TRƯỜNG ĐẠI HỌC TRẦN ĐẠI NGHĨA

KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN MÔN HỌC

MÔN HỌC: AN TOÀN VÀ BẢO MẬT HỆ THỐNG THÔNG TIN

ĐỀ TÀI:

XÂY DỰNG CHƯƠNG TRÌNH CHỮ KÝ SỐ ĐỂ XÁC THỰC NỘI DUNG CỦA MỘT VĂN BẢN BẰNG MÃ HÓA BẤT ĐỐI XỨNG

TP. HỒ CHÍ MINH, THÁNG 11 NĂM 2020

TRƯỜNG ĐẠI HỌC TRẦN ĐẠI NGHĨA

KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN MÔN HỌC

MÔN HỌC: AN TOÀN VÀ BẢO MẬT HỆ THỐNG THÔNG TIN

ĐỀ TÀI:

XÂY DỰNG CHƯƠNG TRÌNH CHỮ KÝ SỐ ĐỂ XÁC THỰC NỘI DUNG CỦA MỘT VĂN BẢN BẰNG MÃ HÓA BẤT ĐỐI XỨNG

Nhóm báo cáo:

Nguyễn Tiểu Phụng

Huỳnh Đức Anh Tuấn

Giảng viên hướng dẫn:

Th.s Đặng Thế Hùng

TP. HỒ CHÍ MINH, THÁNG 11 NĂM 2020

LỜI CẢM ƠN

Chúng em xin gửi lời cảm ơn chân thành đến thầy cô giảng viên trong khoa Công nghệ thông tin trường Đại học Trần Đại Nghĩa. Và đặc biệt là thầy Thạc sĩ Đặng Thế Hùng – giảng viên học phần "An toàn bảo mật hệ thống thông tin" đã tận tình hướng dẫn, truyền đạt kiến thức và kỹ năng cần thiết để chúng em có thể hoàn thành đồ án môn học này.

Tuy nhiên, trong quá trình tìm hiểu và nghiên cứu đề tài, do kiến thức chuyên ngành và thời gian còn hạn chế chúng em vẫn còn nhiều thiếu sót trong quá trình tìm hiểu, thực hiện, đánh giá và trình bày về đề tài. Rất mong được sự quan tâm, góp ý của các thầy cô và giảng viên bộ môn để đồ án môn học của chúng em được hoàn chỉnh hơn.

Xin chân thành cảm ơn!

MỞ ĐẦU

1. Lý do chọn đề tài

- Mật mã học là một trong những vấn đề quan trọng trong lĩnh vực bảo mật và an toàn thông tin. Với sự bùng nổ mạnh mẽ của internet hiện nay, mạng máy tính đang ngày càng đóng vai trò thiết yếu trong mọi lĩnh vực hoạt động của toàn xã hội, và khi nó trở thành phương tiện điều hành các hệ thống thì nhu cầu bảo mật thông tin được đặt lên hàng đầu.
- Việc sử dụng chữ ký số là một giải pháp hữu hiệu, ngày càng được ứng dụng nhiều trong thực tế, không chỉ giới hạn trong lĩnh vực công nghệ thông tin, mật mã học mà còn được áp dụng trong nhiều lĩnh vực khác như tài chính ngân hàng, viễn thông,...
- Mật mã học khóa công khai tạo ra chữ ký số và úng dụng vào các tài liệu. Hệ mã hóa RSA – hệ mã hóa điểm hình của mật mã công khai cùng với hàm băm mật mã SHA chính là những công cụ chính tạo ra chữ ký số.

2. Cấu trúc đồ án

- Chương 1: Tổng quan
- Chương 2: Hệ mã RSA – Hàm băm SHA và Chữ ký số
- Chương 3: Xây dựng ứng dụng

MỤC LỤC

CHƯƠNG 1 – TỔNG QUAN

1.1. Tổng quan

- Vấn đề bảo đảm an toàn cho các hệ thống thông tin là một trong những vấn đề quan trọng cần cân nhắc trong suốt quá trình thiết kế, thi công, vận hành và bảo dưỡng hệ thống thông tin.

- Cũng như tất cả các hoạt động khác trong đời sống xã hội, từ khi con người có nhu cầu lưu trữ và xử lý thông tin, đặc biệt là từ khi thông tin được xem như một bộ phận của tư liệu sản xuất, thì nhu cầu bảo vệ thông tin càng trở nên bức thiết. Bảo vệ thông tin là bảo vệ tính bí mật của thông tin và tính toàn vẹn của thông tin. Một số loại thông tin chỉ còn ý nghĩa khi chúng được giữ kín hoặc giới hạn trong một số các đối tượng nào đó, ví dụ như thông tin về chiến lược quân sự chẳng hạn. Đây là tính bí mật của thông tin. Hơn nữa, thông tin không phải luôn được con người ghi nhớ do sự hữu hạn của bộ óc, nên cần phải có thiết bị để lưu trữ thông tin. Nếu thiết bị lưu trữ hoạt động không an toàn, thông tin lưu trữ trên đó bị mất đi hoặc sai lệch toàn bộ hay một phần, khi đó tính toàn vẹn của thông tin không còn được bảo đảm.

- Khi máy tính được sử dụng để xử lý thông tin, hiệu quả xử lý thông tin được nâng cao lên, khối lượng thông tin được xử lý càng ngày càng lớn lên, và kéo theo nó, tầm quan trọng của thông tin trong đời sống xã hội cũng tăng lên. Nếu như trước đây, việc bảo vệ thông tin chỉ chú trọng vào vấn đề dùng các cơ chế và phương tiện vật lý để bảo vệ thông tin theo đúng nghĩa đen của từ này, thì càng về sau, vấn đề bảo vệ thông tin đã trở nên đa dạng hơn và phức tạp hơn. Có thể kể ra hai điều thay đổi lớn sau đây đối với vấn đề bảo vệ thông tin:

+ Sự ứng dụng của máy tính trong việc xử lý thông tin làm thay đổi dạng lưu trữ của thông tin và phương thức xử lý thông tin. Cần thiết phải xây dựng các cơ chế bảo vệ thông tin theo đặc thù hoạt động của máy tính. Từ đây xuất hiện yêu cầu bảo vệ sự an toàn hoạt động của máy tính (Computer Security) tồn tại song song với yêu cầu bảo vệ sự an toàn của thông tin (Information Security).

+ Sự phát triển mạnh mẽ của mạng máy tính và các hệ thống phân tán làm thay đổi phạm vi tổ chức xử lý thông tin. Thông tin được trao đổi giữa các thiết bị xử lý thông qua một khoảng cách vật lý rất lớn, gần như không giới hạn, làm xuất hiện thêm nhiều nguy cơ hơn đối với sự an toàn của thông tin. Từ đó xuất hiện yêu cầu bảo vệ sự an toàn của hệ thống mạng (Network Security), gồm các cơ chế và kỹ thuật phù hợp với việc bảo vệ sự an toàn của thông tin khi chúng được trao đổi giữa các thiết bị trên mạng.

1.2. Các đặc trưng của một hệ thống thông tin bảo mật

- Một hệ thống thông tin bảo mật (Secure Information System) là một hệ thống mà thông tin được xử lý trên nó phải đảm bảo được 3 đặc trưng sau đây:
  + Tính bí mật của thông tin (Confidentiality)
  + Tính toàn vẹn của thông tin (Integrity)
  + Tính khả dụng của thông tin (Availability)

Hình 1.1 Mô hình CIA

- Ba đặc trưng này được liên kết lại và xem như là mô hình tiêu chuẩn của các hệ thống thông tin bảo mật, hay nói cách khác, đây là 3 thành phần cốt yếu của một hệ thống thông tin bảo mật.

1.2.1. Tính bảo mật

- Một số loại thông tin chỉ có giá trị đối với một đối tượng xác định khi chúng không phổ biến cho các đối tượng khác. Tính bí mật của thông tin là tính giới hạn về đối tượng được quyền truy xuất đến thông tin. Đối tượng truy xuất có thể là con người, là máy tính hoặc phần mềm, kể cả phần mềm phá hoại như virus, worm, spyware, …
- Tuỳ theo tính chất của thông tin mà mức độ bí mật của chúng có khác nhau. Ví dụ: các thông tin về chính trị và quân sự luôn được xem là các thông tin nhạy cảm nhất đối với các quốc gia và được xử lý ở mức bảo mật cao nhất. Các thông tin khác như thông tin về hoạt động và chiến lược kinh doanh của doanh nghiệp, thông tin cá nhân, đặc biệt của những người nổi tiếng, thông tin cấu hình hệ thống của các mạng cung cấp dịch vụ, v.v… đều có nhu cầu được giữ bí mật ở từng mức độ.
- Để đảm bảo tính bí mật của thông tin, ngoài các cơ chế và phương tiện vật lý như nhà xưởng, thiết bị lưu trữ, dịch vụ bảo vệ, … thì kỹ thuật mật mã hoá (Cryptography) được xem là công cụ bảo mật thông tin hữu hiệu nhất trong môi trường máy tính. Các kỹ thuật mật mã hoá sẽ được trình bày cụ thể ở chương II. Ngoài ra, kỹ thuật quản lý truy xuất (Access Control) cũng được thiết lập để bảo đảm chỉ có những đối tượng được cho phép mới có thể truy xuất thông tin. Access control sẽ được trình bày ở phần 3 của chương này.
- Sự bí mật của thông tin phải được xem xét dưới dạng 2 yếu tố tách rời: sự tồn tại của thông tin và nội dung của thông tin đó.
- Đôi khi, tiết lộ sự tồn tại của thông tin có ý nghĩa cao hơn tiết lộ nội dung của nó. Ví dụ: chiến lược kinh doanh bí mật mang tính sống còn của một công ty đã bị tiết lộ cho một công ty đối thủ khác. Việc nhận thức được rằng có điều đó tồn tại sẽ quan trọng hơn nhiều so với việc biết cụ thể về nội dung thông tin, chẳng hạn như ai đã tiết lộ, tiết lộ cho đối thủ nào và tiết lộ những thông tin gì,…
- Cũng vì lý do này, trong một số hệ thống xác thực người dùng (user authentication) ví dụ như đăng nhập vào hệ điều hành Netware hay đăng nhập

vào hộp thư điện tử hoặc các dịch vụ khác trên mạng, khi người sử dụng cung cấp một tên người dùng (user-name) sai, thay vì thông báo rằng user-name này không tồn tại, thì một số hệ thống sẽ thông báo rằng mật khẩu (password) sai, một số hệ thống khác chỉ thông báo chung chung là "Invalid user name/password" (người dùng hoặc mật khẩu không hợp lệ). Dụng ý đằng sau câu thông báo không rõ ràng này là việc từ chối xác nhận việc tồn tại hay không tồn tại một user-name như thế trong hệ thống. Điều này làm tăng sự khó khăn cho những người muốn đăng nhập vào hệ thống một cách bất hợp pháp bằng cách thử ngẫu nhiên.

1.2.2. Tính toàn vẹn

- Đặc trưng này đảm bảo sự tồn tại nguyên vẹn của thông tin, loại trừ mọi sự thay đổi thông tin có chủ đích hoặc hư hỏng, mất mát thông tin do sự cố thiết bị hoặc phần mềm. Tính toàn vẹn được xét trên 2 khía cạnh:
  + Tính nguyên vẹn của nội dung thông tin.
  + Tính xác thực của nguồn gốc của thông tin.
- Nói một cách khác, tính toàn vẹn của thông tin phải được đánh giá trên hai mặt: toàn vẹn về nội dung và toàn vẹn về nguồn gốc.
- Ví dụ: một ngân hàng nhận được lệnh thanh toán của một người tự xưng là chủ tài khoản với đầy đủ những thông tin cần thiết. Nội dung thông tin được bảo toàn vì ngân hàng đã nhận được một cách chính xác yêu cầu của khách hàng (đúng như người xưng là chủ tài khoản gởi đi).
- Tuy nhiên, nếu lệnh thanh toán này không phải cho chính chủ tài khoản đưa ra mà do một người nào khác nhờ biết được thông tin bí mật về tài khoản đã mạo danh chủ tài khoản để đưa ra, ta nói nguồn gốc của thông tin đã không được bảo toàn.
- Một ví dụ khác, một tờ báo đưa tin về một sự kiện vừa xảy ra tại một cơ quan quan trọng của chính phủ, có ghi chú rằng nguồn tin từ người phát ngôn của cơ quan đó. Tuy nhiên, nếu tin đó thật sự không phải do người phát ngôn công bố

mà được lấy từ một kênh thông tin khác, không xét đến việc nội dung thông tin có đúng hay không, ta nói rằng nguồn gốc thông tin đã không được bảo toàn.

- Sự tòan vẹn về nguồn gốc thông tin trong một số ngữ cảnh có ý nghĩa tương đương với sự đảm bảo tính không thể chối cãi (non-repudiation) của hệ thống thông tin.
- Các cơ chế đảm bảo sự toàn vẹn của thông tin được chia thành 2 loại: các cơ chế ngăn chặn (Prevention mechanisms) và các cơ chế phát hiện (Detection mechanisms).
- Cơ chế ngăn chặn có chức năng ngăn cản các hành vi trái phép làm thay đổi nội dung và nguồn gốc của thông tin. Các hành vi này bao gồm 2 nhóm: hành vi cố gắng thay đổi thông tin khi không được phép truy xuất đến thông tin và hành vi thay đổi thông tin theo cách khác với cách đã được cho phép.
- Ví dụ: một người ngoài công ty cố gắng truy xuất đến cơ sở dữ liệu kế toán của một công ty và thay đổi dữ liệu trong đó. Đây là hành vi thuộc nhóm thứ nhất. Trường hợp một nhân viên kế toán được trao quyền quản lý cơ sở dữ liệu kế toán của công ty, và đã dùng quyền truy xuất của mình để thay đổi thông tin nhằm biển thủ ngân quỹ, đây là hành vi thuộc nhóm thứ hai.
- Nhóm các cơ chế phát hiện chỉ thực hiện chức năng giám sát và thông báo khi có các thay đổi diễn ra trên thông tin bằng cách phân tích các sự kiện diễn ra trên hệ thống mà không thực hiện chức năng ngăn chặn các hành vi truy xuất trái phép đến thông tin.
- Nếu như tính bí mật của thông tin chỉ quan tâm đến việc thông tin có bị tiết lộ hay không, thì tính toàn vẹn của thông tin vừa quan tâm tới tính chính xác của thông tin và cả mức độ tin cậy của thông tin. Các yếu tố như nguồn gốc thông tin, cách thức bảo vệ thông tin trong quá khứ cũng như trong hiện tại đều là những yếu tố quyết định độ tin cậy của thông tin và do đó ảnh hưởng đến tính toàn vẹn của thông tin. Nói chung, việc đánh giá tính toàn vẹn của một hệ thống thông tin là một công việc phức tạp.

1.2.3. Tính khả dụng

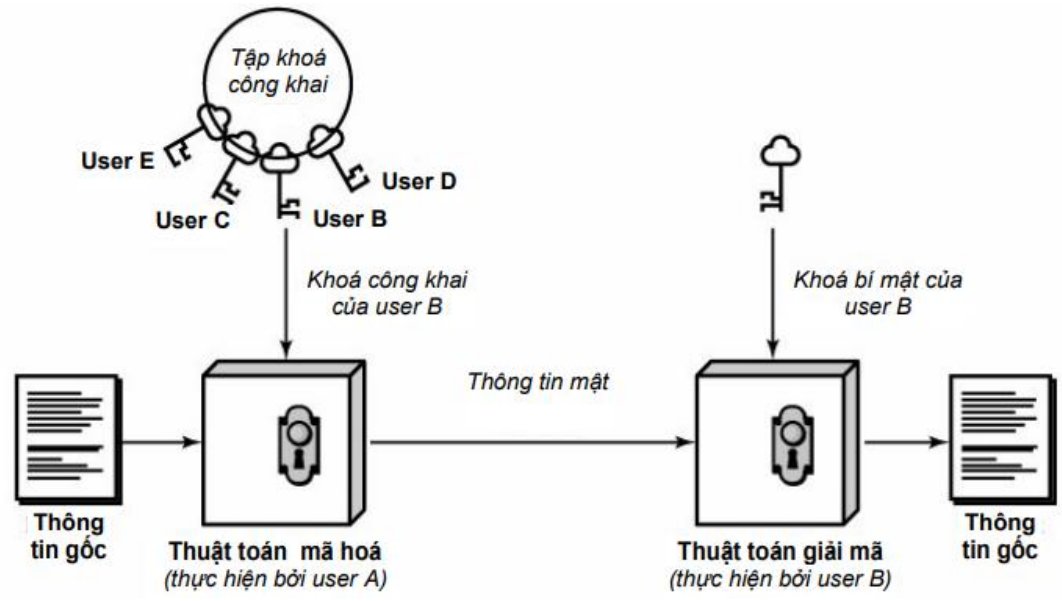
- Tính khả dụng của thông tin là tính sẵn sàng của thông tin cho các nhu cầu truy xuất hợp lệ.
- Ví dụ: các thông tin về quản lý nhân sự của một công ty được lưu trên máy tính, được bảo vệ một cách chắc chắn bằng nhiều cơ chế đảm bảo thông tin không bị tiết lộ hay thay đổi. Tuy nhiên, khi người quản lý cần những thông tin này thì lại không truy xuất được vì lỗi hệ thống. Khi đó, thông tin hoàn toàn không sử dụng được và ta nói tính khả dụng của thông tin không được đảm bảo.
- Một hệ thống khả dụng là một hệ thống làm việc trôi chảy và hiệu quả, có khả năng phục hồi nhanh chóng nếu có sự cố xảy ra.
- Hiện nay, các hình thức tấn công từ chối dịch vụ DoS (Denial of Service) và DDoS (Distributed Denial of Service) được đánh giá là các nguy cơ lớn nhất đối với sự an toàn của các hệ thống thông tin, gây ra những thiệt hại lớn và đặc biệt là chưa có giải pháp ngăn chặn hữu hiệu. Các hình thức tấn công này đều nhắm vào tính khả dụng của hệ thống.
- Một số hướng nghiên cứu đang đưa ra các mô hình mới cho việc mô tả các hệ thống an toàn. Theo đó, mô hình CIA không mô tả được đầy đủ các yêu cầu an toàn của hệ thống mà cần phải định nghĩa lại một mô hình khác với các đặc tính của thông tin cần được đảm bảo như:
  + Tính khả dụng (Availability)
  + Tính tiện ích (Utility)
  + Tính toàn vẹn (Integrity)
  + Tính xác thực (Authenticity)
  + Tính bảo mật (Confidentiality)
  + Tính sở hữu (Possession)

CHƯƠNG 2 HỆ MÃ RSA – HÀM BĂM SHA VÀ CHỮ KÝ SỐ

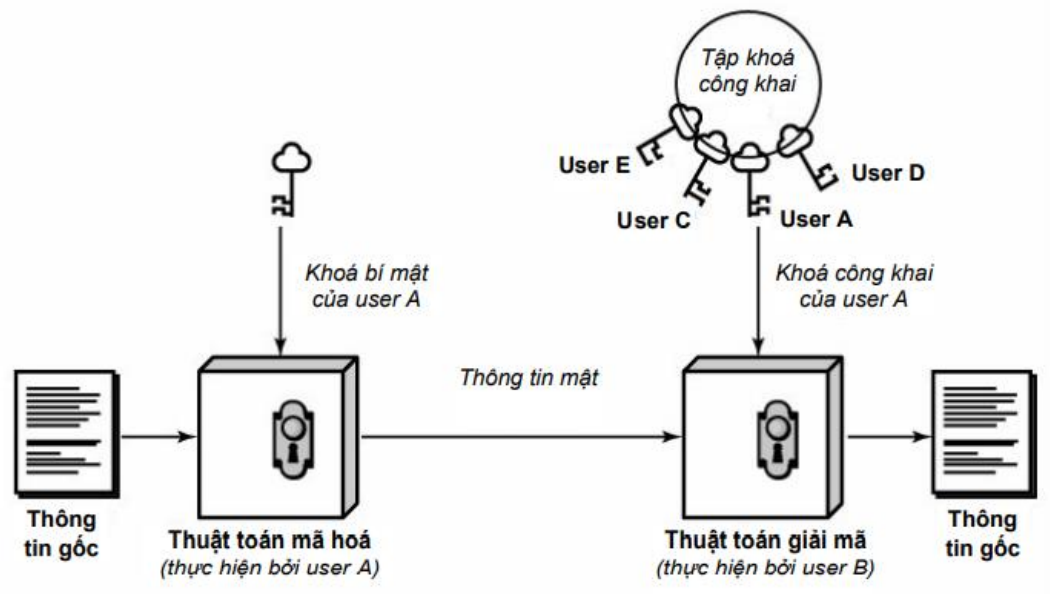
2.1.Hệ mã RSA

2.1.1. Cấu trúc hệ thống mật mã bất đối xứng

- Đặc trưng của kỹ thuật mật mã bất đối xứng là dùng 2 khóa riêng biệt cho hai quá trình mã hóa và giải mã, trong đó có một khóa được phổ biến công khai (public key hay PU) và khóa còn lại được giữ bí mật (private key hay PR). Cả hai khoá đều có thể được dùng để mã hoá hoặc giải mã. Việc chọn khoá công khai hay khoá bí mật cho quá trình mã hoá sẽ tạo ra hai ứng dụng khác nhau của kỹ thuật mật mã bất đối xứng:
  + Nếu dùng khoá công khai để mã hoá và khoá bí mật để giải mã, ta có ứng dụng bảo mật trên thông tin (confidentiality).
  + Nếu dùng khoá bí mật để mã hoá và khoá công khai để giải mã, ta có ứng dụng xác thực nội dung và nguồn gốc thông tin (authentication).
- Thuật toán mật mã bất đối xứng dựa chủ yếu trên các hàm toán học hơn là dựa vào các thao tác trên chuỗi bit. Mật mã hóa bất đối xứng còn được gọi bằng một tên thông dụng hơn là mật mã hóa dùng khóa công khai (public key encryption).
- Nói chung, mật mã hóa bất đối xứng không phải là một kỹ thuật mật mã an tòan hơn so với mật mã đối xứng, mà độ an tòan của một thuật toán mã nói chung phụ thuộc vào 2 yếu tố: Độ dài của khóa và mức độ phức tạp khi thực hiện thuật tóan (trên máy tính). Hơn nữa, mặc dù được ra đời sau nhưng không có nghĩa rằng mật mã bất đối xứng hòan tòan ưu điểm hơn và sẽ được sử dụng thay thế cho mật mã đối xứng. Mỗi kỹ thuật mã có một thế mạnh riêng và mật mã đối xứng vẫn rất thích hợp cho các hệ thống nhỏ và đơn giản.
- Ngoài ra, vấn đề phân phối khóa trong mật mã bất đối xứng cũng được đánh giá là một trong những vấn đề phức tạp khi triển khai kỹ thuật mật mã này trong thực tế.



Hình a- ứng dụng trong bảo mật thông tin



Hình b- ứng dụng trong xác thực thông tin

Hình 2.1 Cấu trúc hệ thông mật mã bất đối xứng

- Cấu trúc một hệ thống mật mã bất đối xứng được trình bày trong hình 2.22.
- Các bước cơ bản của một hệ thống mật mã dùng khóa công khai bao gồm:

+ Mỗi thực thể thông tin (user) tạo ra một cặp khóa (public/private) để dùng cho việc mã hóa và giải mã.

+ Mỗi user thông báo một trong hai khoá của mình cho các user khác biết, khóa này được gọi là khóa công khai (public key). Khóa còn lại được giữ bí mật, và gọi là khóa riêng (private key).

+ Nếu một user A muốn gởi thông tin cho user B, user A sẽ thực hiện mã hóa thông tin cần gởi bằng khóa công khai của user B.

+ Khi nhận được thông tin đã mã hóa từ user A, user B thực hiện giải mã thông tin đó bằng khóa riêng của mình. Do khóa riêng không phổ biến công khai nên chỉ có một mình user B có khả năng giải mã được.

Mật mã hóa bất đối xứng được sử dụng trong các ứng dụng: che giấu thông tin, tạo chữ ký số (digital signature) và trao đổi khóa trong các thuật tóan mật mã đối xứng (key exchange).

2.1.2. Thuật toán mật mã RSA

- RSA là thuật toán mật mã bất đối xứng được xây dựng bởi Ron Rivest, Adi Shamir và Len Adleman tại viện công nghệ Massachusetts (MIT), do đó được đặt tên là Rivest – Shamir – Adleman hay RSA. Thuật toán này ra đời năm 1977 và cho đến nay đã được ứng dụng trong nhiều lĩnh vực. Cũng như các thuật toán mật mã bất đối xứng khác, nguyên lý của RSA dựa chủ yếu trên lý thuyết số chứ không dựa trên các thao tác xử lý bit.
- RSA là một thuật toán mật mã khối, kích thước khối thông thường là 1024 hoặc 2048 bit.
- Thông tin gốc của RSA được xử lý như các số nguyên. Ví dụ, khi chọn kích thước khối của thuật toán là 1024 bit thì số nguyên này có giá trị từ 0 đến $2^{1024} – 1$, tương đương với số thập phân có 309 chữ số. Chú ý rằng đây là những số nguyên cực lớn, không thể xử lý được bằng cách sử dụng các cấu trúc dữ liệu có sẵn của các ngôn ngữ lập trình phổ biến.

- Thuật toán RSA được mô tả như sau:

B 5. Để tạo ra một cặp khóa RSA, trước hết, chọn hai số nguyên tố đủ lớn p và q. Gọi N là tích của p và q (N = pq).

B 6. Tiếp theo, chọn một số e sao cho e và (p-1)(q-1) là hai số nguyên tố cùng nhau. Sau đó tìm số d sao cho ed = 1 mod (p-1)(q-1). Ký hiệu mod m biểu diễn phép modulo trên cơ số m.

B 7. Bây giờ, bỏ qua vai trò của p và q. Với 3 thành phần còn lại là N, e và d, ta đó:

- Khóa công khai (public key) là tổ hợp (N, e)
- Khóa bí mật (private) là tổ hợp (N, d).

B 8. Việc mã hóa một khối thông tin gốc M được thực hiện theo công thức:

$$\mathbf{C = M^e \bmod N}$$ (với M là số nguyên nhỏ hơn N)

B 9. Và quá trình giải mã C được thực hiện theo công thức:

$$\mathbf{M = C^d \bmod N}$$

Ví dụ: Cặp số nguyên tố p = 11 và q = 3 được chọn để tạo ra cặp khoá RSA cho user A.

- Khi đó, N = pq = 3*11 = 33

$$(p-1)\,(q-1) = (11 - 1)\,(3 - 1) = 20$$

- Tiếp theo, chọn e = 3 thoả điều kiện 3 và 20 là cặp số nguyên tố cùng nhau.
- Với e = 3, ta xác định được d = 7 vì ed = 3*7 = 1 mod 20. Thật ra, có nhiều giá trị d thỏa mãn yêu cầu này, nhưng để cho đơn giản, ta chọn giá trị nhỏ nhất.
- Khi đó, ta xác định được cặp khóa như sau:
  + Khóa công khai: (N, e) = (33, 3)
  + Khóa bí mật: (N, d) = (33, 7)
- Giả sử, user B muốn gởi đoạn thông tin M = 15 cho user A, dựa trên khóa công khai của A, B thực hiện như sau:

$$C = M^e \bmod N = 15^3 \bmod 33 = 3375 \bmod 33 = 9 \bmod 33.$$

- Khi đó, thông tin mật gởi cho A là C = 9.
- Khi nhận được thông tin này, A giải mã bằng khóa riêng của mình (d = 7) như sau:

$$M = Cd \text{ mob } N = 9^7 \bmod 33 = 4.782.969 \bmod 33 = 15 \bmod 33.$$

- Như vậy, thông tin giải mã được là M = 15, đúng với thông tin gốc ban đầu.
- Tóm lại, thuật toán mật mã RSA được thực hiện gồm 3 quá trình tách rời: tạo khoá, mã hoá và giải mã được tóm tắt như sau:

1-Tạo khoá:

- ***Chọn p, q*** *(p và q là số nguyên tố, $p \neq q$)*
- ***Tính N = p.q***
- ***Tính $\phi(N) = (p - 1)(q - 1)$***
- ***Chọn e sao ước số chung lớn nhất của e và $\phi(N)$ là 1***
- ***Chọn d sao cho $e.d \bmod \phi(N) = 1$***
- ***Cặp khoá RSA được tạo ra là PU = (N, e), PR = (N, d)***

2- Mã hoá:

- ***$C = M^e \bmod N$*** *(M là số nguyên nhỏ hơn N)*

3- Giải mã:

- ***$M = C^d \bmod N$***

- Trong thực tế, để đạt được độ an tòan cao, cặp khóa phải được chọn trên các số p và q đủ lớn (N nhỏ nhất phải là 1024 bit), do vậy, vấn đề thực thi RSA bao gồm các phép tóan lũy thừa trên các số rất lớn.
- Vấn đề giảm chi phí tính tóan và tăng tốc độ thực hiện thuật tóan RSA là một trong những vấn đề quan trọng cần phải giải quyết. Trên các hệ thống máy tính hiện nay, hiệu suất thực hiện giải thuật RSA là chấp nhận được.
- Độ an toàn của RSA:
- Theo lý thuyết, hệ thống RSA có thể bị tấn công bằng những phương thức sau đây:

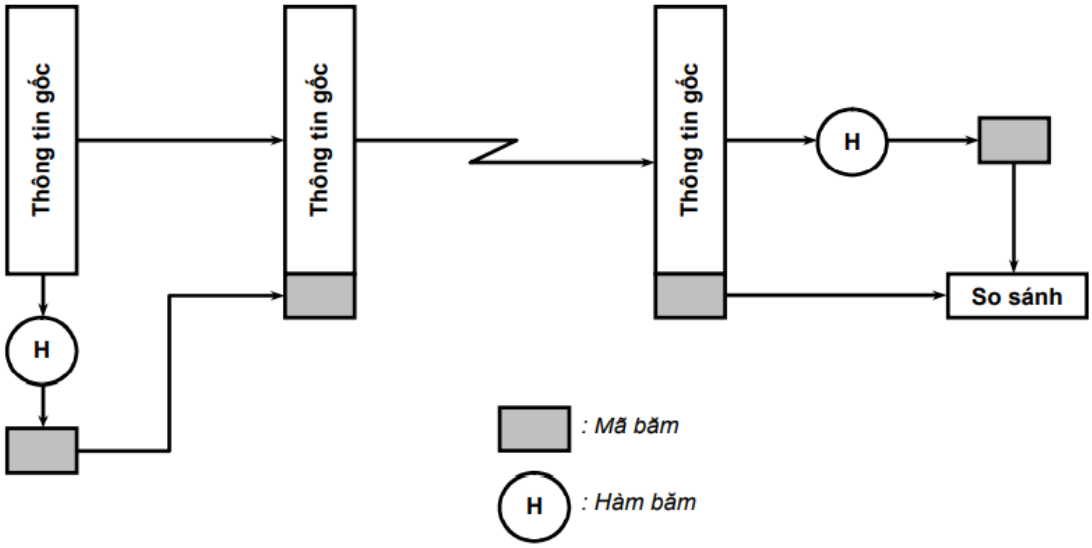
+ Brute-force attack: tìm lần lượt khoá riêng PR
+ Mathematical attack: xác định p và q bằng cách phân tích N thành tích của các thừa số nguyên tố rồi từ đó xác định e và d.
+ Timing attack: dựa trên thời gian thực thi của thuật toán giải mã.
+ Chosen ciphertext attack: sử dụng các đoạn thông tin mật (ciphertext) đặc biệt để khôi phục thông tin gốc.

- Tuy nhiên trong thực tế, nguy cơ tấn công các hệ thống mật mã RSA là rất thấp, do RSA là một thuật toán linh động, kích thước khối dữ liệu gốc và chiều dài khoá dễ dàng được thay đổi mà không ảnh hưởng đến thuật toán mã.

2.2.Thuật toán băm SHA

2.2.1. Hàm băm

- Các hàm băm bảo mật (secure hash functions) hay gọi tắt là hàm băm là một trong những kỹ thuật cơ bản để thực hiện các cơ chế xác thực thông tin (message authentication). Ngoài ra, hàm băm cũng còn được sử dụng trong nhiều thuật toán mật mã, trong chữ ký số (digital signature) và nhiều ứng dụng khác.
- Nguyên tắc của hàm băm là biến đổi khối thông tin gốc có độ dài bất kỳ thành một đoạn thông tin ngắn hơn có độ dài cố định gọi là mã băm (hash code hay message digest). Mã băm được dùng để kiểm tra tính chính xác của thông tin nhận được. Thông thường, mã băm được gởi kèm với thông tin gốc. Ở phía nhận, hàm băm lại được áp dụng đối với thông tin gốc để tìm ra mã băm mới, giá trị này được so sánh với mã băm đi kèm với thông tin gốc. Nếu hai mã băm giống nhau, nghĩa là thông tin gởi đi không bị thay đổi.
- Chỉ có thể dùng hàm băm để tính mã băm từ thông tin gốc chứ không thể tính được thông tin gốc từ mã băm. Do đặc tính này, các hàm băm bảo mật cũng còn được gọi là hàm băm một chiều (one way hash fntion).



Hình 2.2 Một ứng dụng điểm hình của hàm băm

- Hình 2.2 mô tả nguyên lý hoạt động của một giải thuật xác thực thông tin sử dụng hàm băm đơn giản.
- Các yêu cầu của một hàm băm bảo mật H:
  + H có thể được áp dụng cho khối thông tin với chiều dài bất kỳ.
  + Kết quả của hàm H luôn có chiều dài cố định.
  + Việc tính giá trị của H(x) với một giá trị x cho trước phải đơn giản, có thể thực hiện được bằng cả phần cứng hoặc phần mềm.
  + Cho trước một giá trị h, không thể tìm được một giá trị x sao cho H(x) = h, đây được gọi là thuộc tính một chiều của hàm băm (one-way property).
  + Cho trước khối thông tin x, không thể tìm được một khối thông tin y khác x sao cho H(y) = H(x). Thuộc tính này được gọi là weak collision resistance.
  + Không thể tìm được hai khối thông tin x và y khác nhau sao cho H(x) = H(y).
- Thuộc tính này được gọi là strong collision resistance.
- Tấn công trên các hàm băm:

+ Nguyên lý làm việc của hàm băm là biểu diễn một khối thông tin có kích thước lớn bởi một đoạn thông tin có kích thước nhỏ hơn nhiều gọi là mã băm, và trong trường hợp lý tưởng nhất thì các biểu diễn này là các ánh xạ 1:1, tức sẽ không xảy ra tình huống 2 khối thông tin khác nhau cùng cho ra một mã băm. Trường hợp có 2 khối thông tin khác nhau cùng cho ra một mã băm, ta nói thuật tóan băm bị đụng độ (collision). Mục tiêu tấn công vào một hàm băm bảo mật là tạo ra các tình huống đụng độ này.

+ Xác suất để hai khối thông tin có cùng mã băm phụ thuộc vào kích thước của mã băm, tức phụ thuộc vào số lượng mã băm có thể có. Kích thước này càng nhỏ thì khả năng xảy ra càng lớn, và do đó xác suất tấn công thành công càng lớn. Bài toán ngày sinh (Birthday problem)(*) chỉ ra rằng: với kích thước mã băm là n bit, để xác suất xảy ra đụng độ là 50% thì cần có khoảng 2n/2 khối thông tin được xử lý. Người ta thường dùng nguyên lý này để tấn công vào các ứng dụng có sử dụng hàm băm, các tấn công này được gọi là Birthday attack.

+ Nói chung, độ an toàn của một hàm băm phụ thuộc vào kích thước ngõ ra của nó .

2.2.2. Thuật toán SHA

- SHA (Secure Hash Function) được chuẩn hoá năm 1993, sau đó được chỉnh sửa năm 1995 và đặt tên là SHA-1, từ đó phiên bản cũ được gọi là SHA-0.
- SHA-1 tạo ra mã băm có chiều dài cố định là 160 bit. Về sau, có nhiều nâng cấp đối với SHA, chủ yếu là tăng chiều dài mã băm, từ đó xuất hiện các phiên bản khác nhau của SHA, bao gồm: SHA-256 (mã băm dài 256 bit), SHA-384 (mã băm dài 385 bit) và SHA-512 (mã băm dài 512 bit).

Thông số	SHA-1	SHA-256	SHA-384	S
Kích thước mã băm (bit)	160	256	384	
Kích thước thông tin gốc (bit)	$< 2^{64}$	$< 2^{64}$	$< 2^{128}$	
Kích thước khối (bit)	512	512	1024	
Độ dài từ (bit)	32	32	64	
Số bước thực hiện (bước)	80	64	80	

Hình 2.3 Các phiên bản SHA

- Phần này chỉ mô tả thuật toán băm SHA-1, các phiên bản khác của SHA cũng được thiết kế theo nguyên lý tương tự.
- SHA-1 chấp nhận các khối thông tin có kích thước tối đa là 264 bit để tạo ra mã băm với độ dài cố định 160 bit. Tòan bộ khối thông tin được xử lý theo từng khối 512 bit, qua 5 công đoạn như sau:

B 1. Gắn bit đệm – Append padding bit: thông tin gốc được gắn thêm các bit thừa để có chiều dài (448 modulo 512) bit, tức là tất cả các khối trước có chiều dài bằng nhau là 512 bit, riêng khối cuối cùng là 448 bit. Chú ý rằng việc chèn thêm bit vào khối thông tin được thực hiện đối với tất cả các khối thông tin gốc, kể cả khi khối thông tin gốc có số bit chính xác bằng 448 mod 512 (khi đó chuỗi bit chèn vào sẽ có chiều dài là 512 bit).

B 2. Gắn chiều dài – Append length: một chuỗi 64 bit được gắn thêm vào khối thông tin. 64 bit này được xử lý như một số nguyên không dấu, cho biết chiều dài của khối thông tin gốc (tức chiều dài thật sự khi chưa thực hiện công đoạn 1). Sau công đoạn này, khối thông tin nhận được có chiều dài là bội số của 512 bit, được chia thành các nhóm, mỗi nhóm tương đương với 16 thanh ghi 32 bit (16*32 = 512 bit).

B 3. Khởi tạo bộ đệm MD – Initialize MD buffer: bộ đệm MD (message digest) là bộ nhớ có dung lượng 160 bit dùng để chứa các kết quả trung gian và kết quả cuối cùng của mã băm. Bộ nhớ này được tổ chức thành 5 thanh ghi 32 bit và được khởi tạo các giá trị ban đầu như sau (Hex):

A = 67452301

B = EFCDAB89

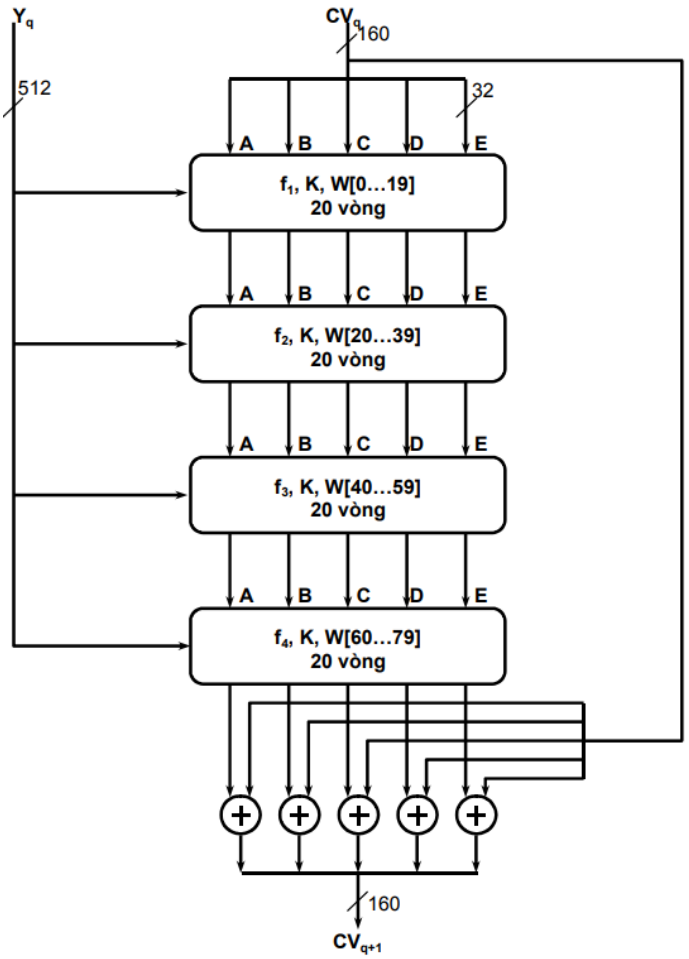
C = 98BADCFE

D = 10325476

E = C3D2E1F0

B 4. Xử lý thông tin theo từng khối 512 bit – Process message: đây là công đoạn trung tâm quả hàm băm, còn được gọi là hàm nén (compress function), bao gồm 4 vòng, mỗi vòng 20 bước.

- Hình 2.4 trình trình bày sơ đồ khối của bước 4.



Hình 2.4 Xử lý thông tin trong SHA-1

- Cả 4 vòng có cấu trúc tương tự nhau, nhưng mỗi vòng sử dụng một hàm luận lý khác nhau là f1, f2, f3 và f4.
- Ngõ vào của mỗi vòng là khối bit Y (512 bit) đang xử lý cùng với giá trị của bộ đệm MD. Mỗi vòng sử dụng một biến cộng Kt khác nhau, với 0 □ t □ 79 biểu diễn cho 80 bước của 4 vòng.

  Tuy nhiên, thực tế chỉ có 4 giá trị K khác nhau như sau:

Bước	**Giá trị K (Hexa)**
$0 \le t \le 19$	Kt = 5A827999
$20 \le t \le 39$	Kt = 6ED9EBA11
$40 \le t \le 59$	Kt = 8F1BBCDC
$60 \le t \le 79$	Kt = CA62C1D6

  - Ngõ ra của vòng thứ tư (tức bước 80) được cộng với ngõ vào của vòng đầu tiên để tạo ra CV_{q+1}. Thao tác cộng được thực hiện một cách độc lập, ứng với từng thanh ghi trong bộ đệm MD với một từ tương ứng trong CV_q, sử dụng phép cộng modulo 2^{32}.

B 5. Xuất kết quả - Output: Sau khi tất cả các khối 512 bit đã được xử lý, ngõ ra của bước cuối cùng chính là giá trị của mã băm.

- Một thuộc tính quan trọng của giải thuật băm SHA-1 là mỗi bit trong mã băm đều có quan hệ với tất cả các bit trong thông tin gốc. Việc lặp lại các hàm f một cách phức tạp như vậy nhằm mục đích đảm bảo rằng dữ liệu đã được trộn một cách kỹ lưỡng và do đó rất khó tìm được 2 khối thông tin gốc khác nhau có thể tạo ra cùng một mã băm.

2.3.Chữ ký số

2.3.1. Nguyên lý hoạt động của chữ ký số

- Chữ ký số là một cơ chế xác thực cho phép người tạo ra thông tin (message creator) gắn thêm một đoạn mã đặc biệt vào thông tin có tác dụng như một chữ ký. Chữ ký được tạo ra bằng cách áp dụng một hàm băm lên thông gốc, sau đó mã hóa thông tin gốc dùng khóa riêng của người gởi. Chữ ký số có mục đích đảm bảo tính tòan vẹn về nguồn gốc và nội dung của thông tin.
- Tại sao phải dùng chữ ký số trong khi các cơ chế xác thực thông tin (message authentication) đã thực hiện chức năng xác thực nguồn gốc thông tin? Các cơ chế xác thực thông tin sử dụng các hàm băm một chiều có tác dụng bảo vệ thông tin trao đổi giữa hai thực thể thông tin khỏi sự xâm phạm của một thực thể thứ 3, tuy nhiên nó không có tác dụng ngăn chặn được sự xâm phạm của chính hai thực thể.
- Ví dụ:

Thực thể A gởi một bản tin X cho thực thể B sử dụng một cơ chế xác thực nào đó, cơ chế này đảm bảo chỉ có A và B dùng chung một khoá bí mật K để tạo ra các mã xác thực từ thông tin gốc. Tuy nhiên, thực thể B có thể đổi bản tin X thành một bản tin Y, và với khóa bí mật K, thực thể B hòan tòan có thể tạo ra thông tin xác thực mới để gắn vào Y, làm cho nó trở thành một bản tin hợp lệ mặc dù thực chất đây không phải là bản tin do thực thể A tạo ra.

Một ví dụ khác, thực thể A có thể từ chối xác nhận việc mình đã gởi bản tin X cho thực thể B, vì với các cơ chế xác thực như trên, thực thể B hoàn toàn có khả năng giả mạo thông tin đưa ra từ thực thể A.

Giống như một chữ ký thông thường (chữ ký bằng tay), một chữ ký số phải có đầy đủ các thuộc tính sau đây:

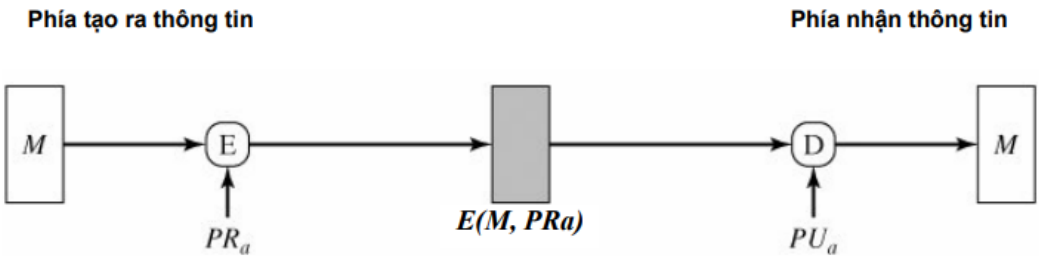
+ Phải xác nhận chính xác người ký và ngày giờ phát sinh chữ ký.
+ Phải xác thực nội dung thông tin ngay tại thời điểm phát sinh chữ ký.

+ Phải có khả năng cho phép kiểm chứng bởi một người thứ 3 để giải quyết các tranh chấp nếu có.

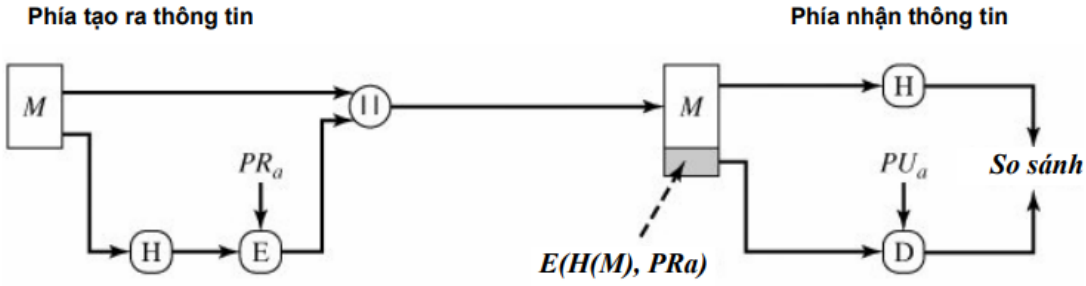
- Như vậy, chức năng của chữ ký số bao gồm chức năng của xác thực thông tin.
- Các yêu cầu đối với chữ ký số:
  + Là một chuỗi bit phát sinh từ khối thông tin cần được xác nhận (thông tin gốc).
  + Chữ ký phải chứa thông tin nhận dạng riêng của người ký để tránh giả mạo và tránh phủ nhận.
  + Quy trình tạo ra chữ ký cũng như xác minh chữ ký phải đơn giản, nhanh chóng
  + Chữ ký thông thể bị giả mạo bằng bất cứ cách nào.
  + Có thể sao chép một bản sao của chữ ký dành cho mục đích lưu trữ.
- **Phân loại chữ ký số**: Có nhiều thuật toán phát sinh chữ ký số khác nhau. Có thể phân loại các thuật toán này theo các cách như sau:
  + Chữ ký cố định và chữ ký ngẫu nhiên: thuật toán tạo chữ ký cố định (deterministic) tạo ra một chữ ký duy nhất ứng với một khối thông tin gốc xác định, nghĩa là nếu thực hiện nhiều lần thuật toán tạo chữ ký trên một bản tin thì vẫn cho ra một kết quả duy nhất. Ngược lại, chữ ký ngẫu nhiên (probabilistic) tạo ra những chữ ký khác nhau đối với cùng một bản tin.
  + Chữ ký phục hồi được và chữ ký không phục hồi được: cơ chế tạo chữ ký phục hồi được (reversible signature) cho phép người nhận phục hồi lại thông tin gốc từ chữ ký, điều này cũng có nghĩa là chữ ký phải có chứa thông tin gốc trong nó dưới một dạng mã hoá nào đó, và kết quả là chữ ký số sẽ có kích thước lớn hơn thông tin gốc. Khi đó, người gởi chỉ cần gởi đi chữ ký là đủ. Do vậy, cơ chế tạo chữ ký này cũng còn được gọi là chữ ký khôi phục bản tin (signature with message recovery). Ngược lại, cơ chế tạo chữ ký không phục hồi được (non-reversible signature) không cho phép phục hồi thông tin gốc từ chữ ký, do vậy, chữ ký chỉ là một khối thông tin cộng thêm có kích thước

nhỏ hơn thông tin gốc. Người gởi cần phải gởi chữ ký đi kèm với thông tin gốc như một dạng phụ lục, do đó cơ chế tạo chữ ký này cũng còn được gọi là chữ ký với phụ lục (signature with appendix).

- **Các phương pháp thực hiện chữ ký số**: Có hai phương pháp thực hiện chữ ký số là ký trực tiếp (direct signature) và ký thông qua trọng tài (arbitrated signature).
  + *Ký trực tiếp (direct signature):* Ở phương pháp này, giả thiết rằng phía nhận biết được khóa công khai của phía gởi. Do đó, chữ ký có thể được tạo ra bằng cách mã hóa tòan bộ bản tin bằng khóa riêng của người tạo ra thông tin, hoặc là chỉ mã hóa phần mã băm (kết quả tạo ta từ hàm băm đối với thông tin gốc) dùng khóa riêng của người tạo thông tin.



a- Tạo chữ ký trực tiếp bằng cách mã hóa tòan bộ thông tin gốc



b- Tạo chữ ký trực tiếp bằng cách mã hóa phần mã băm của thông tin gốc

M: thông tin gốc E: Thuật tóan mã hóa D: Thuật tóan giải mã
H: Hàm băm ||: Nối mã băm vào thông tin gốc
PRa: Khóa bí mật của người ký PUa: Khóa công khai của người ký

Hình 2.5 Chữ ký trực tiếp

- Để đạt được tính bảo mật của thông tin thì thông tin gốc cùng với chữ ký vừa được tạo ra sẽ được mã hóa sử dụng khóa công khai của thực thể nhận chữ ký (trong trường hợp dùng mật mã bất đối xứng) hoặc dùng khóa bí mật (trong trường hợp dùng mật mã đối xứng).

- Một nhược điểm rất dễ thấy của phương thức ký trực tiếp đó là độ an tòan của chữ ký phụ thuộc cao độ vào khóa riêng của người tạo ra chữ ký. Do vậy, nếu khóa riêng này bị mất hoặc bị tiết lộ thì ý nghĩa của chữ ký số sẽ không còn.

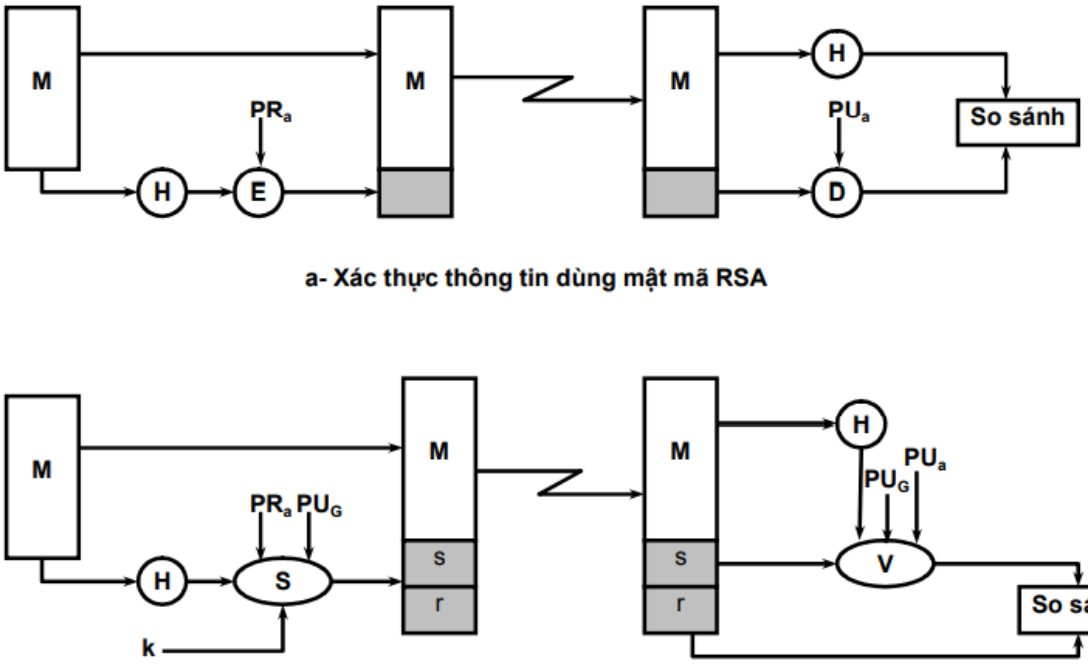
+ Ký thông qua trọng tài (arbitrated signature): đây là một giải pháp được xây dựng để khắc phục nhược điểm của chữ ký trực tiếp. Khi thực thể A muốn gởi một bản tin cho thực thể B, quá trình tạo ra một chữ ký được thực hiện bình thường như đối với chữ ký trực tiếp. Tuy nhiên, trước khi bản tin này được gởi đến B, nó phải được gởi đến một thực thể thứ 3 gọi là trọng tài (arbiter). Trọng tài thực hiện việc kiểm tra, xác nhận tính chính xác của thông tin và chữ ký, sau đó ghi lại ngày giờ rồi mới gởi cho thực thể B, kèm theo thông tin xác nhận của trọng tài. Sự xuất hiện của trọng tài trong quy trình đảm bảo được thực thể A sẽ không phủ nhận được thông tin mình đã gởi.

2.3.2. Chuẩn chữ ký DSS

- DSS (Digital Signature Standard) là một chuẩn về chữ ký số, được chuẩn hóa năm 1991, sửa đổi năm 1993 và 1996, sau đó mở rộng vào năm 2000. DSS sử dụng hàm băm SHA và thuật toán tạo chữ ký DSA (Digital Signature Algorithm). DSS thuộc loại chữ ký ngẫu nhiên và không phục hồi được.
- Hình 2.6 so sánh cấu trúc DSS so với phương thức xác thực thông tin sử dụng mật mã bất đối xứng RSA.
- Trong thuật toán xác thực thông tin dùng mật mã RSA, thông tin gốc được đưa vào hàm băm SHA để tạo ra mã băm (tức message digest) có kích thước cố định. Mã băm này sau đó được mã hóa (bằng thuật toán RSA) dùng khóa riêng của thực

thể tạo thông tin (phía gởi). Kết quả của phép mã hóa được gắn vào thông tin gốc và gởi đi. Phía thu nhận được thông tin, tách phần mã băm ra khỏi thông tin gốc và giải mã nó bằng khóa công khai của phía gởi. Chú ý rằng khóa công khai là thông tin được công bố rộng rãi cho bất kỳ thực thể nào có quan tâm. Đồng thời, thông tin gốc cũng được đưa vào hàm băm để tính mã băm, sau đó đem so sánh với mã băm vừa nhận được. Nếu hai mã này giống nhau thì thông tin vừa nhận được chấp nhận như là thông tin hợp lệ.

- Hoạt động của DSS cũng bao gồm việc đưa thông tin gốc vào hàm băm để tạo ra mã băm có kích thước cố định. Tuy nhiên, mã băm này sẽ không được mã hóa trực tiếp bằng một giải thuật mã hóa mà được sử dụng làm ngõ vào của một hàm tạo chữ ký S (Signature function). Các thông tin đưa vào hàm tạo chữ ký bao gồm:
  + Mã băm của thông tin gốc
  + Một số ngẫu nhiên k
  + Khóa riêng của người ký (PR_a)
  + Khóa công khai của nhóm các thực thể liên quan đến giao dịch chữ ký (PU_G).



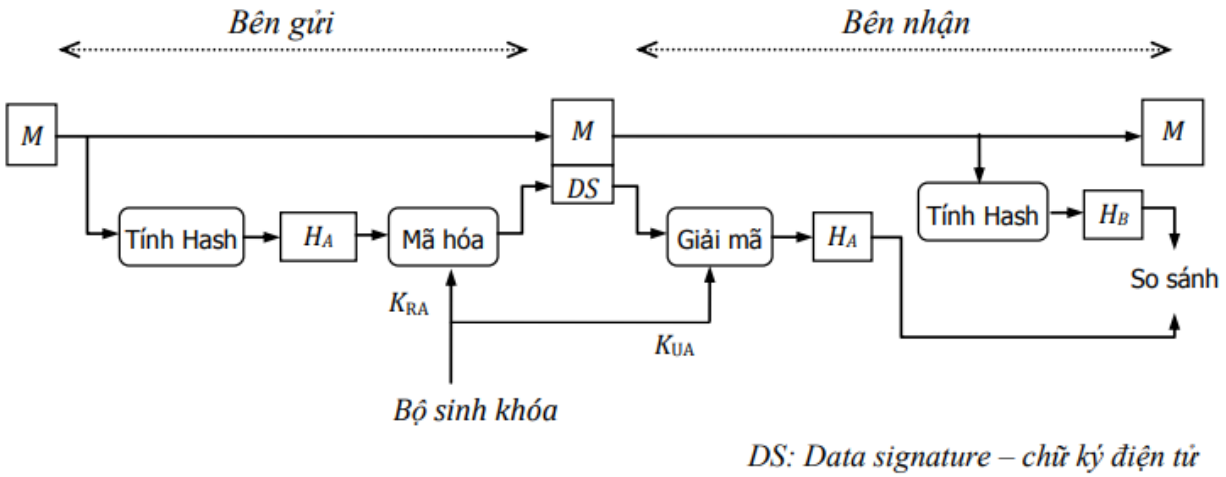
a- Xác thực thông tin dùng mật mã RSA

Hình 2.6 Xác thực thông tin RSA và dùng chữ ký số DSS

CHƯƠNG 3 XÂY DỰNG ỨNG DỤNG

3.1.Xác định mô hình

Mô hình chữ ký số RSA trong các hệ thống quản lý: Quá trình gửi và nhận các tệp văn bản phục vụ quản lý dựa vào thuật toán SHA-1 và thuật toán RSA.



Hình 3.1 Mô hình chữ ký số

- Quán trình ký và gửi các tệp văn bản.
  + Từ file cần gửi ban đầu, chương trình sẽ sử dụng hàm băm SHA-1 để mã hóa chuỗi ký tự dài 128 bit. Chương trình sử dụng thuật toán RSA để mã hóa giá trị băm thu được với khóa bí mật của người gửi để nhận được một giá trị gọi là chữ ký điện tử. Kết hợp file ban đầu với chữ ký điễn tử thành một thông điệp đã ký và gửi đi cho người nhận.
- Quá trình nhận tệp văn bản.
  + Sau khi người nhận nhận được văn bản. Hệ thống sẽ tách thông điệp đã ký ra thành file văn bản và chữ ký điện tử. Đến giai đoạn này có 2 quá trình kiểm tra:
  + Kiểm tra file văn bản có đúng người gửi hay không. Sử dụng thuật toán RSA để giải mã chữ ký điện tử bằng khóa công khai của người gửi. Nếu giải mã không được file thì file nhận được là không đúng người gửi. Nếu giải mã

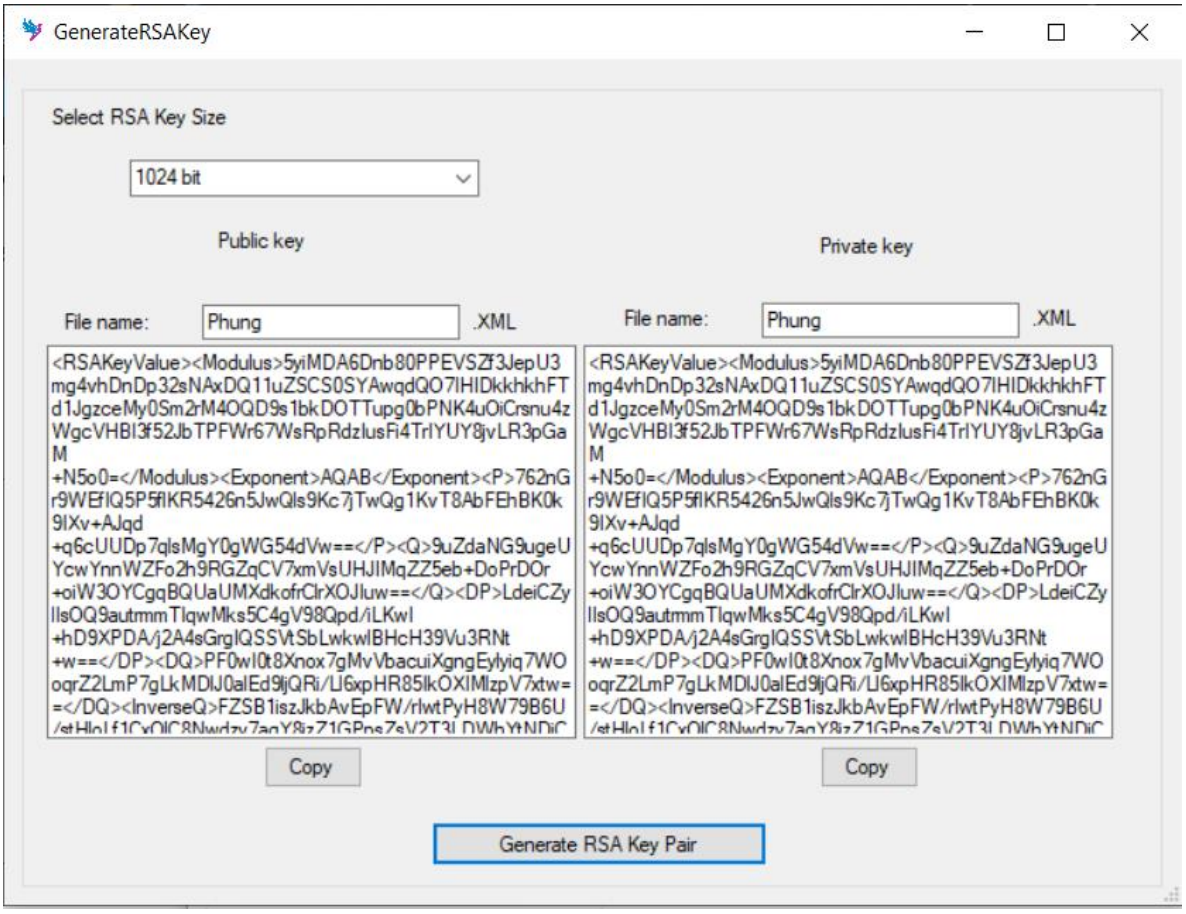
thành công thì file nhận được là đúng người gửi và ta nhận được giá trị băm 1.

+ Kiểm tra file văn bản có bị thay đổi hay không: Từ file văn bản ban đầu ta sử dụng hàm băm SHA-1 mã hóa thành giá trị băm 2. Kiểm tra giá trị băm 1 và giá trị băm 2 có giống nhau hay không? Nếu giống nhau thì file nhận được là toàn vẹn không bị thay đổi, nếu ngược lại thì file văn bản đã bị thay đổi.

3.2. Cài đặt

3.2.1. Modul tạo khóa.

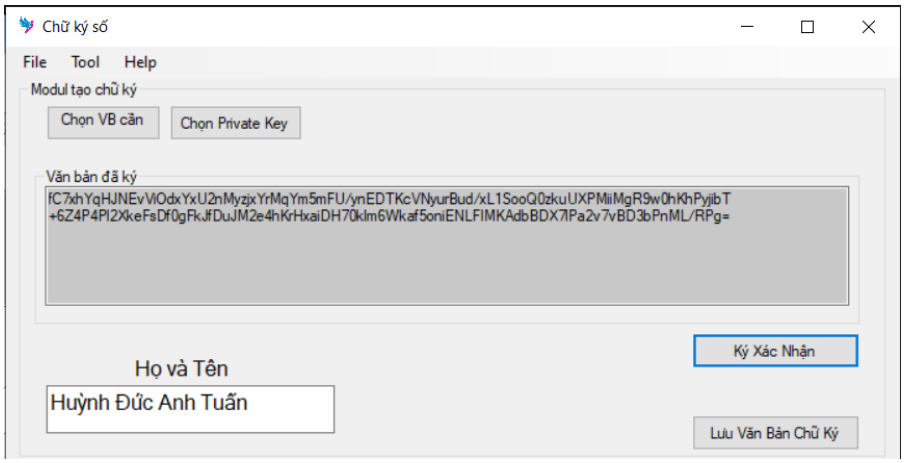
- Nhiệm vụ của modul này chính là tạo cặp khóa công khai và khóa bí mật cho người dùng. Mặc định chương trình tạo cặp khóa có độ dài 1024 bit.



Hình 3.2 Giao diện modul tạo cặp khóa

3.2.2. Modul tạo chữ ký cho file văn bản.

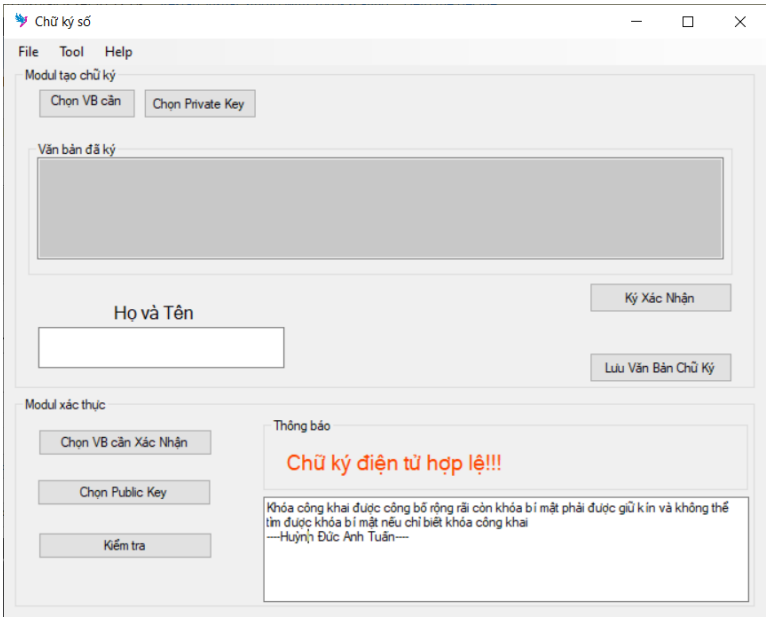
- Nhiệm vụ cơ bản của modul này là tạo ra file chữ ký (*sig) có tên trùng với tên file được chọn để ký. Người dùng muốn sử dụng chứng năng này phải thông qua việc tạo khóa.



Hình 3.3 Giao diện modul tạo chữ ký số

3.2.3. Modul kiểm tra xác thực.

- Đầu vào của modul này là file cần xác thực + với chữ ký của nó (thông điệp). Nhiệm vụ của modul này chính là kiểm tra tính đúng đắn của chữ ký số.



Hình 3.4 Giao diện xác thực chữ ký số

TÀI LIỆU THAM KHẢO

1. An toàn và Bảo mật thông tin – Trần Minh Văn
2. Bảo mật hệ thống thông tin – Lê Phúc
3. Cryptography and Network Security Principles and Practices - William Stallings