật đều dựa trên việc sử dụng mật mã. Chúng không phải chỉ để che dấu thông tin mà còn để xác nhận thông tin.

Các phương pháp mã hoá dữ liệu được chia thành hai lớp:

- Mã hoá bằng khoá bí mật như thuật toán DES, AES.
- Mã hoá bằng khoá công khai, RSA.

Mã hoá bằng khoá công khai thuận tiện hơn mã hoá bằng khoá bí mật vì nó tránh được việc truyền khoá bí mật trên mạng. Nhưng sự thực hiện nó thì chậm và kém hiệu quả hơn nhiều so với mã hoá bằng khoá bí mật. Vì lý do đó, mã hoá bằng khoá công khai thường chỉ được sử dụng để truyền các khoá bí mật, rồi sau đó các khoá này lại được dùng trong các thuật toán khoá bí mật để mã hoá hoặc giải mã các thông điệp.

Trong các thuật toán mã hoá, kể cả khoá bí mật và khoá công khai, do độ dài của các thông điệp cần mã hoá là rất lớn. Để tối ưu cho các thuật toán mã hoá thông thường các thông điệp được chia thành các khối với kích thước phù hợp cho việc mã hoá. Các khối đó có thể được mã hoá độc lập hay mã hoá liên tiếp nhau tuỳ thuộc vào phương pháp mã hoá của thuật toán đó. Mã hoá theo khối sẽ làm tăng tốc độ mã hoá, nó có thể đáp ứng cho việc mã dữ liệu, tiếng nói.

Sau một thời gian nghiên cứu thử nghiệm, luận văn đã có được những kết quả nhất định:

- Nghiên cứu thực trạng an toàn thông tin trên mạng máy tính, các nguy cơ mất an toàn thông tin cũng như việc bảo vệ các thông tin bằng kỹ thuật mã hoá.
- Nghiên cứu lý thuyết về các thuật toán mã hoá, cả mã hoá bằng khoá bí mật và mã hoá bằng khoá công khai. Cùng các cách thức để tạo các dòng khoá giả ngẫu nhiên.
- Nghiên cứu một số các thuật toán mã khối đạng được sử dụng rộng rãi hiện nay như DES, PGP, AES,... cũng như vai trò của nó trong việc mã hoá dữ liệu.
- Nghiên cứu và đưa ra thuật toán hoạt động dựa trên các thanh ghi phản hồi phi tuyến nhằm sinh ra dòng bít giả ngẫu nhiên, để ứng dụng cho việc mã hoá dữ liệu.
- Nghiên cứu một số các tiêu chuẩn thống kê để kiểm tra các dãy bít giả ngẫu nhiên được sinh ra từ thuật toán nào đó.
- Xây dựng chương trình mã khối đáp ứng yêu cầu mã hoá dữ liệu.

Hướng nghiên cứu tiếp theo.

- Trong thời gian tới, tiếp tục hoàn thiện chương trình mã hoá để phục vụ cho các yêu cầu mã hoá dữ liệu tại cơ quan.
- Nghiên cứu các giao thức trao đổi khoá để ứng dụng cho chương trình mã hoá.

Dù đã đạt được những kết quả bước đầu, tuy nhiên sẽ không thể tránh khỏi những điểm chưa thật hợp lý. Do vậy, em rất mong được sự góp ý, bổ sung của các thầy/cô để tiếp tục hoàn thiện thuật toán cũng như chương trình.

Em xin chân thành cảm ơn./.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Nguyễn Đình Diệu (2000), Giáo trình lý thuyết mật mã, NXB Đại học Quốc gia, Hà Nội.
2. Thái Hồng Nhị, Phạm Thị Việt (2004), "*An toàn thông tin*", NXB Khoa học và Kỹ thuật.
3. Đinh Thế Kỷ (2001), "*Một số suy nghĩ bước đầu về giải pháp bảo mật, an toàn thông tin trong hệ thống thương mại điện tử của Việt Nam*", Kỷ yếu tuần lễ tin học X, trang 71.

Tiếng Anh

4. Bruce Schneier, Jonh Wiley & Sons, "*Applied Cryptography: Protocols, Algorithms, and Source Code in C*", ISBN: 0-471-12845-7
5. A. Menezes, P. van Oorschot, and S. Vanstone, "*Handbook of Applied Cryptography*" CRC Press, 1996.
6. Gil Held , "*Learn Encryption Techniques with BASIC and C++*", ISBN: 1556225989. Publication Date: 10/01/98 .
7. Charlie Kaufman, Radia Perlman, Mike Speciner, "*Network Security*", Prentice Hall PTR 2002.

Các Website

8. www.google.com
9. www.schneier.com
10. www.rsasecurity.com
11. www.en.wikipedia.org
12. www.williamstallings.com
13. http://csrc.nist.gov
14. www2.mat.dtu.dk