ĐẠI HỌC QUỐC GIA HÀ NỘI

KHOA CÔNG NGHỆ

Phan Đình Diệu

Lý thuyết mật mã
&
AN TOÀN THÔNG TIN

NXB ĐẠI HỌC QUỐC GIA HÀ NỘI - 2002

Lý thuyết mật mã

&

An toàn thông tin

Lý thuyết mật mã
&
An toàn thông tin

Phan Đình Diệu

Đại học Quốc gia Hà Nội

Khoa Công nghệ- ĐHQG Hà nội

NỘI DUNG

Chương 3

Các hệ mật mã khoá đối xứng 55

Chương 4

Các hệ mật mã khoá công khai 92

Chương 5

Bài toán xác nhận và Chữ ký điện tử.....115

Lời mở đầu

Từ khi con người có nhu cầu trao đổi thông tin, thư từ cho nhau thì nhu cầu giữ bí mật và bảo vệ tính riêng tư của những thông tin, thư từ được trao đổi đó cũng nẩy sinh. Hình thức thông tin được trao đổi phổ biến và sớm nhất là dưới dạng các văn bản, để giữ bí mật của thông tin người ta đã sớm nghĩ đến cách che dấu nội dung các văn bản bằng cách biến dạng các văn bản đó để người ngoài không đọc hiểu được, đồng thời có cách khôi phục lại nguyên dạng ban đầu để người trong cuộc vẫn đọc hiểu được; theo cách gọi ngày nay thì dạng biến đổi của văn bản được gọi là *mật mã* của văn bản, cách lập mật mã cho một văn bản được gọi là *phép lập mật mã*, còn cách khôi phục lại nguyên dạng ban đầu của văn bản từ bản mật mã được gọi là *phép giải mã*. Phép lập mật mã và phép giải mã được thực hiện nhờ một chìa khoá riêng nào đó mà chỉ những người trong cuộc được biết, sau đây ta sẽ gọi là *khoá mật mã*. Người ngoài cuộc không được biết khoá mật mã, nên dù có "ăn cắp" được bản mật mã trên đường truyền tin, về nguyên tắc cũng không thể giải mã để hiểu được nội dung của văn bản truyền đi.

Hiển nhiên, tiêu chuẩn của một bản mật mã là tạo được tính bí mật cho văn bản; vì vậy khái niệm *bí mật* là khái niệm cốt lõi nhất đối với một lý thuyết về mật mã. Có thể có một định nghĩa khoa học cho khái niệm *bí mật* hay không? Đã có nhiều cách tiếp cận để tìm hiểu nội dung của khái niệm bí mật, nhưng một định nghĩa khoa học, hay hơn nữa, một định nghĩa toán học cho khái niệm đó thì chưa có. Một cách tiếp cận khá phổ biến là gắn khái niệm bí mật với khái niệm "ngẫu nhiên", nếu một văn bản rõ có một nội dung xác định thì điều ta mong muốn là bản mật mã của nó phải là một bản gồm các ký tự được sắp xếp hỗn độn, có vẻ như ngẫu nhiên khiến

người ngoài nhìn vào không thể xác định được nội dung của văn bản gốc. Tuy nhiên, nếu "bí mật" là khái niệm chưa định nghĩa được, thì khái niệm "ngẫu nhiên", hay cụ thể hơn, khái niệm "dãy bit ngẫu nhiên", cũng khó định nghĩa như vậy, ta chưa qui định được một tiêu chuẩn toán học để xác định một dãy bit có là "ngẫu nhiên" hay không, mà chỉ mới tìm hiểu được một số thuộc tính gần với "ngẫu nhiên", dùng làm căn cứ để tạm xác định một dãy bit có là "giả ngẫu nhiên" theo nghĩa có các thuộc tính đó hay không mà thôi.

Từ mấy thập niên gần đây, bước vào kỷ nguyên máy tính, cũng như đối với nhiều lĩnh vực khác, lĩnh vực mật mã cũng đã có những chuyển biến to lớn từ giai đoạn mật mã truyền thống sang giai đoạn *mật mã máy tính;* máy tính điện tử được sử dụng ngày càng phổ biến trong việc lập mật mã, giải mật mã, và những chuyển biến đó đã kích thích việc nghiên cứu các giải pháp mật mã, biến việc nghiên cứu mật mã thành một khoa học có đối tượng ngày càng rộng lớn và được sử dụng có hiệu quả trong nhiều phạm vi hoạt động của cuộc sống. Vì các nghiệp vụ chủ yếu của mật mã được thực hiện bằng máy tính, nên các khái niệm bí mật, ngẫu nhiên cũng dần được "máy tính hoá", và với sự ra đời của *Lý thuyết về độ phức tạp tính toán* vào giữa những năm 1960, các khái niệm đó tìm được một nội dung chung có thể được nghiên cứu một cách toán học là tính *phức tạp*. Bây giờ ta có thể nói, một bản mật mã đối với anh là *bí mật*, nếu từ bản mật mã đó để tìm ra bản rõ anh phải thực hiện một tiến trình tính toán mà độ phức tạp của nó vượt quá mọi năng lực tính toán (kể cả mọi máy tính) của anh; một dãy bit có thể xem là *ngẫu nhiên* , nếu dựa vào một đoạn bit đã biết để tìm một bit tiếp theo của dãy anh cũng phải thực hiện một tiến trình tính toán có độ phức tạp cực lớn tương tự như nói trên.

Việc chuyển sang giai đoạn mật mã máy tính trước hết đã có tác dụng phát triển và hiện đại hoá nhiều hệ thống mật mã theo kiểu truyền thống, làm cho các hệ thống đó có các cấu trúc tinh tế hơn, đòi hỏi lập mật mã và giải mã phức tạp hơn, do đó hiệu quả giữ bí mật của các giải pháp mật mã được nâng cao hơn trước rất nhiều. Tuy nhiên, một bước chuyển có tính chất cách mạng mà mật mã máy tính mang lại là việc phát minh ra các hệ mật mã *có khoá công khai*, bắt đầu từ cuối những năm 1970, cơ sở lý thuyết của các phát

minh đó là sự tồn tại của các *hàm một phía* (one-way function), tức là những hàm số số học $y = f(x)$ mà việc tính theo phía thuận từ x tính y là tương đối dễ, nhưng việc tính theo phía ngược từ y tìm lại x ($x = f^{-1}(y)$) là cực kỳ phức tạp. Các hệ mật mã có khoá công khai đã làm thay đổi về bản chất việc tổ chức các hệ truyền thông bảo mật, làm dễ dàng cho việc bảo mật trên các hệ truyền thông công cộng, và do tính chất đặc biệt đó chúng đã là cơ sở cho việc phát triển nhiều giao thức an toàn thông tin khác khi sử dụng mạng truyền thông công cộng, chẳng hạn các loại giao thức về xác nhận nguồn tin và định danh người gửi, chữ ký điện tử, các giao thức xác nhận không để lộ thông tin gì khác ngoài việc xác nhận, các giao thức trao đổi khoá trong tổ chức truyền tin bảo mật và trong xác nhận, v.v..., và gần đây trong việc phát triển nhiều giao thức đặc thù khác trong các giao dịch ngân hàng và thương mại điện tử, phát hành và mua bán bằng tiền điện tử,... Cũng cần nói thêm là lý thuyết mật mã hiện đại, tức là mật mã máy tính trên cơ sở lý thuyết về độ phức tạp tính toán tuy có nhiều ứng dụng đặc sắc và có triển vọng to lớn, nhưng cũng mới đang trong giai đoạn phát triển bước đầu, còn phải khắc phục nhiều khó khăn và tìm kiếm thêm nhiều cơ sở vững chắc mới để tiếp tục hoàn thiện và phát triển. Chẳng hạn, như trên đã nói, một cơ sở quan trọng của lý thuyết mật mã hiện đại là sự tồn tại của các hàm một phía, nhưng ngay có thật tồn tại các hàm một phía hay không cũng còn là một bài toán chưa có câu trả lời! Ta chỉ mới *đang có* một số hàm một phía *theo sự hiểu biết của con người hiện nay*, nhưng chưa chứng minh được có một hàm cụ thể nào đó *chắc chắn* là hàm một phía! Tuy nhiên, nếu theo quan điểm khoa học hiện đại, ta không xem mục đích khoa học là đi tìm những chân lý chắc chắn tuyệt đối, mà là đi tìm những cách giải quyết vấn đề (problem solving) gặp trong thực tiễn, thì ta vẫn có thể tin vào những giải pháp "tương đối" rất có hiệu quả mà lý thuyết hiện đại về mật mã đang cống hiến cho con người hiện nay.

Tập giáo trình *Lý thuyết mật mã và an toàn thông tin* này được soạn để phục vụ cho việc học tập của sinh viên các lớp theo chương trình đại học hoặc cao học thuộc ngành Công nghệ thông tin của Đại học Quốc gia Hà nội. Trong khoảng mươi năm gần đây, trên thế giới đã xuất hiện nhiều sách và tài liệu có tính chất giáo khoa

hoặc tham khảo về lý thuyết mật mã hiện đại và ứng dụng. Người viết tập giáo trình này chỉ có cố gắng lựa chọn và sắp xếp một số nội dung mà mình nghĩ là cần thiết và thích hợp nhất để trong một phạm vi hạn chế về thời gian (và không gian) trình bày và giới thiệu được cho người học một cách tương đối hệ thống những kiến thức cơ bản về lý thuyết mật mã hiện đại, bao gồm cả một số kiến thức toán học cần thiết. Giáo trình này đã được giảng dạy cho sinh viên các khoá cao học về Công nghệ thông tin thuộc Đại học Bách khoa Hà nội và khoa Công nghệ Đại học Quốc gia Hà nội từ năm 1997 đến 2004. Người viết chân thành cảm ơn các bạn đồng nghiệp và người đọc chỉ cho những chỗ thiếu sót để có thể kịp thời sửa chữa cho những lần in sau, nếu có.

Tháng 12 năm 2002

Phan Đình Diệu

CHƯƠNG I

Giới thiệu chung về mật mã

1.1. Sơ lược lịch sử về mật mã.

Như đã giới thiệu trong *Lời mở đầu*, nhu cầu sử dụng mật mã đã xuất hiện từ rất sớm, khi con người biết trao đổi và truyền đưa thông tin cho nhau, đặc biệt khi các thông tin đó đã được thể hiện dưới hình thức ngôn ngữ, thư từ. Lịch sử cho ta biết, các hình thức mật mã sơ khai đã được tìm thấy từ khoảng bốn nghìn năm trước trong nền văn mịnh Ai cập cổ đại. Trải qua hàng nghìn năm lịch sử, mật mã đã được sử dụng rộng rãi trên khắp thế giới từ Đông sang Tây để giữ bí mật cho việc giao lưu thông tin trong nhiều lĩnh vực hoạt động giữa con người và các quốc gia, đặc biệt trong các lĩnh vực quân sự, chính trị, ngoại giao. Mật mã trước hết là một loại hoạt động thực tiễn, nội dung chính của nó là để giữ bí mật thông tin (chẳng hạn dưới dạng một văn bản) từ một người gửi A đến một người nhận B, A phải tạo cho văn bản đó một bản mã mật tương ứng, và thay vì gửi văn bản rõ thì A chỉ gửi cho B bản mã mật, B nhận được bản mã mật và sẽ có cách từ đó khôi phục lại văn bản rõ để hiểu được thông tin mà A muốn gửi cho mình. Vì bản gửi đi thường được chuyển qua các con đường công khai nên người ngoài có thể "lấy trộm" được, nhưng do đó là bản mật mã nên không đọc hiểu được, còn A có thể tạo ra bản mã mật và B có thể giải bản mã mật thành bản rõ để hiểu được là do giữa hai người đã có một thỏa thuận về một *chìa khóa chung*, chỉ với chìa khóa chung này thì A mới tạo được bản mã mật từ bản rõ, và B mới từ bản mã mật khôi phục lại được bản rõ. Sau này ta sẽ gọi đơn giản chìa khóa chung đó là *khóa mật mã.* Tất nhiên để thực hiện được một phép mật mã, ta

còn cần có một thuật toán biến bản rõ, cùng với khóa mật mã, thành bản mã mật, và một thuật toán ngược lại, biến bản mã mật, cùng với khóa mật mã, thành bản rõ. Các thuật toán đó được gọi tương ứng là thuật toán *lập mật mã* và thuật toán *giải mật mã*. Các thuật toán này thường không nhất thiết phải giữ bí mật, mà cái cần được giữ tuyệt mật luôn luôn là khóa mật mã. Trong thực tiễn, đã có hoạt động bảo mật thì cũng có hoạt động ngược lại là khám phá bí mật từ các bản mã mật "lấy trộm" được, ta thường gọi hoạt động này là *mã thám*, hoạt động này quan trọng không kém gì hoạt động bảo mật! Vì các thuật toán lập mật mã và giải mật mã không nhất thiết là bí mật, nên mã thám thường được tập trung vào việc tìm khóa mật mã, do đó cũng có người gọi công việc đó là *phá khóa.*

Suốt mấy nghìn năm lịch sử, các thông báo, thư từ được truyền đưa và trao đổi với nhau thường là các văn bản, tức là có dạng các dãy ký tự trong một ngôn ngữ nào đó; vì vậy, các thuật toán lập mật mã thường cũng đơn giản là thuật toán xáo trộn, thay đổi các ký tự được xác định bởi các phép chuyển dịch, thay thế hay hoán vị các ký tự trong bảng ký tự của ngôn ngữ tương ứng; khóa mật mã là thông tin dùng để thực hiện phép lập mật mã và giải mật mã cụ thể, thí dụ như số vị trí đối với phép chuyển dịch, bảng xác định các cặp ký tự tương ứng đối với phép thay thế hay hoán vị,... Mật mã chưa phải là một khoa học, do đó chưa có nhiều kiến thức sách vở để lại, tuy nhiên hoạt động bảo mật và thám mã trong lịch sử các cuộc đấu tranh chính trị, ngoại giao và quân sự thì hết sức phong phú, và mật mã đã có nhiều tác động rất quan trọng đưa đến những kết quả lắm khi có ý nghĩa quyết định trong các cuộc đấu tranh đó. Do trong một thời gian dài, bản thân hoạt động mật mã cũng được xem là một bí mật, nên các tài liệu kỹ thuật về mật mã được phổ biến đến nay thường chỉ ghi lại các kiến thức kinh nghiệm, thỉnh thoảng mới có một vài "phát minh" như các hệ mật mã Vigenère vào thế kỷ 16 hoặc hệ mật mã Hill ra đời năm 1929 là các hệ mã thực hiện phép chuyển dịch (đối với mã Vigenère) hay phép thay thế (mã Hill) đồng thời trên một nhóm ký tự chứ không phải trên từng ký tự riêng rẽ. Vấn đề thám mã, ngược lại, khi thành công thường đưa đến những cống hiến nổi trội và ấn tượng trong những

tình huống gay cấn của các cuộc đấu tranh, và cũng thường đòi hỏi nhiều tài năng phát hiện với những kinh nghiệm và suy luận tinh tế hơn, nên để lại nhiều chuyện hấp dẫn hơn. Nhiều câu chuyện kỳ thú của lịch sử thám mã đã được thuật lại trong quyển sách nổi tiếng của David Kahn *The Codebreakers . The Story of Secret Writing* , xuất bản năm 1967 (sách đã được dịch ra nhiều thứ tiếng, có bản dịch tiếng Việt *Những người mã thám*, 3 tập, xuất bản tại Hà nội năm 1987).

Bước sang thế kỷ 20, với những tiến bộ liên tục của kỹ thuật tính toán và truyền thông, ngành mật mã cũng đã có những tiến bộ to lớn. Vào những thập niên đầu của thế kỷ, sự phát triển của các kỹ thuật biểu diễn, truyền và xử lý tín hiệu đã có tác động giúp cho các hoạt động lập và giải mật mã từ thủ công chuyển sang cơ giới hóa rồi điện tử hóa. Các văn bản, các bản mật mã trước đây được viết bằng ngôn ngữ thông thường nay được chuyển bằng kỹ thuật số thành các dãy tín hiệu nhị phân, tức các dãy bit, và các phép biến đổi trên các dãy ký tự được chuyển thành các phép biến đổi trên các dãy bit, hay các dãy số, việc thực hiện các phép lập mã, giải mã trở thành việc thực hiện các hàm số số học. Toán học và kỹ thuật tính toán bắt đầu trở thành công cụ cho việc phát triển khoa học về mật mã. Khái niệm trung tâm của khoa học mật mã là khái niệm *bí mật*. Đó là một khái niệm phổ biến trong đời sống, nhưng liệu có thể cho nó một nội dung có thể định nghĩa được một cách toán học không? Như đã lược qua trong *Lời mở đầu,* khái niệm *bí mật* thoạt đầu được gắn với khái niệm *ngẫu nhiên*, rồi về sau trong những thập niên gần đây, với khái niệm *phức tạp*, cụ thể hơn là khái niệm *độ phức tạp tính toán*. Việc sử dụng lý thuyết xác suất và ngẫu nhiên làm cơ sở để nghiên cứu mật mã đã giúp C.Shannon đưa ra khái niệm *bí mật hoàn toàn* của một hệ mật mã từ năm 1948, khởi đầu cho một lý thuyết xác suất về mật mã. Trong thực tiễn làm mật mã, các *dãy bit ngẫu nhiên* được dùng để trộn với bản rõ (dưới dạng một dãy bit xác định) thành ra bản mật mã. Làm thế nào để tạo ra các dãy bit ngẫu nhiên? Có thể tạo ra bằng phương pháp vật lý đơn giản như sau: ta tung đồng xu lên, nếu đồng xu rơi xuống ở mặt sấp thì ta ghi bit 0, ở mặt ngửa thì ta ghi bit 1; tung n lần ta sẽ được một dãy n

bit, dãy bit thu được như vậy có thể được xem là dãy bit ngẫu nhiên. Nhưng tạo ra theo cách như vậy thì khó có thể sử dụng một cách phổ biến, vì không thể tìm ra *qui luật* để theo đó mà sinh ra dãy bit ngẫu nhiên được. Ở đây ta gặp một khó khăn có tính bản chất: nếu có qui luật thì đã không còn là ngẫu nhiên nữa rồi! Như vậy, nếu ta muốn tìm theo qui luật, thì không bao giờ có thể tìm ra các dãy bit ngẫu nhiên, mà cùng lắm cũng chỉ có thể được các dãy bit gần ngẫu nhiên, hay *giả ngẫu nhiên*, mà thôi. Từ nhiều chục năm nay, người ta đã nghiên cứu đề xuất nhiều thuật toán toán học để sinh ra các dãy bit giả ngẫu nhiên, và cũng đã đưa ra nhiều thuộc tính để đánh giá một dãy bit giả ngẫu nhiên có đáng được xem là "gần" ngẫu nhiên hay không. Một vài thuộc tính chủ yếu mà người ta đã đề xuất là: cho một dãy bit $X = (x_1,x_2,.....,x_n,...)$; dãy đó được xem là giả ngẫu nhiên "tốt" nếu xác suất xuất hiện bit 0 hay bit 1 trong toàn dãy đó cũng như trong mọi dãy con bất kỳ của nó đều bằng 1/2; hoặc một tiêu chuẩn khác: nếu mọi chương trình sinh ra được đoạn đầu n bit của dãy đều phải có độ phức tạp (hay độ dài) cỡ n ký tự ! Về sau này, khi lý thuyết về độ phức tạp tính toán đã được phát triển thì tiêu chuẩn về ngẫu nhiên cũng được qui về tiêu chuẩn phức tạp tính toán, cụ thể một dãy bit X được xem là giả ngẫu nhiên "tốt" nếu mọi thuật toán tìm được bit thứ n (x_n) khi biết các bit trước đó ($x_1,,...,x_{n-1}$) với xác suất đúng > 1/2 đều phải có độ phức tạp tính toán thuộc lớp *NP*-khó!

Lý thuyết về độ phức tạp tính toán ra đời từ giữa những năm 1960 đã cho ta một cách thích hợp để qui yêu cầu bí mật hoặc ngẫu nhiên về một yêu cầu có thể định nghĩa được là yêu cầu về *độ phức tạp tính toán.* Bây giờ ta có thể nói: một giải pháp mật mã là bảo đảm bí mật, nếu mọi thuật toán thám mã, nếu có, đều phải được thực hiện với độ phức tạp tính toán cực lớn! Cực lớn là bao nhiêu? Là vượt quá giới hạn khả năng tính toán (bao gồm cả máy tính) mà người thám mã có thể có. Về lý thuyết, có thể xem đó là những độ phức tạp tính toán với tốc độ tăng vượt quá hàm mũ, hoặc thuộc loại *NP*-khó. Tuy nhiên, lý thuyết độ phức tạp tính toán không chỉ cống hiến cho ta một khái niệm để giúp chính xác hóa tiêu chuẩn bí mật của các giải pháp mật mã, mà còn mở ra một giai đoạn mới của ngành mật mã, biến ngành mật mã thành một khoa học có nội dung

lý luận phong phú và có những ứng dụng thực tiễn quan trọng trong nhiều lĩnh vực của đời sống hiện đại. Bước ngoặt có tính cách mạng trong lịch sử khoa học mật mã hiện đại xẩy ra vào năm 1976 khi hai tác giả Diffie và Hellman đưa ra khái niệm về *mật mã khóa công khai* và một phương pháp trao đổi c*ông khai* để tạo ra một khóa bí mật chung mà tính an toàn được bảo đảm bởi độ khó của một bài toán toán học cụ thể (là bài toán tính "lôgarit rời rạc"). Hai năm sau, năm 1978, Rivest, Shamir và Adleman tìm ra một hệ mật mã khóa công khai và một sơ đồ *chữ ký điện tử* hoàn toàn có thể ứng dụng trong thực tiễn, tính bảo mật và an toàn của chúng được bảo đảm bằng độ phức tạp của một bài toán số học nổi tiếng là bài toán phân tích số nguyên thành các thừa số nguyên tố. Sau phát minh ra hệ mật mã đó (mà nay ta thường gọi là hệ RSA), việc nghiên cứu để phát minh ra các hệ mật mã khóa công khai khác, và ứng dụng các hệ mật mã khóa công khai vào các bài toán khác nhau của an toàn thông tin đã được tiến hành rộng rãi, lý thuyết mật mã và an toàn thông tin trở thành một lĩnh vực khoa học được phát triển nhanh trong vài ba thập niên cuối của thế kỷ 20, lôi cuốn theo sự phát triển của một số bộ môn của toán học và tin học. Trong các chương về sau của tập giáo trình này ta sẽ lần lượt làm quen với một số thành quả chủ yếu của lý thuyết đó.

1.2. Các hệ thống mật mã.

1.2.1. Sơ đồ hệ thống mật mã.

Mật mã được sử dụng để bảo vệ tính bí mật của thông tin khi thông tin được truyền trên các kênh truyền thông công cộng như các kênh bưu chính, điện thoại, mạng truyền thông máy tính, mạng Internet, v.v... Giả thử một người gửi A muốn gửi đến một người nhận B một văn bản (chẳng hạn, một bức thư) p, để bảo mật A lập cho p một bản mật mã c, và thay cho việc gửi p, A gửi cho B bản mật mã c, B nhận được c và "giải mã" c để lại được văn bản p như A định gửi. Để A biến p thành c và B biến ngược lại c thành p, A và B phải thỏa thuận trước với nhau các thuật toán lập mã và giải mã, và đặc biệt một *khóa mật mã chung K* để thực hiện các thuật toán đó. Người ngoài, không biết các thông tin đó (đặc biệt, không biết khóa

K), cho dù có lấy trộm được *c* trên kênh truyền thông công cộng, cũng không thể tìm được văn bản *p* mà hai người A, B muốn gửi cho nhau. Sau đây ta sẽ cho một định nghĩa hình thức về sơ đồ mật mã và cách thức thực hiện để lập mật mã và giải mật mã.

Định nghĩa 1.2.1. *Một sơ đồ hệ thống mật mã là một bộ năm*

$$\mathcal{S} = (\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D}) \qquad (1)$$

thỏa mãn các điều kiện sau đây:

$\mathcal{P}$ *là một tập hữu hạn các ký tự bản rõ,*

$\mathcal{C}$ *là một tập hữu hạn các ký tự bản mã,*

$\mathcal{K}$ *là một tập hữu hạn các khóa,*

$\mathcal{E}$ *là một ánh xạ từ* $\mathcal{K}x\mathcal{P}$ *vào* $\mathcal{C}$, *được gọi là phép lập mật mã; và* $\mathcal{D}$ *là một ánh xạ từ* $\mathcal{K}x\mathcal{C}$ *vào* $\mathcal{P}$, *được gọi là phép giải mã. Với mỗi* $K \in \mathcal{K}$, *ta định nghĩa* $e_K : \mathcal{P} \to \mathcal{C}$, $d_K : \mathcal{C} \to \mathcal{P}$ *là hai hàm cho bởi*:

$$\forall x \in \mathcal{P} : e_K(x) = \mathcal{E}(K,x) ; \forall y \in \mathcal{C} : d_K(y) = \mathcal{D}(K,y).$$

e_K *và* d_K *được gọi lần lượt là hàm lập mã và hàm giải mã ứng với khóa mật mã K. Các hàm đó phải thỏa mãn hệ thức*:

$$\forall x \in \mathcal{P} : d_K(e_K(x)) = x.$$

Về sau, để thuận tiện ta sẽ gọi một danh sách (1) thoả mãn các tính chất kể trên là một *sơ đồ hệ thống mật mã* , còn khi đã chọn cố định một khoá *K*, thì danh sách ($\mathcal{P}$, $\mathcal{C}$, e_K , d_K) là một *hệ mật mã* thuộc sơ đồ đó.

Trong định nghĩa này, phép lập mật mã (giải mã) được định nghĩa cho từng ký tự bản rõ (bản mã). Trong thực tế, bản rõ của một thông báo thường là một dãy ký tự bản rõ, tức là phần tử của tập $\mathcal{P}^*$, và bản mật mã cũng là một dãy các ký tự bản mã, tức là phần tử của tập $\mathcal{C}^*$, việc mở rộng các hàm e_K và d_K lên các miền tương ứng $\mathcal{P}^*$ và $\mathcal{C}^*$ để được các thuật toán lập mật mã và giải mã dùng trong thực tế sẽ được trình bày trong tiết sau. Các tập ký tự bản rõ và bản mã thường dùng là các tập ký tự của ngôn ngữ thông thường như tiếng Việt, tiếng Anh (ta ký hiệu tập ký tự tiếng Anh là $\boldsymbol{A}$ tức $\boldsymbol{A}$ = {*a,b,c,...,x,y,z* } gồm 26 ký tự; tập ký tự nhị phân $\boldsymbol{B}$ chỉ gồm hai ký tự

0 và 1; tập các số nguyên không âm bé hơn một số n nào đó (ta ký hiệu tập này là $\boldsymbol{Z_n}$ **tức** $\boldsymbol{Z_n = \{0,1,2,....,n-1\}}$**).** Chú ý rằng có thể xem $\boldsymbol{B = Z_2}$. Để thuận tiện, ta cũng thường đồng nhất tập ký tự tiếng Anh $\boldsymbol{A}$ với tập gồm 26 số nguyên không âm đầu tiên $\boldsymbol{Z_{26}}$ = {0,1,2,...., 24,25} với sự tương ứng sau đây:

a b c d e f g h i j k l m n o p q r s t u v w x y z

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25.

Đôi khi ta cũng dùng với tư cách tập ký tự bản rõ hay bản mã là các tập tích của các tập nói trên, đặc biệt là các tập $\boldsymbol{A}^m$, $\boldsymbol{B}^m$, $\boldsymbol{Z}_n^m$.

1.2.2. Mã theo khối và mã theo dòng.

Như nói ở trên, bản rõ của thông báo mà ta muốn gửi đi thường là một dãy ký tự, trong khi theo định nghĩa của sơ đồ mật mã, hàm lập mật mã và hàm giải mã được định nghĩa cho từng ký tự. Từ các định nghĩa của hàm lập mật mã và hàm giải mã, ta mở rộng thành thuật toán lập mã (và giải mã) xác định cho mọi bản rõ (bản mã) như sau:

Theo cách *mã theo khối* (block cipher), trước hết ta xác định một độ dài khối (chẳng hạn là k), tiếp đó mở rộng không gian khóa từ $\mathcal{K}$ thành $\mathcal{K}^k$, và với mỗi $K = K_1...K_k \in \mathcal{K}^k$, ta mở rộng e_K và d_K thành các thuật toán $e_K : \mathcal{P}^k \to \mathcal{C}^k$ và $d_K : \mathcal{C}^k \to \mathcal{P}^k$ như sau: với mọi $x_1...x_k \in \mathcal{P}^k$ và $y_1...y_k \in \mathcal{C}^k$ ta có

$$e_K(x_1....x_k) = e_{K_1}(x_1)....e_{K_k}(x_k); \quad d_K(y_1....y_k) = d_{K_1}(y_1)....d_{K_k}(y_k).$$

Giả thử bản rõ mà ta muốn lập mật mã cho nó là dãy ký tự $X \in \mathcal{P}^*$. Ta cắt X thành từng khối, mỗi khối có độ dài k, khối cuối cùng có thể có độ dài $< k$, ta luôn có thể giả thiết là có thể bổ sung vào phần cuối của khối một số ký tự qui ước nào đó để nó cũng có độ dài k. Do đó ta có thể giả thiết $X = X_1....X_m$, trong đó mỗi $X_1,...,X_m$ là một khối có độ dài k. Và ta định nghĩa bản mật mã của X là:

$$e_K(X) = e_K(X_1....X_m) = e_K(X_1)....e_K(X_m).$$

Đặt $Y = e_K(X_1)....e_K(X_m)$, ta có thể viết $Y = Y_1.... Y_m$ với $Y_i = e_K(X_i)$, và do đó có

$$d_K(Y) = d_K(Y_1)....d_K(Y_m) = X_1....X_m = X.$$

Cách mã theo khối đơn giản và thông dụng nhất là khi ta chọn độ dài khối k=1. Khi đó với mọi bản rõ $X = x_1...x_m \in \mathcal{P}^*$ ta có

$$e_K(X) = e_K(x_1....x_m) = e_K(x_1)....e_K(x_m).$$

Với cách *mã theo dòng* (stream cipher), trước hết ta phải xác định một *dòng khóa*, tức là một phần tử $K = K_1...K_m \in \mathcal{K}^*$, với dòng khóa đó ta xác định với mọi bản rõ $X = x_1...x_m \in \mathcal{P}^*$ bản mã tương ứng là

$$e_K(X) = e_K(x_1...x_m) = e_{K_1}(x_1)...e_{K_m}(x_m).$$

Giải mã $Y = e_K(X)$ ta được

$$d_K(Y) = d_{K_1}(e_{K_1}(x_1))....d_{K_m}(e_{K_m}(x_m)) = x_1....x_m = X .$$

Để sử dụng cách lập mật mã theo dòng, ngoài sơ đồ mật mã gốc ta còn phải có một dòng khóa, tức là một dãy có độ dài tùy ý các ký tự khóa. Đó thường là các dãy các ký tự khóa được sinh ra bởi một bộ "tạo dãy ngẫu nhiên" nào đó xuất phát từ một "mầm" chọn trước. Trong các ứng dụng thực tế, người ta thường dùng cách mã theo dòng có sơ đồ mật mã gốc là sơ đồ Vernam với

$$\mathcal{P} = \mathcal{C} = \mathcal{K} = \{0,1\}$$

và các hàm lập mã và giải mã được xác định bởi

$$e_K(x) = x + K \bmod 2,\ \ d_K(y) = y + K \bmod 2\ (K = 0 \text{ hoặc } 1);$$

dòng khóa là dãy bit ngẫu nhiên được sinh ra bởi một bộ tạo dãy bit ngẫu nhiên nào đó.

1.3. Mật mã khóa đối xứng và mật mã có khóa công khai.

Theo định nghĩa 1.2.1 về sơ đồ mật mã, cứ mỗi lần truyền tin bảo mật, cả người gửi A và người nhận B phải cùng thỏa thuận trước với nhau một khóa chung K, sau đó người gửi dùng e_K để lập mật mã cho thông báo gửi đi, và người nhận dùng d_K để giải mã bản mật mã nhận được. Người gửi và người nhận cùng có một khóa

chung K, được giữ như bí mật riêng của hai người, dùng cả cho lập mật mã và giải mã, ta gọi những hệ mật mã với cách sử dụng đó là *mật mã khóa đối xứng*, đôi khi cũng gọi là mật mã truyền thống, vì đó là cách đã được sử dụng từ hàng ngàn năm nay.

Tuy nhiên, về nguyên tắc hai hàm lập mã và giải mã là khác nhau, không nhất thiết phải phụ thuộc cùng một khóa. Nếu ta xác định mỗi khóa K gồm có hai phần $K = (K', K'')$, K' dành cho việc lập mật mã (và ta có hàm lập mã $e_{K'}$), K'' dành cho việc giải mã (và có hàm giải mã $d_{K''}$), các hàm lập mã và giải mã thỏa mãn hệ thức

$$d_{K''}(e_{K'}(x)) = x \text{ với mọi } x \in \mathcal{P},$$

thì ta được một hệ *mật mã khóa phi đối xứng*. Như vậy, trong một hệ mật mã khóa phi đối xứng, các khóa lập mã và giải mã (K' và K'') là khác nhau, nhưng tất nhiên có quan hệ với nhau. Trong hai khóa đó, khóa cần phải giữ bí mật là khóa giải mã K'', còn khóa lập mã K' có thể được công bố công khai; tuy nhiên điều đó chỉ có ý nghĩa thực tiễn khi việc *biết* K' *tìm* K'' là cực kỳ khó khăn đến mức hầu như không thể thực hiện được. Một hệ mật mã khóa phi đối xứng có tính chất nói trên, trong đó khóa lập mật mã K' của mỗi người tham gia đều được công bố công khai, được gọi là *hệ mật mã khóa công khai*. Khái niệm mật mã khóa công khai mới được ra đời vào giữa những năm 1970, và ngay sau đó đã trở thành một khái niệm trung tâm của khoa học mật mã hiện đại. Ta sẽ dành phần lớn nội dung giáo trình này cho các hệ mật mã đó và những ứng dụng của chúng vào các vấn đề an toàn thông tin.

1.4. Các bài toán về an toàn thông tin.

Chúng ta đang sống trong một thời đại bùng nổ thông tin. Nhu cầu trao đổi thông tin và các phương tiện truyền đưa thông tin phát triển một cách nhanh chóng. Và cùng với sự phát triển đó, đòi hỏi bảo vệ tính bí mật và an toàn của thông tin cũng càng ngày càng to lớn và có tính phổ biến. Có nhiều bài toán khác nhau về yêu cầu an toàn thông tin tùy theo những tình huống khác nhau, nhưng tựu

trung có một số bài toán chung nhất mà ta thường gặp trong thực tiễn là những bài toán sau đây:

- *bảo mật* : giữ thông tin được bí mật đối với tất cả mọi người, trừ một ít người có thẩm quyền được đọc, biết thông tin đó;

- *toàn vẹn thông tin* : bảo đảm thông tin không bị thay đổi hay xuyên tạc bởi những kẻ không có thẩm quyền hoặc bằng những phương tiện không được phép;

- *nhận thực một thực thể*: xác nhận danh tính của một thực thể, chẳng hạn một người, một máy tính cuối trong mạng, một thẻ tín dụng,... ;

- *nhận thực một thông báo* : xác nhận nguồn gốc của một thông báo được gửi đến ;

- *chữ ký* : một cách để gắn kết một thông tin với một thực thể, thường dùng trong bài toán nhận thực một thông báo cũng như trong nhiều bài toán nhận thực khác ;

- *ủy quyền* : chuyển cho một thực thể khác quyền được đại diện hoặc được làm một việc gì đó ;

- *cấp chứng chỉ* : cấp một sự xác nhận thông tin bởi một thực thể được tín nhiệm ;

- *báo nhận* : xác nhận một thông báo đã được nhận hay một dịch vụ đã được thực hiện ;

- *làm chứng* : kiểm thử việc tồn tại một thông tin ở một thực thể khác với người chủ sở hữu thông tin đó ;

- *không chối bỏ được* : ngăn ngừa việc chối bỏ trách nhiệm đối với một cam kết đã có (thí dụ đã ký vào một văn bản) ;

- *ẩn danh* : che giấu danh tính của một thực thể tham gia trong một tiến trình nào đó (thường dùng trong giao dịch tiền điện tử) ;

- *thu hồi* : rút lại một giấy chứng chỉ hay ủy quyền đã cấp;

- vân vân.....

Cơ sở của các giải pháp cho các bài toán kể trên là các phương pháp mật mã, đặc biệt là mật mã khóa công khai, ta sẽ xem xét kỹ một vài bài toán đó trong các chương tiếp theo.