ĐẠI HỌC QUỐC GIA HÀ NỘI

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

VƯƠNG THỊ HẠNH

NGHIÊN CỨU CÁC PHƯƠNG PHÁP MẬT MÃ ĐẢM BẢO TÍNH TOÀN VẸN DỮ LIỆU TRONG TRƯỜNG HỌC THÔNG MINH

LUẬN VĂN THẠC SĨ NGÀNH HỆ THỐNG THÔNG TIN

HÀ NỘI 2017

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

VƯƠNG THỊ HẠNH

NGHIÊN CỨU CÁC PHƯƠNG PHÁP MẬT MÃ ĐẢM BẢO TÍNH TOÀN VẸN DỮ LIỆU TRONG TRƯỜNG HỌC THÔNG MINH

Ngành**:** Hệ thống thông tin
Chuyên ngành: Hệ thống thông tin
Mã số: 60.48.01.04

LUẬN VĂN THẠC SĨ NGÀNH HỆ THỐNG THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. LÊ PHÊ ĐÔ
TS. PHÙNG VĂN ỔN

HÀ NỘI 2017

LỜI CẢM ƠN

Trước tiên tôi xin gửi lời cảm ơn sâu sắc nhất đến thầy Lê Phê Đô và thầy Phùng văn Ổn. Các thầy đã tận tâm, tận lực hướng dẫn, định hướng phương pháp nghiên cứu khoa học cho tôi, đồng thời cũng đã cung cấp nhiều tài liệu và tạo điều kiện thuận lợi trong suốt quá trình học tập và nghiên cứu để tôi có thể hoàn thành luận văn này.

Tôi xin được gửi lời cảm ơn đến các thầy, cô trong bộ môn Hệ thống thông tin và khoa công nghệ thông tin, Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội đã nhiệt tình giảng dạy và truyền đạt những kiến thức, kinh nghiệm quý giá trong suốt thời gian tôi học tập tại trường.

Tôi xin gửi lời cảm ơn đến các bạn học viên lớp K22-HTTT, những người đồng hành trong suốt khóa học và có nhiều góp ý bổ ích cho tôi. Cảm ơn gia đình, bạn bè đã quan tâm và động viên giúp tôi có nghị lực phấn đấu để hoàn thành tốt luận văn này.

Do kiến thức và thời gian có hạn nên luận văn chắc chắn không tránh khỏi những thiếu sót nhất định.

Một lần nữa xin gửi lời cảm ơn chân thành và sâu sắc.

Hà Nội, tháng 7 năm 2017
Học viên thực hiện

Vương Thị Hạnh

LỜI CAM ĐOAN

Luận văn thạc sĩ đánh dấu cho những thành quả, kiến thức tôi đã tiếp thu được trong suốt quá trình rèn luyện, học tập tại trường. Tôi xin cam đoan luận văn "***Nghiên cứu các phương pháp mật mã đảm bảo tính toàn vẹn dữ liệu trong trường học thông minh".*** được hoàn thành bằng quá trình học tập và nghiên cứu của tôi dưới sự hướng dẫn của TS. Lê Phê Đô và TS. Phùng văn Ổn.

Trong toàn bộ nội dung nghiên cứu của luận văn, các vấn đề được trình bày là những tìm hiểu và nghiên cứu của cá nhân tôi hoặc là trích dẫn các nguồn tài liệu và một số trang web đều được đưa ra ở phần Tài liệu tham khảo.

Tôi xin cam đoan những lời trên là sự thật và chịu mọi trách nhiệm trước thầy cô và hội đồng bảo vệ luận văn thạc sĩ.

Hà Nội, tháng 7 năm 2017

Vương Thị Hạnh

DANH MỤC VIẾT TẮT

TT	Từ viết tắt	Tiếng anh	Tiếng Việt
1	RSA	Rivest Shamir Adleman	Thuật toán mật mã hóa khóa công khai
2	SHA 3	Secure Hash Algorithm -3	Thuật toán băm an toàn SHA3
3	SHAKE	Secure Hash Algorithm Keccak	Thuật toán băm an toàn Keccak
4	CKĐT	Electronic Signature	Chữ ký điện tử
5	CA	Certificate Authority	Ủy quyền chứng chỉ
6	UCLN		Ước chung lớn nhất
7	CNTT		Công nghệ thông tin
8	PRF	pseudo random function	Chức năng giả ngẫu nhiên
9	MDC	Modification Detection Code	Mã phát hiện sửa đổi
10	NFC	Near Field Communications	Công cụ kết nối không dây
11	MAC	Message Authentication Code	Mã xác thực thông điệp
12	HMAC	Hash Message Authentication Code	Mã xác thực thông điệp bởi hàm băm
13	NIST	National Institute of Standards and Technology	Viện tiêu chuẩn và công nghệ Quốc gia (Mĩ)
14	CRHF	Collision Resistant Hash Function	Hàm băm kháng xung đột
15	OWHF	One way hash function	Hàm băm một chiều
16	IP	Internet Protocol	Giao thức giao tiếp mạng Internet
17	LFSR	Linear Feedback Shift Register	Thanh ghi phản hồi dịch tuyến tính
18	IV	Initial Value	Giá trị ban đầu

MỤC LỤC

MỤC LỤC HÌNH

MỤC LỤC BẢNG

MỞ ĐẦU

1. Tính cấp thiết của đề tài luận văn

Mô hình trường học thông minh là trường học có các thiết bị hiện đại bao gồm: Máy chủ, máy chiếu Projector, màn hình LCD, camera ghi hình cùng với hệ thống internet được kết nối đồng bộ. Sử dụng các phần mềm hỗ trợ học tập, phần mềm quản lý học tập có thể giao tiếp hai chiều giữa giáo viên, học sinh và gia đình. Giải pháp trường học thông minh đã được triển khai thành công tại nhiều trường học ở Mỹ, Trung Đông và một số nước Châu Âu, Châu Á. Trường học thông minh tạo môi trường tốt cho giáo viên và học sinh học tập; nâng cao chất lượng dạy và học.

Tuy nhiên, cùng với lợi ích của việc sử dụng phần mềm quản lý học sinh thông qua mạng internet là vấn đề mất an toàn thông tin như: mất mát dữ liệu, rò rỉ thông tin làm ảnh hưởng nghiêm trọng đến nhà trường, thầy cô và học sinh. Vì vậy đảm bảo an toàn thông tin trường học thông minh là nhiệm vụ rất quan trọng mà tôi đề cập trong luận văn này. Trường học Việt Nam tuy có quan tâm nhiều nhưng chưa được toàn diện, nên mục tiêu và đối tượng mà tôi hướng đến là "***Nghiên cứu các phương pháp mật mã đảm bảo tính toàn vẹn dữ liệu trong trường học thông minh".***

2. Mục đích nghiên cứu:

❖ Mục tiêu của đề tài.

- Nghiên cứu mô hình trường học thông minh.
- Nhận diện những thách thức và biện pháp giải quyết đảm bảo toàn vẹn dữ liệu nói chung và toàn vẹn dữ liệu trường học nói riêng.
- Xây dựng chương trình mô phỏng

❖ Đối tượng và phạm vi nghiên cứu.

- Các trường học và lớp học thông minh.

❖ Phương pháp nghiên cứu.

- Tìm hiểu các mô hình lớp học thông minh trên thế giới và Việt Nam cũng như các nguy cơ của công nghệ ảnh hưởng đến trường học.
- Tìm hiểu các phương pháp mật mã để đảm bảo toàn vẹn dữ liệu trường học.
  - Phương pháp sử dụng hàm băm SHA
  - Phương pháp dùng mã xác thực dữ liệu MAC
  - Phương pháp dùng chữ ký số

3. Nội dung của đề tài, các vấn đề cần giải quyết.

a. Hướng nghiên cứu:

- Mô hình trường học thông minh
- Các nguy cơ và các công nghệ đảm bảo an toàn dữ liệu.

b. Nội dung:

Chương 1: An toàn thông tin ở trường học thông minh

Chương 2: Các phương pháp mật mã đảm bảo toàn vẹn dữ liệu

Chương 3: Ứng dụng chữ ký điện tử đảm bảo tính toàn vẹn dữ liệu trong trường học

Chương trình demo.

CHƯƠNG 1: AN TOÀN THÔNG TIN Ở TRƯỜNG HỌC THÔNG MINH

1.1 TỔNG QUAN VỀ TRƯỜNG HỌC THÔNG MINH

Bước sang thế kỷ 21, công nghệ thông tin được ứng dụng mạnh trong quá trình tổ chức đào tạo, thay đổi nội dung, phương pháp giảng dạy bám sát yêu cầu thực tiễn theo xu thế chung thế giới hình thành trường học nền tảng số hóa. Công nghệ là động lực then chốt cho sự phát triển giáo dục, chuyển đổi mô hình giáo dục thụ động sang giáo dục thông minh. Trường học thông minh đã triển khai thành công ở nhiều nước Mỹ, Trung Đông và một số quốc gia Châu Âu, Châu Á. Giáo dục ngày một phát triển kéo theo có nhiều phần mềm giáo dục ra đời, sách điện tử và các ứng dụng khác sẵn có ngày một đa dạng và phong phú hơn. Trường học thông minh so với phương pháp giảng dạy truyền thống là tối ưu hóa về thiết bị, tài liệu. Điều này tạo giao tiếp hai chiều giữa giáo viên và học sinh thuận lợi hơn. Việc truyền đạt của giáo viên dễ dàng, đầy đủ và đa dạng, học sinh tiếp nhận kiến thức dễ hiểu hơn, sâu hơn, chủ động hơn.

Lớp học thông minh – trường học thông minh chú trọng vào giờ dạy tương tác và quản lý học tập. Giảng dạy tương tác đó là giáo viên – học sinh có thể trao đổi bài giảng, tài liệu, bài tập, câu hỏi – trả lời,… thông qua màn hình tương tác. Ngoài ra, giáo viên có thể quản lý, theo dõi được quá trình học tập của từng em thông qua tính năng quản lý từ máy tính PC giáo viên.

- Giảng dạy tương tác: hỗ trợ bằng cách sử dụng các chức năng chia sẻ màn hình, màn hình giám sát các hoạt động nhóm, bài kiểm tra và thăm dò ý kiến,…
- Quản lý học tập: hỗ trợ giáo viên lập kế hoạch quản lý khóa học, bài học.

Hầu hết các phòng học được kết nối internet thông qua wifi hoặc băng thông rộng không dây và có trang thiết bị máy tính để bàn, máy tính xách tay, máy tính bảng,..

Mục tiêu của nhà giáo dục là cải tiến cơ bản giáo dục thông qua công cụ giảng dạy tương tác trong và ngoài lớp học.

Hình 1.1 Mô hình lớp học thông minh

1.2 XÂY DỰNG TRƯỜNG HỌC THÔNG MINH Ở VIỆT NAM

Xây dựng mô hình "Trường học thông minh" nhằm tối ưu hóa các thiết bị dạy học điện tử hiện đại như: Máy tính chủ kết nối màn hình tương tác điện tử, máy chiếu Projecto hoặc màn hình LCD, camera ghi hình, laptop, ibad, …cùng với hệ thống

internet được kết nối đồng bộ giữa bục giảng thông minh có bộ xử lý điện tử tiêu chuẩn.

- Tạo môi trường tốt cho tất cả học sinh học tập có chất lượng, hướng học sinh vào sự đa dạng, phong phú của tri thức nhân loại.
- Là nơi phát huy tính tích cực, trí thông minh của học sinh trên con đường nhận thức.
- Tạo môi trường làm việc hiện đại cho Ban giám hiệu nhà trường có thể quản lý chặt chẽ được từng giáo viên, từng học sinh… bằng một phần mềm quản lý học tập thông qua hệ thống internet.

Trên cơ sở các chi phí, thông tư hướng dẫn và nhiệm vụ từng năm học của các cấp, của ngành về ứng dụng công nghệ thông tin; kế hoạch phát triển tổng thể nguồn nhân lực và triển khai kế hoạch xây dựng mô hình trường hoc thông minh, lớp học thông minh với một số nội dung sau:

- Bám sát nội dung các chỉ thị, nghị quyết, quyết định, thông tư hướng dẫn của nhà nước, Bộ GD&ĐT. Các văn bản chỉ đạo của ngành để làm căn cứ xây dựng kế hoạch thực hiện mô hình "Lớp học thông minh" từng bước theo lộ trình phù hợp với điều kiện của nhà trường và địa phương;
- Nội dung kế hoạch và việc triển khai thực hiện kế hoạch xây dựng "trường học thông minh" cần đảm bảo: "Công nghệ hiện đại -Tính năng vượt trội - Thân thiện dễ sử dụng", tiết kiệm, có hiệu quả, đảm bảo tính khả thi cao;
- Việc lựa chọn các trang thiết bị, phần mềm cho lớp học thông minh và quản lý của nhà trường thực hiện theo các hướng dẫn của huyện, tỉnh và của ngành,

Tuy nhiên, khó khăn lớn nhất hiện nay là kinh phí. Điều này thể hiện rõ trong việc ngành giáo dục khuyến khích xây dựng trường học thông minh và chỉ được hỗ trợ 50% kinh phí, còn 50% là nhà trường xã hội hóa. Một số trường ĐH lớn và các trường ở trung tâm được sự hỗ trợ của công ty Samsung đã xây dựng hệ thống trường học thông minh. Còn lại các trường bình dân, ở huyện hay miền núi thì khó hay không thể thực hiện được.Vì số tiền đầu tư vào trang thiết bị không hề nhỏ, trong khi đó ngân sách của nhà trường thì rất hạn hẹp. Điển hình là các trường xây dựng mô hình "trường học thông minh" như: Đại họcY Dược TP. HCM, trường tiểu học Hoàng Hoa Thám, trường THPT Trần Phú,…

Các tiêu chí " Trường học thông minh" như:

1. Có lắp đặt hệ thống internet băng thông rộng.
2. Có ít nhất 30% số lớp học trong nhà trường đạt tiêu chí là "Lớp học thông minh"
3. Có hệ thống camera giám sát tại các lớp học.
4. Quản lý nhà trường theo mô hình hiện đại, ứng dụng các phần mềm quản lý trực tuyến trong công tác điều hành nhà trường, phát huy tốt các chức năng trang thông tin điện tử của nhà trường.

Dưới đây là hình ảnh giảng đường ĐH Y Dược TP/CHM

Hình 1.2: Giảng đường thông minh trường ĐH Y Dược TP.HCM

1.3 CÁC NGUY CƠ MẤT AN TOÀN THÔNG TIN TRONG TRƯỜNG HỌC

Trong những năm qua, tình hình an toàn thông tin trên thế giới nói chung và Việt Nam nói riêng tiếp tục diễn biến phức tạp. Tuy nhiên các tổ chức giáo dục dường như vẫn chưa thực sự được quan tâm, trong khi đây là môi trường chứa rất nhiều thông tin nhạy cảm.

Điểm khác biệt trong việc quản lý một hệ thống trường học thông minh gồm nhiều hệ điều hành sẽ khó kiểm soát hơn. Do đó nhà trường cần phải thay đổi cách tiếp cận, chuyển từ phòng vệ sang giám sát mạng để phát hiện kịp thời dữ liệu bị truy xuất bất hợp pháp. Không những thế, việc theo dõi chặt chẽ dữ liệu vào ra qua hệ thống giúp nhà trường nhận biết khi nào và biết trước được những bất ổn vừa mới hoặc đang xảy ra, nhằm bám sát vụ tấn công ngay từ khi mới xuất hiện để giảm thiểu những tổn thất, thiệt hại do nó mang lại.

Các tài liệu khi được truyền tin trên mạng thường đối mặt với nguy cơ bị mất an toàn như: bị truy cập bất hợp pháp, sao chép, lưu trữ hoặc chuyển đến cho những người không được phép.

Bằng cách sử dụng công nghệ kỹ thuật như bắt gói tin trên đường truyền, thâm nhập trực tiếp vào máy tính chứa dữ liệu quan trọng, hacker có thể dễ dàng lấy được các văn bản này. Việc lấy cắp, truy cập lại càng dễ dàng hơn nếu những cá nhân có mục đích xấu này lại là những người có hiểu biết về công nghệ thông tin hoặc là những người quản trị hệ thống, quản trị ứng dụng trong cơ quan.

Đối với các tài liệu có các thông tin bí mật khi trao đổi trong hệ thống mà không có biện pháp nào để bảo vệ thì nguy cơ bị mất ATTT là vô cùng lớn và như vậy hậu quả của việc mất ATTT là không thể lường được. Điều gì sẽ xảy ra nếu các tài liệu, thông tin này lọt vào tay đối thủ cạnh tranh hay những cá nhân, tổ chức có mục đích xấu.

1.3.1 Những mối đe dọa về an toàn thông tin trong trường học

Dữ liệu số là dữ liệu cốt lõi của các hoạt động trong trường học, chúng đảm bảo hệ thống máy tính vận hành tốt. Dữ liệu trường học cũng đa dạng và được xếp vào ba nhóm chính:

- Dữ liệu thuộc về tài sản trí tuệ của trường cần được lưu trữ, truy cập và sử dụng thích hợp để phục vụ công tác học tập, nghiên cứu hay thương mại. Dữ liệu này gồm các chương trình đào tạo, giảng dạy, các nghiên cứu của giáo viên, học sinh.
- Dữ liệu có được do liên kết với các tổ chức bên ngoài như tổ chức y tế, chính trị, các viện nghiên cứu hoặc doanh nghiệp, cơ quan thương mại ngoài nước. Việc liên kết với nhiều tổ chức khác nhau không còn xa lạ với các trường Đại học – Cao đẳng, tùy thuộc vào mục đích liên kết mà nhà trường có được thông tin từ tổ chức và ngược lại.
- Dữ liệu phát sinh do hoạt động nhà trường gồm các thông tin giáo viên, sinh viên, nhân viên, số liệu tài chính. Nhóm dữ liệu này có thể coi là thông tin pháp luật nhạy cảm và là thông tin mang rủi ro lớn.

Nguồn dữ liệu nhà trường đa dạng và có nhiều đối tượng có thể truy cập khiến nhà trường phải đối mặt với rất nhiều rủi ro. Những kỹ thuật tấn công ngày càng đang dạng và phong phú. Dữ liệu từ giáo viên, nhân viên, học sinh truy cập là không hề nhỏ rất nhiều trong số đó là các thông tin nhạy cảm. Bất kỳ dữ liệu riêng tư nào bị truy cập trái phép hoặc sử dụng sai mục đích cũng sẽ dẫn đến hậu quả không thể lường trước, một số ảnh hưởng có thể kể đến như:

- Danh tiếng: các thông tin về hành vi trộm cắp, phá hoại dữ liệu số có thể gây hại nghiêm trọng tới danh tiếng của trường trong mắt học sinh – phụ huynh, các đối tác, các doanh nghiệp và chính phủ.
- Pháp lý: việc đánh mất thông tin nhạy cảm khiến nhà trường có nguy cơ bị truy tố, xử phạt và phải bồi thường thiệt hại.
- Kinh tế: đánh mất thông tin về các nghiên cứu, sở hữu trí tuệ hoặc chuyển giao công nghệ có thể gây thiệt hại rất lớn về kinh tế, trực tiếp làm suy yếu đến trường học.
- Hoạt động: các cuộc tấn công có thể làm tê liệt hệ thống, cản trở hoạt động của nhà trường, ảnh hưởng tới cơ sở hạ tầng của nhà trường.

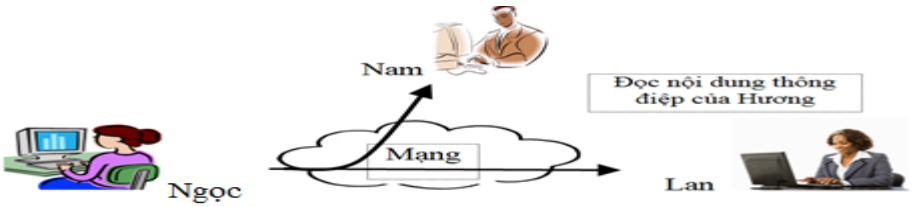
Tùy thuộc vào quy mô của từng trường mà lượng lưu trữ, mức độ quan trọng của thông tin cũng như hậu quả khi bị tấn công sẽ khác nhau.

1.3.2 Những loại hình tấn công dữ liệu

Để xem xét những vấn đề bảo mật thông tin trên đường truyền mạng, chúng ta lấy bối cảnh sau: ba nhân vật tên là Ngọc, Lan và Nam, trong đó Ngọc và Lan thực hiện trao đổi thông tin với nhau còn Nam là kẻ xấu, đặt thiết bị can thiệp vào kênh truyền tin giữa Ngọc và Lan. Sau đây là các loại hành động tấn công của Nam ảnh hưởng đến quá trình truyền tin.

- *Xem trộm thông tin*

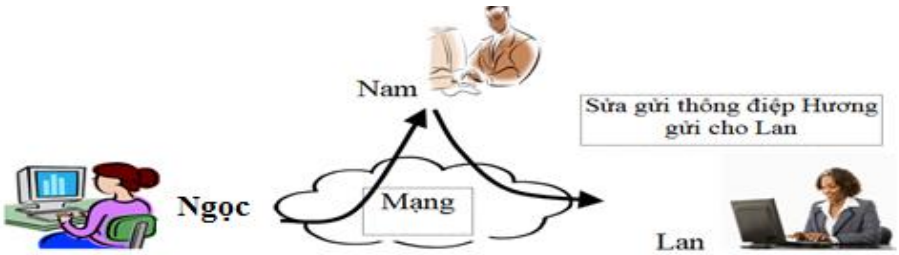
Trong trường hợp này Nam chặn các thông điệp Ngọc gửi cho Lan, và xem nội dung của thông điệp



Hình 1.3 Xem trộm thông điệp

- *Thay đổi thông điệp*

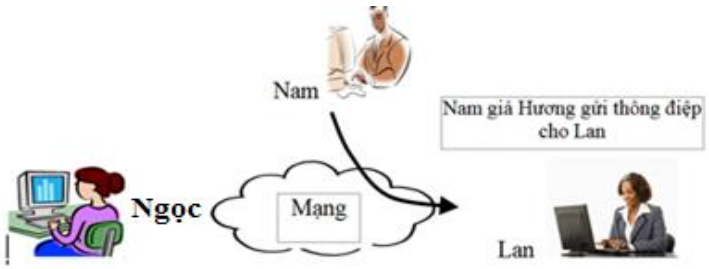
Nam chặn các thông điệp Ngọc gửi cho Lan và ngăn không cho các thông điệp này đến đích. Sau đó Nam thay đổi nội dung của thông điệp và gửi tiếp cho Lan. Lan nghĩ rằng nhận được thông điệp nguyên bản ban đầu của Ngọc mà không biết rằng chúng đã bị sửa đổi.



Hình 1.4 Sửa thông điệp

- *Mạo danh*

Trong trường hợp này Nam giả là Ngọc gửi thông điệp cho Lan. Lan không biết điều này và nghĩ rằng thông điệp là của Ngọc.

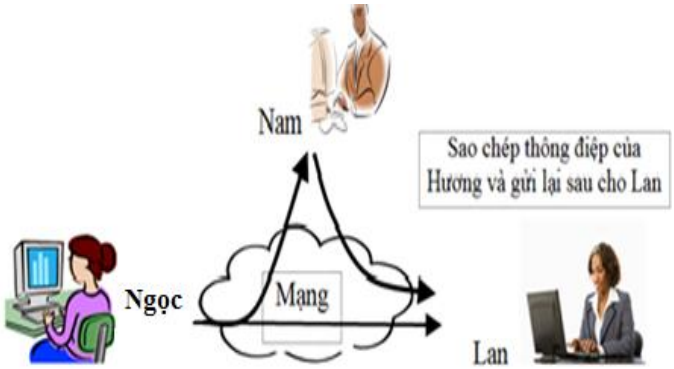


Hình 1.5 Mạo danh

- *Phát lại thông điệp*

Nam sao chép lại thông điệp Ngọc gửi cho Lan. Sau đó một thời gian Nam gửi bản sao chép này cho Lan. Lan tin rằng thông điệp thứ hai vẫn là từ Ngọc, nội dung hai thông điệp là giống nhau. Thoạt đầu có thể nghĩ rằng việc phát lại này là vô hại, tuy nhiên trong nhiều trường hợp cũng truy ra tác hại không kém so với việc giả mạo thông điệp.

Xét tình huống sau: giả sử Lan là nhân viên ngân hàng còn Ngọc là một khách hàng. Ngọc gửi thông điệp đề nghị Lan chuyển cho Nam 2000$. Nếu Nam sao chép và phát lại thông điệp, Lan tin rằng Ngọc gửi tiếp một thông điệp mới để chuyển thêm cho Nam 2000$ nữa.

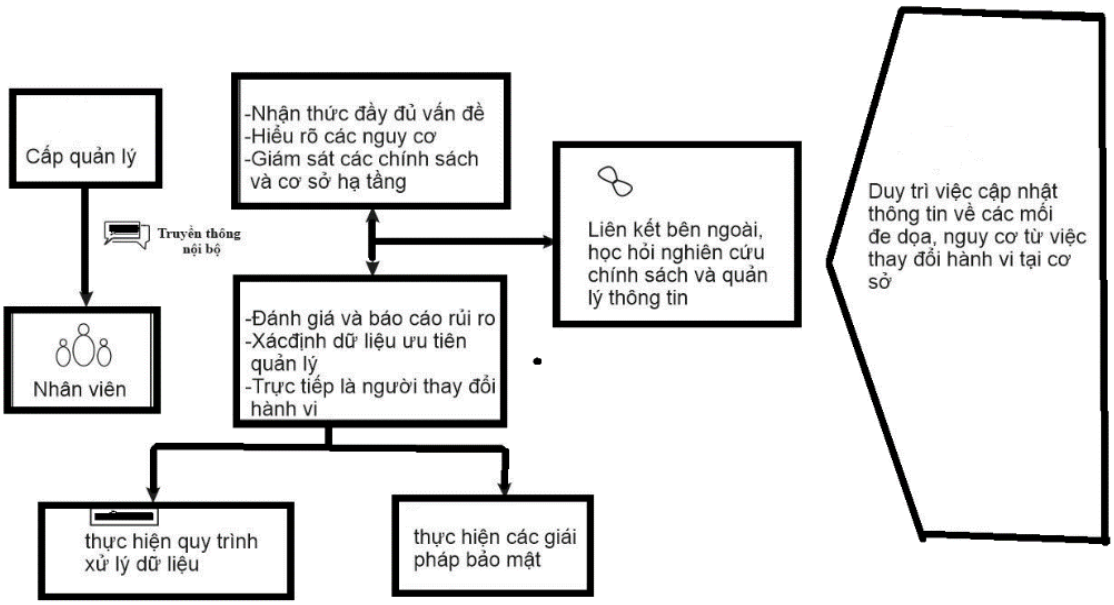


Hình 1.6 Sao chép thông điệp

1.4 GIẢI PHÁP ĐẢM BẢO AN TOÀN THÔNG TIN TRONG TRƯỜNG

Để dữ liệu trong trường học được an toàn, mạng internet được ổn định lâu dài thì cán bộ quản lý hệ thống công nghệ thông tin trường học cần thực hiện tốt các nhiệm vụ sau:

- *Đánh giá rủi ro* bằng cách xác định các nhóm dữ liệu, phát hiện lỗ hổng và thiết lập các ưu tiên quản lý.
- *Xây dựng và thực hiện nghiêm túc* các giải pháp, chính sách quản trị dữ liệu, thiết lập các lớp an toàn mạng.
- *Giám sát và báo cáo định kỳ* dựa trên đặc thù của từng đơn vị, như thay đổi hành vi, nhận thức người dùng, đồng thời chia sẻ và cập nhật kiến thức thường xuyên.



Hình 1.7 Qui trình quản lý ATTT

1.4.1 Đánh giá rủi ro

Đánh giá, phân loại được dữ liệu là bước đầu tiên trong quá trình xây dựng kế hoạch ứng phó phù hợp. Vai trò của dữ liệu như thế nào? Tầm ảnh hưởng của chúng đến đâu? Đó là những điều cần làm rõ để đánh giá rủi ro. Những dữ liệu quan trọng có thể chia thành ba nhóm giá trị sau:

- Dữ liệu sử dụng nội bộ: những dữ liệu này sử dụng trong các hoạt động học tập, nghiên cứu và làm việc của nhà trường như tài liệu trong thư viện, giáo trình, các nghiên cứu đã được công bố.
- Dữ liệu liên quan đến pháp luật, các hợp đồng có hiệu lực: việc quản lý số lượng học sinh, liên kết với các trường, tổ chức doanh nghiệp bên ngoài đòi hỏi nhà trường cần nhận thức đầy đủ trách nhiệm của mình với những dữ liệu là định danh, thông tin cá nhân, doanh nghiệp, hợp đồng đã kí kết qua các chương trình hợp tác với nhà trường.
- Dữ liệu có giá trị kinh tế hoặc chính trị: xác định loại dữ liệu này là thách thức lớn nhất đối với nhà trường. Có thể không nhiều người có quyền biết và sử dụng nhóm dữ liệu này, tuy nhiên việc quản lý và thực hiện chuyển giao kiến thức không hề đơn giản.

1.4.2 Xây dựng và thực hiện nghiêm túc các giải pháp quản trị dữ liệu, thiết lập các lớp an toàn mạng.

Quản lý có hiệu quả là yếu tố then chốt cho sự thành công trong việc đảm bảo ATTT của các tổ chức trong mọi lĩnh vực không riêng các trường học. Cán bộ, đặc biệt đội ngũ quản lý hệ thống công nghệ cần phải nhận thức được nhiệm vụ của mình liên quan đến việc bảo vệ dữ liệu số và có biện pháp thích hợp để đảm bảo rằng chúng

phù hợp với quy định của pháp luật. Tuy nhiên, mỗi đơn vị nhà trường sẽ có cấu trúc khác nhau vì vậy khi thiết lập chính sách, đội ngũ quản trị phải tìm cách trả lời đầy đủ các câu hỏi: Dữ liệu nhạy cảm nằm ở đâu? Ai sở hữu hoặc kiểm soát dữ liệu? Thiết lập quản lý dữ liệu như thế nào? Các kênh theo dõi và quản lý rủi ro là gì?

Một số chính sách nhà trường tham khảo:

- Đánh giá, lường trước những rủi ro khi mất ATTT, có phương án phản ứng kịp thời.
- Đảm bảo không có kênh thông tin liên lạc không rõ ràng giữa các bộ phận nắm giữ, điều khiển dữ liệu quan trọng khác.
- Xem xét vai trò của kiểm toán nội bộ và bên ngoài để đánh giá hiệu quả công tác quản trị dữ liệu và quản lý an ninh mạng.
- Xem xét việc thành lập một ban quản trị chuyên duy trì giám sát, quản lý dữ liệu về thể chế và rủi ro an ninh mạng.

Đối với các trường học thông minh phương pháp phân đoạn sẽ mang lại hiệu quả cao hơn cho quản lý dữ liệu. Tuy nhiên, nhà trường cũng có thể xem xét xây dựng "kho dữ liệu" tập trung, thực hiện toàn bộ chính sách và công nghệ có tiêu chuẩn an ninh trên "kho dữ liệu" đó. Giải pháp này có nhược điểm là chi phí cao, đòi hỏi kinh nghiệm quản lý các loại dữ liệu khác nhau.

1.4.3 Giám sát và báo cáo định kỳ

Hiệu quả quản lý hệ thống công nghệ sau khi đã đi vào hoạt động phụ thuộc hoàn toàn vào quá trình giám sát, từ đó đưa ra những báo cáo tổng hợp để có được cái nhìn toàn diện. Dựa trên kết quả đó, ban quản lý sẽ nhìn ra ưu, nhược điểm để có những điều chỉnh phù hợp. Công việc này cần phải thực hiện định kỳ.

Do tính chất phát triển nhanh chóng nên có nhiều mối đe dọa, quá trình rà soát liên tục và cập nhật thực tế cần thực hiện thường xuyên. Hợp tác với doanh nghiệp, tổ chức, các trường trong và ngoài nước để chia sẻ kiến thức, học hỏi kinh nghiệm.

- Phối hợp chặt chẽ giữa hiệu trưởng, chuyên gia nghiên cứu và nhân viên quản lý dữ liệu, an ninh mạng để hiểu rõ các mối đe dọa, kịp thời đưa ra giải pháp.
- Khuyến khích trao đổi giữa các phòng ban, xây dựng diễn đàn trường học, học hỏi kinh nghiệm để tăng cường hiểu biết các mối đe dọa đã từng gặp phải và cùng nhau phòng tránh hoặc tìm cách giải quyết.
- Đưa ra các chương trình đào tạo, liên kết giữa giáo viên và học sinh trong trường, tích hợp thực hành quản lý an toàn thông tin vào chương trình đào tạo.

Việc đảm bảo ATTT trong các tổ chức nói chung và trường học nói riêng là một thách thức không hề đơn giản, nhất là khi các mối đe dọa đang không ngừng phát triển về cách thức, quy mô. Các tổ chức cần một hệ thống quản lý phù hợp để bảo vệ thông tin nhạy cảm. Hệ thống này phải đảm bảo: có thể xác định, đánh giá và giám sát rủi ro

an ninh mạng. Thực hiện quản lý dữ liệu hiệu quả và an toàn. Kiểm soát hệ thống CNTT sát sao, cập nhật thường xuyên các kiến thức mới.

Bảo mật thông tin là trách nhiệm của toàn bộ tổ chức, dựa trên sự phối hợp hoạt động giữa các cấp lãnh đạo, quản trị dữ liệu, hệ thống và người dùng.

1.4.4 Kiểm soát truy cập internet

Cán bộ quản lý CNTT của nhà trường cần đảm bảo hệ thống mạng được an toàn. Dưới đây là một số vấn đề cho máy chủ ứng dụng và dịch vụ:

- Đặt các máy chủ trong vùng DMZ (là một vùng mạng trung lập giữa mạng nội bộ và mạng internet). Thiết lập filewall không cho các kết nối với máy chủ trên toàn bộ các cổng ngoại trừ được sử dụng cho các dịch vụ và ứng dụng mà máy chủ sử dụng.
- Loại bỏ toàn bộ các dịch vụ không cần thiết khỏi máy chủ. Mỗi dịch vụ không cần thiết sẽ bị lợi dụng để tấn công hệ thống nếu không có chế dộ bảo mật tốt.
- Không cho phép quản trị hệ thống từ xa, trừ khi nó được đăng nhập theo kiểu mật khẩu chỉ sử dụng một lần hay đường kết nối đã được mã hóa.
- Giới hạn số người có quyền quản trị hay truy cập mức tối đa
- Tạo các log file theo dõi hoạt động của người sử dụng và duy trì các log file này trong môi trường được mã hóa.
- Hệ thống điều khiển log file thông thường được sử dụng cho bất kỳ hoạt đông nào. Cài đặt các bẫy macro để giám sát các cuộc tấn công vào máy chủ. Tạo các macro chạy liên tục hoặc ít ra có thể kiểm tra tính nguyên vẹn của file paswword và các file hệ thống khác. Khi các macro kiểm tra sự thay đổi, chúng gửi email tới nhà quản lý hệ thống.

1.4.5 Đảm bảo an toàn thông tin bằng phương pháp mật mã

Mật mã là một ngành khoa học chuyên nghiên cứu các phương pháp truyền tin bí mật bao gồm hai quá trình: *mã hóa* và *giải mã*. Để bảo vệ thông tin trên đường truyền người ta thường biến đổi thông tin trước khi truyền đi trên mạng gọi là *mã hoá* thông tin, nơi nhận phải thực hiện quá trình *giải mã* thông tin. Để bảo vệ thông tin bằng mật mã theo hai hướng: theo đường truyền (Link Oriented Security) và từ nút đến nút (end_ to_end).

- *Theo đường truyền* thông tin được mã hóa để bảo vệ trên đường truyền giữa hai nút mà không quan tâm đến nguồn và đích của thông tin đó. Lưu ý rằng đối với cách này thông tin chỉ được bảo vệ trên đường truyền, tức là ở mỗi nút đều có quá trình giải mã sau đó mã hóa để truyền đi tiếp, do đó các nút cần phải được bảo vệ tốt.
- *Từ nút đến nút* thông tin trên mạng được bảo vệ trên toàn đường truyền từ nguồn đến đích. Thông tin sẽ được mã hóa ngay sau khi mới tạo ra và chỉ được giải mã

khi về đến đích. Cách này có nhược điểm là chỉ có dữ liệu của người dùng thì mới có thể mã hóa được, còn dữ liệu điều khiển thì giữ nguyên để có thể xử lý tại các nút

Vai trò của hệ mật mã.

- Dùng để che giấu nội dung của văn bản rõ: đảm bảo sao cho chỉ người chủ hợp pháp của thông tin mới có quyền truy cập thông tin, hay nói cách khác là chống truy nhập không đúng quyền hạn.
- Tạo các yếu tố xác thực thông tin: đảm bảo thông tin lưu hành trong hệ thống đến người nhận hợp pháp là xác thực. Tổ chức các sơ đồ chữ ký điện tử, đảm bảo không có hiện tượng giả mạo, mạo danh để gửi thông tin trên mạng.

CHƯƠNG 2. CÁC PHƯƠNG PHÁP MẬT MÃ ĐẢM BẢO TOÀN VẸN DỮ LIỆU

2.1 HỆ MẬT MÃ

2.1.1 Định nghĩa hệ mật mã

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

- P (Plaintext) là một tập hợp hữu hạn các bản rõ và được gọi là không gian bản rõ.
- C (Ciphertext) là tập hữu hạn các bản mã được gọi là không gian các bản mã.
- K (Key) là tập hữu hạn các khóa hay còn gọi là không gian khóa. Đối với mỗi phần tử k của K được gọi là một khóa. Số lượng của không gian khóa phải lớn để không có đủ thời gian thử mọi khóa.
- E (Encrytion) là tập hợp các qui tắc mã hóa có thể.
- D (Decrytion) là tập hợp các qui tắc giải mã có thể.

Đối với mỗi k ∈ K có một quy tắc mã e_k: $P \rightarrow C$ và một quy tắc giải mã tương ứng $d_k \in D$. Mỗi e_k: $P \rightarrow C$ và d_k: $C \rightarrow P$ là những hàm mà: $d_k(e_k(x)) = x$ với mỗi $x \in P$.

Chúng ta biết thông tin thường được tổ chức dưới dạng bản rõ. Người gửi sẽ làm nhiệm vụ mã hóa bản rõ, kết quả thu được gọi là bản mã. Bản mã này được gửi đi trên một đường truyền tới người nhận, sau khi nhận được bản mã người nhận giải mã nó để nhận văn bản.

Thuật toán dùng khi sử dụng định nghĩa hệ mật mã:

$$E_k(P) = C, D_k(C) = P$$

2.1.2 Những yêu cầu đối với một hệ mật mã.

- *Độ tin cậy*: Cung cấp sự bí mật cho các thông tin và dữ liệu được lưu bằng việc sử dụng các kỹ thuật mã hóa.
- *Tính toàn vẹn*: Cung cấp sự bảo đảm với tất cả các bên rằng thông tin không bị thay đổi từ khi gửi cho tới khi người nhận.
- *Không bị chối bỏ*: Người gửi không thể từ chối việc đã gửi thông tin đi.
- *Tính xác thực*: Người nhận có thể xác minh được nguồn tin mình nhận được là đúng đối tác của mình gửi hay không?

2.1.3 Phân loại hệ mật mã

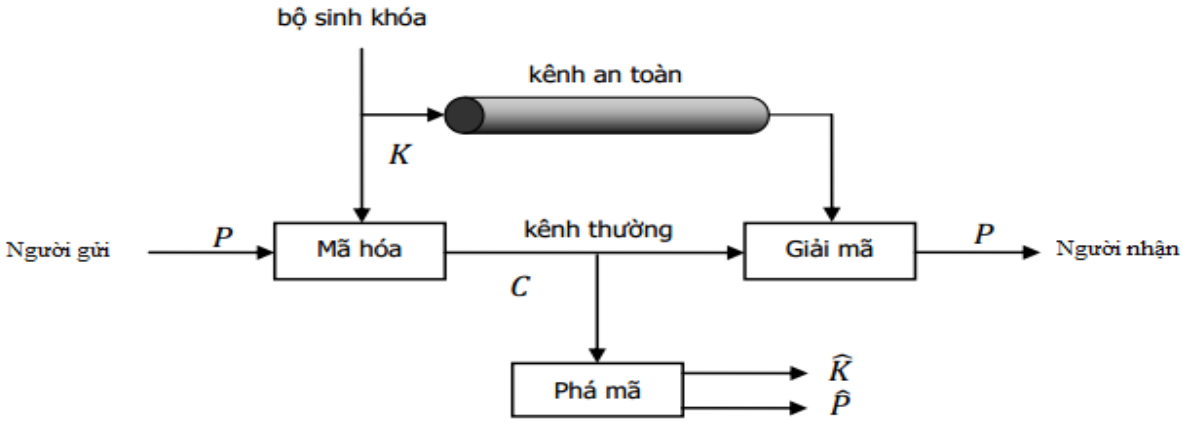
Dựa vào cách sử dụng khóa có thể phân các hệ mật mã thành hai loại:

- *Hệ mật khóa đối xứng* là những hệ mật dùng chung một khóa cả trong quá trình mã hóa dữ liệu và giải mã dữ liệu. Do đó khóa mật mã cần phải có một đường truyền an toàn để truyền đưa khóa mật mã từ phía người truyền tới phía người nhận. Tính an toàn của mật mã này là sự an toàn của khóa mật mã. Nếu như khóa mật mã bị tiết lộ, thì hệ thống mật mã này sẽ bị phá vỡ.

- *Hệ mật khóa bất đối xứng* là hệ mật có 2 khóa, một khóa để mã hóa và một khóa để giải mã là khác nhau. Khóa dùng để mã hóa có thể công khai nhưng khóa dùng để giải mã phải giữ bí mật. Trong các hệ mật này, việc phân phối và thỏa thuận khóa được giải quyết một cách tự động, nếu một người trong hệ thống muốn gửi thông điệp cho B, họ lấy khóa công khai của B trên mạng để mã hóa thông điệp và gửi đến cho B, B dùng khóa bí mật của mình để giải mã thành thông điệp ban đầu.

2.2. HỆ MÃ KHÓA ĐỐI XỨNG

Phương pháp Ceasar là phương pháp mã hóa đơn giản nhất của mã hóa đối xứng. phương pháp mã hóa đối xứng được biểu diễn bằng mô hình sau:



Hình 2.1 Mô hình mã hóa đối xứng

Mô hình gồm 5 yếu tố:

- Bản rõ P (plaitext)
- Thuật toán mã hóa E (encrypt alorithm)
- Khóa bí mật K (secret key)
- Bản mã C (cipheetext)
- Thuật toán giải mã D (decrypt algorithm)

Trong đó: C = E (P,K)

P = D (C,K)

Thuật toán mã hóa và giải mã sử dụng chung một khóa, thuật toán giải mã là phép toán ngược của thuật toán mã hóa. Vì vậy mô hình trên gọi là phương pháp mã hóa đối xứng. Bản mã C được gửi đi trên kênh truyền. Do bản mã C đã được biến đổi so với bản rõ P, cho nên người thứ ba can thiệp vào kênh truyền có thể lấy được bản mã C thì vẫn không hiểu được ý nghĩa của bản mã. Đây chính là đặc điểm quan trọng của mã hóa, cho phép đảm bảo tính bảo mật của một hệ truyền tin. Một đặc tính quan trọng của mã hóa đối xứng là khóa phải được giữ bí mật giữa người gửi và người nhận, hay nói cách khác khóa phải được chuyển một cách an toàn từ người gửi đến người nhận.

Đặc tính quan trọng thứ hai của một hệ mã hóa đối xứng là tính an toàn của hệ mã. Từ một bản mã có thể dễ dàng suy ra được bản rõ ban đầu mà không cần biết khóa bí mật. Hành động đi tìm bản rõ từ bản mã mà không cần khóa như vậy được gọi là hành động phá mã. Do đó một hệ mã hóa đối xứng được gọi là an toàn khi và chỉ khi nó không thể bị phá mã hoặc thời gian phá mã là bất khả thi.

Mã hóa đối xứng có thể được phân thành hai loại:

- Loại thứ nhất tác động trên bản rõ theo từng nhóm bits. Các thuật toán áp dụng được gọi là mã hóa khối (Block Cipher). Theo đó, từng khối dữ liệu trong văn bản ban đầu được thay thế bằng một khối dữ liệu khác có cùng độ dài. Ngày nay các thuật toán có kích thước chung một khối là 64 bits.
- Loại thứ hai tác động lên bản rõ theo từng bits một. Các thuật toán áp dụng được gọi là mã hóa dòng (Stream Cipher). Dữ liệu của văn bản được mã hóa từng bits một. Các thuật toán mã hóa dòng này có tốc độ nhanh hơn các thuật toán mã hóa khối và nó thường được áp dụng khi lượng dữ liệu cần mã hóa chưa biết trước.

Một số thuật toán nổi tiếng trong mã hóa đối xứng là: DES, Triple DES (3DES), RC4, AES…

- DES (Data Encryption Standard), bản rõ (Plaitext) được mã hóa theo từng khối 64 bits và sử dụng một khóa 64 bits, nhưng thực tế thì chỉ có 56 bits là thực sự dùng để tạo khóa, 8 bits còn lại dùng để kiểm tra tính chẵn, lẻ. DES là một thuật toán được sử dụng rộng rãi nhất trên thế giới. Hiện tại DES không còn được đánh giá cao do kích thước của khóa nhỏ 56 bits, và dễ dàng bị phá vỡ.
- Triple DES (3DES): 3DES cải thiện độ mạnh của DES bằng việc sử dụng một quá trình mã hóa và giải mã sử dụng 3 khóa. Khối 64 bits của bản rõ đầu tiên sẽ được mã hóa sử dụng khóa thứ nhất. Sau đó, dữ liệu bị mã hóa được giải mã bằng việc sử dụng khóa thứ hai. Cuối cùng, mã hóa lần nữa với chìa khóa thứ ba để thu được mã hóa cuối.

$$C = E_{K3}(D_{K2}(E_{K1}(P)))$$
$$P = D_{K1}(E_{K1}(D_{K3}(C)))$$

- AES (Advanced Encryption Standard) được sử dụng để thay thế cho DES. Nó hỗ trợ độ dài của khóa từ 128 bits cho đến 256 bits. AES được thiết kế bởi Joan Daemen và Vicent Rijmen, hai nhà khoa học người Bỉ. Phương pháp mã hóa này thích hợp ứng dụng trên nhiều hệ thống khác nhau, từ các thẻ thông minh cho đến máy tính cá nhân. Các thông tin tuyệt mật sẽ phải dùng khóa 192 hoặc 256 bits. Cấu trúc toán học của AES không giống với các thuật toán khác, AES có cấu trúc thuật toán khá đơn giản, rõ ràng. Hiện tại cấu trúc này chưa dẫn đến mối nguy hiểm nào nhưng một số nhà nghiên cứu cho rằng sẽ có nguy hiểm trong tương lai.

Ưu nhược điểm mã hóa khóa đối xứng

- **Ưu điểm**
  - Giải mã và mã hóa nhanh hơn hệ mã hóa khóa công khai.
- **Nhược điểm**

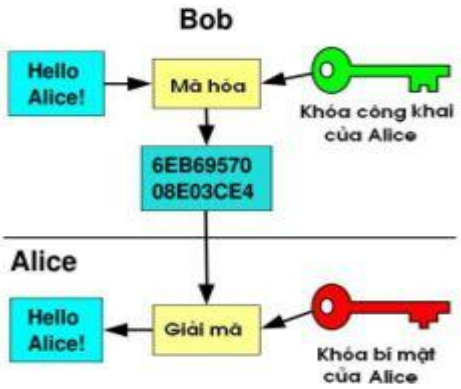
Vấn đề thỏa thuận khóa và quản lý khóa chung là khó khăn và phức tạp. Người gửi và người nhận phải luôn thống nhất với nhau về khóa. Việc thay đổi khóa là rất khó và dễ bị lộ. Khóa chung phải được gửi cho nhau trên kênh an toàn.

2.3 HỆ MÃ KHÓA BẤT ĐỐI XỨNG

2.3.1 Giới thiệu chung

Ý tưởng của hệ mật công khai được Diffie và Hellman đưa ra năm 1976. Còn việc thực hiện hệ mật công khai thì do Rivest, Shamir và Adleman đưa ra đầu tiên năm 1977, họ đề xuất hệ mật RSA[8]. Một số hệ mật khác được công bố sau đó, độ mật của chúng dựa trên tính toán khác nhau, dựa trên độ khó của bài toán phân tích thành nhân tử như hệ mật RSA, dựa trên độ khó của bài toán logarithm rời rạc hệ mật Elgamal, hệ mã dựa trên đường cong Elliptic.

Sơ đồ và nguyên tắc mã và giãi mã của hệ mật công khai.



Hình 2.2. Sơ đồ mã hóa khóa công khai

Hệ mã công khai sử dụng hai khóa có quan hệ toán học với nhau, tức là một khóa này được hình thành từ khóa kia: Người nhận bản mã tạo ra một khóa mật và từ khóa mật tính ra khóa công khai với một thủ tục không phức tạp, còn việc tìm khóa mật khi biết khóa công khai là bài toán khó giải được. Khóa công khai sẽ đưa đến cho người gửi bản tin qua kênh công cộng. Và bản tin được mã hóa bằng khóa công cộng. Bản mã truyền đến người nhận và nó được giải mã bằng khóa mật.

Mật mã khóa công khai cung cấp cơ chế chữ ký số được xác thực với độ bảo mật cao, thông điệp mà người nhận được là chính xác được gửi đi từ người gửi. Các chữ ký như vậy được coi là chữ ký số tương đương với chữ ký thật trên các tài liệu được in ra giấy. Sử dụng hợp thức các thiết kế có chất lượng cao và các bổ sung khác tạo ra độ an toàn cao.

Ưu điểm và nhược điểm của hệ mã hóa khóa công khai

Ưu điểm:

Đơn giản trong việc lưu chuyển khóa: Chỉ cần đăng ký một khóa công khai và mọi người sẽ lấy khóa này về để trao đổi thông tin với người đăng ký. Vì vậy không cần kênh bí mật để truyền khóa.

Mỗi người có cặp khóa công khai – khóa bí mật là có thể trao đổi thông tin với tất cả mọi người. Là tiền đề cho sự ra đời của chữ ký điện tử và các phương pháp chứng thực điện tử.

Nhược điểm:

Mã hóa và giải mã chậm hơn hệ mã hóa khóa đối xứng.

2.3.2 Hệ mật RSA

Thuật toán được Ron Rivest, Adi Shamir và Len Adleman mô tả lần đầu tiên vào năm 1977 tại học viện công nghệ Masachusetts (MIT) [9]. Đây là thuật toán đầu tiên phù hợp với việc tạo ra chữ ký điện tử. Nó đánh dấu một sự tiến bộ vượt bậc của lĩnh vực mật mã học trong việc sử dụng khóa công khai. Về mặt tổng quát RSA là một phương pháp mã hóa theo khối.Trong đó bản rõ M và bản mã C là các số nguyên từ 0 đến 2^i với i số bits của khối. Kích thước thường dùng của i là 1024 bits. RSA sử dụng hàm một chiều phân tích một số thành thừa số nguyên tố. RSA đang được sử dụng phổ biến trong thương mại điện tử đảm bảo an toàn với điều kiện độ dài khóa đủ lớn.

RSA đáp ứng đầy đủ yêu cầu bảo mật thông tin nên được sử dụng trong nhiều phần mềm bảo mật của hệ thống thư điện tử và hệ thống quản lý, điều hành, tác nghiệp. Hiện tại hệ mật mã RSA đảm bảo tính riêng tư và xác thực dữ liệu số. Để đảm bảo an toàn thông tin trong quá trình gửi/nhận Email và truyền tải văn bản qua mạng khi sử dụng hệ mật mã RSA là giải pháp tốt và an toàn nhất hiện nay.

2.3.2.1 Nguyên tắc thực hiện của RSA

Để thực hiện mã hóa và giải mã, RSA dùng phép lũy thừa module của lý thuyết số.

Các bước thực hiện như sau [4]:

1) Chọn hai số nguyên tố lớn p và q và tính $N=p\,q$. Cần chọn p và q sao cho: $M < 2^{i-1} < N < 2^i$. Với i = 1024 thì N là một số nguyên dài khoảng 309 chữ số
2) Tính $n = (p-1)(q-1)$
3) Tìm một số e sao cho e nguyên tố cùng nhau với n
4) Tìm một số d sao cho $d = e^{-1} \bmod n$
5) Lưu giữ bí mật các hệ số n, p và q . Chọn khóa công khai K_u (e, N), khóa bí mật $K_R(d, N)$.
6) Việc mã hóa thực hiện theo công thức:
   - Theo phương án 1, mã hóa bảo mật: $C = E(M, K_U) = M^e \bmod N$

- Theo phương án 2, mã hóa chứng thực: $C = E(M, K_R) = M^d \ mod \ N$

7) Việc giải mã thực hiện theo công thức:

- Theo phương án 1, mã hóa bảo mật: $\bar{M} = D(C, K_R) = C^d \ mod \ N$
- Theo phương án 2, mã hóa chứng thực: $\bar{M} = D(C, K_U) = C^e \ mod \ N$

Bản rõ M có kích thước i-1 bit, bản mã C có kích thước i bit.

Để đảm bảo rằng RSA thực hiện đúng theo nguyên tắc của mã hóa khóa công khai, ta phải chứng minh hai điều sau:

Bản giải mã chính là bản rõ ban đầu: $\bar{M} = M$, xét phương án 1:

Từ bước 4 ta suy ra:

$ed = kn + 1$ với *k* là một số nguyên nào đó

Vậy $\bar{M} = C^d \ mod \ N$

$= M^{ed} \ mod \ N$

$= M^{kn+1}$ mod N

$= M^{k(p-1)(q-1)+1}$ mod N

Ta chứng minh: $M^{k(p-1)(q-1)+1} \equiv M \ mod \ p$. Xét hai trường hợp của M:

- M chia hết cho p: $M \ mod \ p = 0$ do đó $M^{k(p-1)(q-1)+1} \equiv M \equiv 0 \ mod \ p$
- M không chia hết cho p, vì p là số nguyên tố nên suy ra M nguyên tố cùng nhau với p. vậy:

$d \ M^{k(p-1)(q-1)+1} \ mod \ p = M.(M^{p-1})^{k(q-1)} \ mop$

$= M.1^{k(q-1)} \ mod \ p$ (theo định lý Fermat)

$= M \ mod \ p$

Vậy: $M^{k(p-1)(q-1)+1} \equiv M \equiv 0 \ mod \ p$ với mọi M. hay nói cách khác $M^{k(p-1)(q-1)+1} - M \ chia \ hết \ cho \ p$. Chứng minh tương tự ta có $M^{k(p-1)(q-1)+1} - M$ chia hết cho q. Vì p,q là hai số nguyên tố nên suy ra $M^{k(p-1)(q-1)+1} - M$ chia hết cho N = p.q

Tóm lại:

$$M^{k(p-1)(q-1)+1} \equiv M \ mod \ N$$

Suy ra

$$\bar{M} = M^{k(p-1)(q-1)+1} \ mod \ N = M \text{ (do M<N)}$$

Vì e và d đối xứng nên có thể thấy trong phương án 2, ta cũng có $\bar{M} = M$.

Không thể suy ra từ K_R từ K_u, nghĩa là tìm cặp(d,N) từ cặp (e,N):

Có e và N, muốn tìm d, ta phải dựa vào công thức: $e.d \equiv 1 \ mod \ n$. Do đó phải tính được n. Vì n = (p-1)(q-1) nên suy ra phải tính được p và q. Vì N =pq nên ta chỉ có thể tính được p và q từ N. Tuy nhiên điều này là bất khả thi vì N =pq là hàm một chiều. Vậy không thể tính được K_R từ K_U.

2.3.2.2 Sơ đồ

Bước 1. ***Tạo cặp khóa*** (***bí mật, công khai***) (**a**, **b**)

Input: 2 số nguyên tố lớn phân biệt p và q.

Output: Cặp (n,b) là khóa công khai.

Cặp (n,a) là khóa bí mật.

Thuật toán

1. Chọn bí mật số nguyên tố lớn p và q.
2. Tính n = p * q, công khai n, đặt P=C = Zn.
3. Tính bí mật $\phi(n) = (p-1).(q-1)$.
4. Chọn khóa công khai $b < \phi(n)$, nguyên tố với $\phi(n)$.

Khóa bí mật a là phần tử nghịch đảo của b theo mod $\phi(n)$ tức là

$a*b \equiv 1(\text{mod } \phi(n))$.

5. Tập cặp khóa (bí mật, công khai) $K = \{(\mathbf{a}, \mathbf{b})/a,b \in Z_n$,

$a*b \equiv 1 (\text{mod}\phi(n))\}$.

Bước 2. ***Ký số***:

Chữ ký trên $x \in P$ là $y = \text{Sig k}(x) = x^a \pmod n$, $y \in A$.

Bước 3. ***Kiểm tra chữ ký***:

Verk (x,y) = đúng $\Leftrightarrow x \equiv y^b \pmod n$.

2.3.2.3 Ví dụ RSA

Bước 1. Tạo khoá (bí mật, công khai)

Chọn 2 số nguyên tố p = 1009, q = 1153

1. n = p*q = 1009*1153 = 1163377
2. $\phi(n) = (p-1) * (q-1) = 1008*1152 = 1161216$
3. Chọn b = 607 gcd (b, $\phi(n)$) = gcd (607, 1161216) = 1
4. Tính a:

   $a = b^{-1} \text{ mod } 1161216 = 835999$ (bằng thuật toán Euclide mở rộng)
5. Khóa công khai = (n,b) = (1163377,607)

   Khoá bí mật = (n, a) = (1163377,835999).

Bước 2. Mã hóa

Để mã hóa văn bản có giá trị x= 298, ta thự hiện phép tính:

$c = x^b \pmod n = 298^{607} \text{ mod } 1163377 = 549320$

Bước 3. **Giải mã**

Để giải mã văn bản có giá trị 549320, ta thực hiện phép tính:

$y = c^a = 549320^{835999} \bmod 1163377 = 298$

2.3.3 Hệ mật Elgama

Hệ mật Elgama hình thành trên cơ sở bài toán logarith rời rạc. Được đề xuất năm 1984, sau đó chuẩn chữ ký điện tử của Mỹ và Nga hình thành trên cơ sở hệ mật này.

2.3.3.1 Nguyên tắc hoạt động của khóa Elgama

Giả sử A và B muốn trao đổi thông tin mật với nhau bằng hệ mật Elgama thì A thực hiện quá trình hình thành khóa như sau[4,6]:

1. Chọn số nguyên tố đủ lớn p có chiều dài là k sao cho bài toán logarithm rời rạc trong z_p là khó giải.

2. Chọn $\propto \in Z_p^*$ là phần tử nguyên thủy. Chọn x là số ngẫu nhiên sao cho $1 < x < p$.

3. Tính giá trị y thỏa mãn công thức: $y = \propto^x mod\ (p)$.

Khóa mật là x, còn khóa công khai là 3 số $(\propto, p, y)$.

2.3.3.2 Quá trình mã hóa bản tin

1. Chọn số ngẫu nhiên K.
2. Tính $C' = \propto^K\ mod\ (p)$,
3. Tính

$C'' = Ty^K\ mod\ (N)$,

4. Người gửi bãn mã gồm (C', C'') đến người nhận.

2.3.3.3 Quá trình giải mã

1. Tính giá trị:

$$Z = (C')^x$$

2. Tính giá trị nghịch đảo của Z

$$Z^{-1} mod\ (p)$$

3. Giải mã theo bản mã C'':

$$T' = C''.Z^{-1}\ mod\ (p).$$

$$T.\ Y^k.\propto^{-kx}\ mod\ p = T.\propto^{kx}.\propto^{-kx} mod\ p = T$$

Nhận được bản tin T ban đầu.

2.3.3.3 Ví dụ Elgama

Giả sử chọn p = 2357. $\propto = 2,\ X_A = 1751.$

Tính khóa công khai Y_A

$$Y_A = \propto^{X_A}\ (mod) = 2^{1751}\ mod\ 2357 = 1185$$

Khóa công khai của A (p=2357; $\propto = 2$; 1185)

Ký lên bức điện m = 2035, chọn k ngẫu nhiên (k =1520)

Khi đó:

$C_1 = \propto^K mod\ p = \ 2^{1520}\ mod\ 2357 = 1430$

$C_2 = (m\ \times\ Y_A^K)\ mod\ p = (2035\ \times 1185^{1520}) mod\ 2357 = 697$

Người nhận B gửi cho C_1 = 1430 và C_2 = 697 cho A

Giải mã: Để giải mã A cần tính

$$(C_1^{X_A})^{-1}\ mod\ p = \left(C_1^{(p-1-Y_A)}\right) mod\ p$$

$$= \left(1430^{(2357-1-1751)}\right) mod\ 2357 = 872$$

$$m = \left(C_2\ \times \left(C_1{}^{Y_A}\right)^{-1}\right) mod\ p = (872\ \times 697) mod\ 2357 = 2035$$

2.4 CÁC PHƯƠNG PHÁP ĐẢM BẢO TÍNH TOÀN VẸN DỮ LIỆU BẰNG HÀM BĂM

2.4.1 Giới thiệu hàm băm mật mã

Khái niệm

Hàm băm mật mã là hàm toán học chuyển đổi một thông điệp có độ dài bất kỳ thành một dãy bits có độ dài cố định (tùy thuộc vào thuật toán băm). Dãy bits này được gọi là thông điệp rút gọn (message digest) hay giá trị băm (hash value), đại diện cho thông điệp ban đầu[6].

Hàm băm không phải là mã hóa vì nó không được giải mã về văn bản ban đầu, mà nó có chức năng băm "một chiều", và có một kích thước cố định cho bất kỳ một văn bản gốc. Đối tượng chính của hàm băm hướng tới là tính toàn vẹn dữ liệu. Băm thông điệp xem dữ liệu đó có bị thay đổi/ hay đúng là thông điệp gốc.

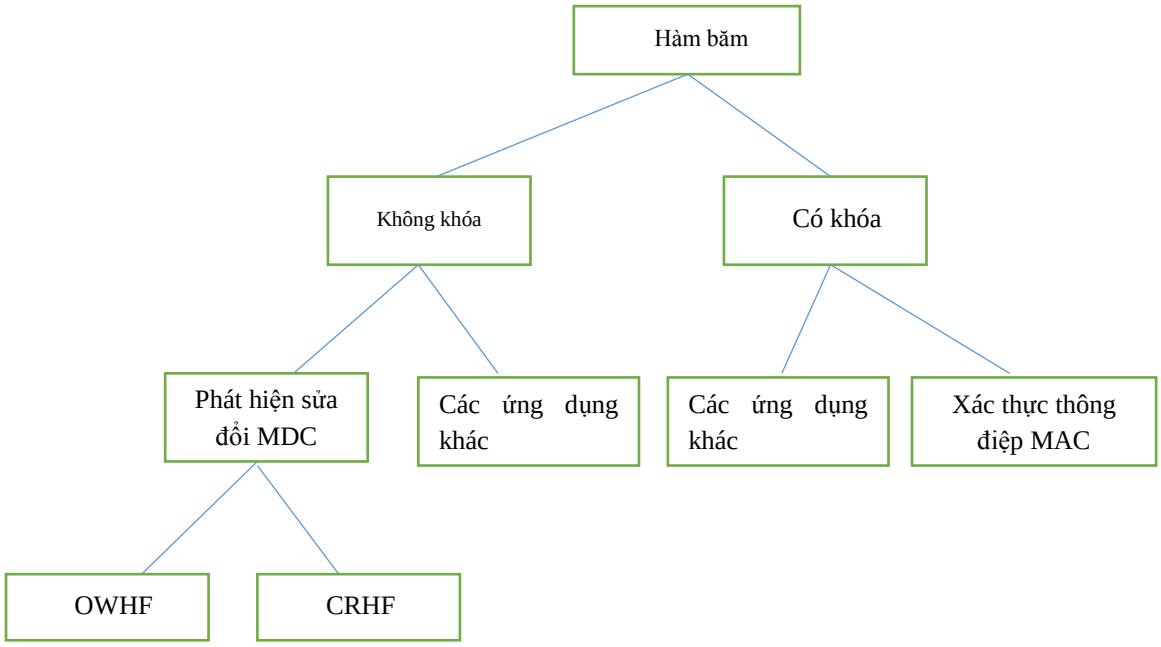
Giá trị băm là ảnh đại diện thu gọn, vân tay số (digital fingerprint), hoặc tóm lược thông báo (message digest) xâu đầu vào. Các hàm băm mật mã đóng vai trò quan trọng trong mật mã hiện đại được dùng để xác thực tính nguyên vẹn dữ liệu và dùng trong quá trình tạo chữ kí số trong giao dịch điện tử. Một lớp các hàm băm riêng được gọi là mã xác thực thông báo (MAC) cho phép xác thực thông báo bằng các kĩ thuật mã đối xứng.

Tính chất cơ bản của hàm băm mật mã

- Giá trị băm của bất kỳ thông điệp nào có thể được tính toán một cách dễ dàng.

- Khó suy ra thông điệp gốc của giá trị băm. Với thông điệp x thì dễ dàng tính được z = h(x), nhưng lại không thể suy ngược lại được x nếu chỉ có giá trị hàm băm h.
- Với thông điệp đầu vào x thu được bản băm z = H(x) là duy nhất.
- Không thể thay đổi một thông điệp nếu không thay đổi giá trị băm. Nếu dữ liệu trong thông điệp x thay đổi hay bị xóa để thành thông điệp x′ thì h(x’) ≠ h(x).
- Không tồn tại hai thông điệp khác nhau có giá trị băm như nhau (tính chất không xung đột).

Phân loại hàm băm mật mã



Hình 2.3: Sơ đồ phân loại hàm băm

Hàm băm mật mã có khóa

Hàm băm mật mã có khóa là hàm băm có dữ liệu đầu vào và kèm thông điệp khác là một khóa bí mật. Các hàm băm có khóa được sử dụng để xác thực thông điệp và thường được gọi là các thuật toán tạo mã xác thực thông báo (MAC).

“Hàm băm có khóa được trình bày trong phần mã xác thực HMAC phần 3.3.4”

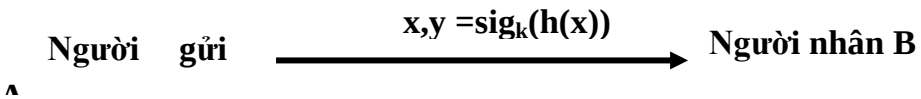
Hàm băm mật mã không khóa (có hàm băm dựa trên mật mã khối)

Hàm băm không khóa là hàm băm có dữ liệu đầu vào chỉ là thông điệp, không chứa khóa. Hàm băm không khóa có một số tính chất như sau: Việc đưa hàm băm vào dùng trong sơ đồ chữ ký số không làm giảm sự an toàn của sơ đồ chữ ký số vì nó là bản tóm lược thông điệp/thông báo – bản đại diện cho thông điệp – được ký chứ không phải là thông điệp gốc. Ngoài ra, hàm băm có khả năng chống cự trước các tấn công mật mã, tức là phải có các tính chất kháng cự sau:

Hàm băm kháng xung đột (được gọi là kháng không xung đột mạnh): Nghĩa là khó tìm được hai thông điệp (x, x') sao cho H(x) = H(x')

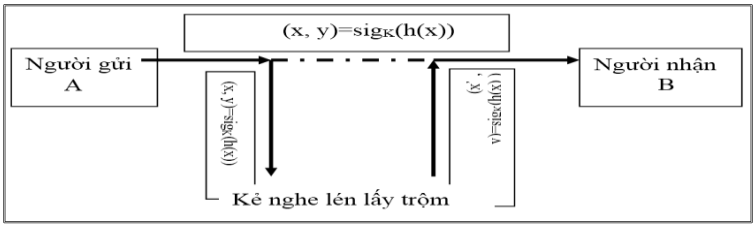
Ví dụ: Ta xét một kiểu tấn công sau:

- *Đáng lẽ*: Thông tin phải được truyền đúng từ A đến B



Hinh(a): Đường đi đúng của thông tin

- *Nhưng*: Trên đường truyền, thông tin bị lấy trộm và bị thay đổi.



Hình (b): Thông tin bị lấy trộm và bi thay đổi trên đường truyền

Hình 2.4: Thông tin trên đường truyền

Người A gửi cho B (x, y) với y = $sig_K(h(x))$ Nhưng trên đường truyền, thông tin bị lấy trộm. Tên trộm bằng cách nào đó tìm được một bản thông điệp x′ sao cho h(x') = h(x) mà x′ ≠ x. Sau đó, hắn đưa x' thay thế x rồi truyền tiếp cho B. Người B nhận được và vẫn xác thực thông tin đúng đắn.

*Hàm băm kháng tiền ảnh v*ới một mã băm h bất kỳ, khó tìm được một thông điệp x nào sao cho h = hash(x). Hàm băm được xem là hàm một chiều khi cho trước giá trị băm, không thể tái tạo lại thông điệp ban đầu, hay còn gọi là "tiền ảnh". Như vậy, trong trường hợp lý tưởng, cần phải thực hiện hàm băm 2^n thông điệp để tìm ra được một phương pháp tấn công cho phép xác định được "tiền ảnh" tương ứng với một giá trị băm cho trước thì thuật toán băm sẽ không còn an toàn nữa.

Kháng tiền ảnh thứ hai (tính chống xung đột yếu kháng) i: với một thông điệp x bất kỳ, khó tìm được một thông điệp x' sao cho x'≠ x. Nếu hàm băm h cho thông điệp đầu vào x (x' ≠ x) tạo ra giá trị băm h(x), thì rất khó để tìm ra giá trị đầu vào khác x' sao cho h(x') = h(x).

Chống va chạm (hay còn được gọi là hàm băm không va chạm): Khó tính toán để tìm được hai thông điệp đầu vào khác nhau (x ≠ x′) sao cho chúng có cùng giá trị băm. Thuộc tính này làm cho kẻ tấn công rất khó tìm thấy hai giá trị đầu vào khác nhau mà cùng giá trị băm. Ngoài ra, nếu một hàm băm là chống va chạm thì nó là hình ảnh thứ hai chống lại hình ảnh trước. Đây chính là nghịch lý ngày sinh.

Giả sử trong phòng có 30 sinh viên. Vậy xác suất để có hai SV có cùng ngày sinh là bao nhiêu? (1 năm 365 ngày khác nhau)

Theo nguyên lý chuồng bồ câu Dirichlet: cần có 365 +1 =366 người thì có thể tìm thấy hai người có cùng ngày sinh với xác suất 100%. Vì vậy với 30 người thì xác xuất này là rất nhỏ. Điều này muốn nói lên rằng, trong nhiều trường hợp xác suất để hai mẫu tin có cùng ảnh băm là không nhỏ.

Ý nghĩa của việc dùng thông điệp và hàm băm

Hàm băm trợ giúp cho các sơ đồ chữ ký nhằm giảm dung lượng của dữ liệu cần thiết khi truyền qua mạng của bức ký số được ký trên bản đại diện của thông điệp gốc.

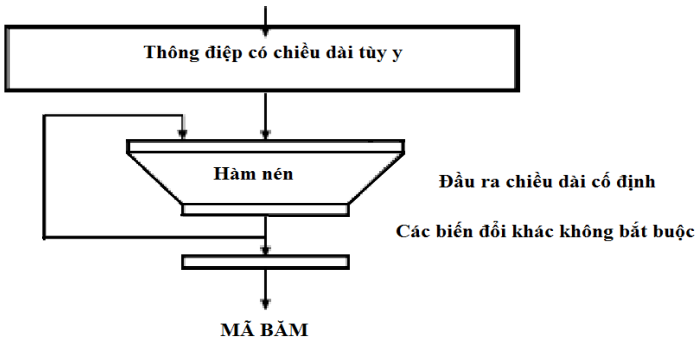
Hàm băm thường kết hợp với chữ ký số để tạo một loại chữ ký điện tử an toàn hơn và dùng để kiểm tra tính toàn vẹn của thông điệp.

Hàm băm không khóa bao gồm các lớp MDC (Modification Detection Code). Các MDC được sử dụng để tạo ra ảnh đặc trưng của thông điệp, đảm bảo sự toàn vẹn của dữ liệu. Bản thân MDC lại được chia thành hai lớp hàm sau:

- Hàm băm một chiều (OWHF – One way hash function) có nghĩa là với một mã băm biết trước, khó có thể tính toán để tìm ra thông điệp đầu vào có mã băm bằng với mã băm đã cho. Hàm băm một chiều thỏa mãn tính chất:
  - Khó tìm nghịch ảnh.
  - Khó tìm nghịch ảnh thứ hai.
- Hàm băm khó va chạm (CRHF _ Collision Resistant Hash Function) có nghĩa là khó có thể tính toán để tìm ra hai thông điệp khác nhau mà có cùng giá trị băm. Hàm băm khó va chạm ngoài hai tính chất cơ bản còn thỏa mãn các tính chất sau:
  - Khó tìm nghịch ảnh thứ hai
  - Khó va chạm.

2.4.2 Cấu trúc của hàm băm mật mã.

Thành phần chính của một hàm băm là một hàm nén và các hàm biến đổi khác. Hàm nén được thực thi nhiều lần để băm thông điệp ban đầu thành một chuỗi có chiều dài cố định.



Hình 2.5: Cấu trúc tổng quát của hàm băm

Có rất nhiều thuật toán hàm băm cho đến nay sử dụng chung một cấu trúc cơ bản cụ thể gồm các bước như sau:

- *Tiền xử lý*
  - Bộ đệm tin nhắn/thông điệp.
  - Phân tích thông điệp đệm thành các khối m bits.
  - Thiết lập giá trị khởi tạo bởi hàm băm.
- *Thuật toán băm bao gồm*:
  - Tạo ra các trạng thái từ thông điệp đệm (cùng với hàm băm, hằng số, chiều dài cố định)
  - Kết quả giá trị băm dùng để xác định bản tóm tắt của thông điệp.

Bước 1: Phân chia thông điệp đầu vào có chiều dài hữu hạn thành các khối thông điệp con liên tiếp có chiều dài cố định r (giả sử $m_1,m_2,m_{3,..,}m_{k)}$.

Bước 2: Do m có độ dài bất kỳ nên luôn có một bước thêm bits phụ sao cho chiều dài chuỗi mới m chia hết cho r (trong các bits thêm thường thêm 64 bits để lưu lại chiều dài ban đầu của chuỗi trước khi chèn).

Bước 3: Đưa khối thông điệp con $m_1,m_2,m_{3,..,}m_k$ sẽ lần lượt đi qua một hàm nén f của hàm băm b(m).

Bước 4: Kết quả của khối thứ m_{i-1} sau khi đi qua hàm nén f sẽ là nguồn dữ liệu vào cho bước thứ i tiếp theo.

Thiết kế thuật toán Hash

Một hàm băm là một hàm toán học gồm hai khối kích thước cố định của dữ liệu để tạo ra một mã băm. Mã băm này tạo thành một phần của thuật toán băm.Thành phần đầu tiên là hàm nén nhận đầu vào là một chuỗi có chiều dài bất kì và giá trị chaining variable và cho đầu ra là chuỗi có chiều dài cố định. Thành phần thứ hai là hàm chuẩn chuỗi đầu vào, hàm này có nhiệm vụ biến chuỗi đầu vào có chiều dài bất kì thành chuỗi các bits, mà chuỗi này là có chiều dài là bội số của các khối message block (có chiều dài là 512 bit hoặc 1024 bits). Ở thời điểm bắt đầu giá trị khởi tạo và giá trị cuối cùng của các chaining variable chính là giá trị của hàm băm. Hình minh họa hàm băm.

Hình 2.6: Mô hình các khối dữ liệu sử dụng hàm băm

Thuật toán chung:

Given: compression function C: $\{0,1\}^n \times \{0,1\}^m \rightarrow \{0,1\}^n$;

n – bit constant IV

Input: message M

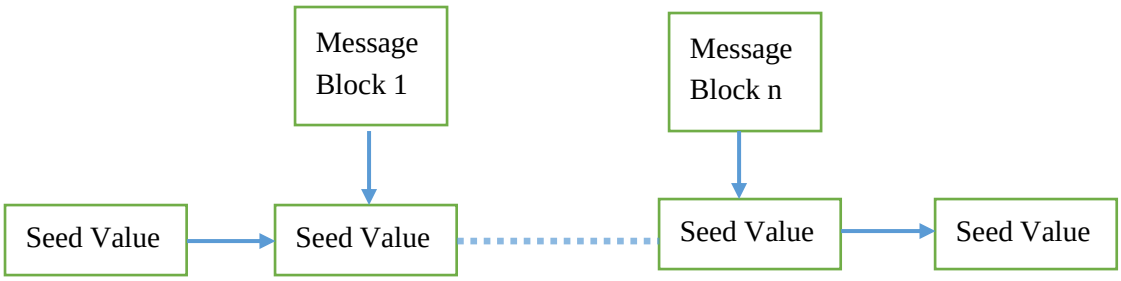
1. Break M into m –bit block $M_1,\ldots,M_k$, padding if necessary;
2. Let M_{k+1} be encoding of |M|;

3.Let h_0 = IV;

4. for I =1 to k+1 let $h_i = C(h_{i-1}, M_i)$;

5. Output h_{k+1}.

Thuật toán hash bao gồm các vòng hàm băm ở trên như một mật mã khối. Mỗi vòng lấy một đầu vào có kích thước cố định, điển hình là một sự kết hợp các khối tin nhắn mới nhất và đầu ra là vòng cuối cùng. Quá trình này được lặp lại bao nhiêu vòng như vậy sẽ băm hết toàn bộ tin nhắn. Sơ đồ của thuật toán băm được miêu tả trong minh họa sau đây.



Hình 2.7: Mô hình thuật toán băm

2.4.3 Hàm băm SHA (secure hash algorithm)

SHA hay thuật toán băm bảo mật là một họ những thuật toán băm mật mã do viện tiêu chuẩn và công nghệ Quốc gia (NIST) công bố thuộc tiêu chuẩn xử lý thông tin Liên Bang Hoa Kỳ (FIPS)[6,8,9]. Hiện tại có ba thuật toán SHA1, SHA2, SHA3 được định nghĩa.

Dưới đây các thuật toán băm SHA

- **SHA 1**
- **SHA 2**
  - + SHA -224
  - + SHA -256
  - + SHA -384
  - + SHA -512
- **SHA3**
  - + SHA3 - 224
  - + SHA3 – 256
  - + SHA3 – 384
  - + SHA3 - 512

2.4.3.1. SHA1 & SHA2

Đối với SHA 1 và SHA – 256, thông điệp mở rộng được phân tích thành N khối 512 bits $M^{(1)}$, $M^{(2)}$, …, $M^{(N)}$. Do đó 512 bits của khối dữ liệu đầu vào có thể được thể hiện bằng 16 từ 32 – bits, $M_0^{(i)}$ chứa 32 bits đầu của khối thông điệp i, $M_1^{(i)}$ chứa 32 bits kế tiếp…

Đối với SHA 384, SHA – 512 thông điệp mở rộng được phân tích thành N khối 1024 bits $M^{(1)}$, $M^{(2)}$,.., $M^{(N)}$. Do đó 1024 bits của khối dữ liệu ban đầu vào có thể được thể hiện bằng 16 từ 64 bits, $M_0^{(i)}$ chứa 64 bit đầu của khối thông điệp i, $M_1^{(i)}$ chứa 64 bits kế tiếp… $M_{16}^{(i)}$ chứa 64 bits cuối cùng.

Trước khi thực hiện băm, với mỗi thuật toán băm an toàn, giá trị băm ban đầu $H^{(0)}$ phải được thiết lập. Kích thước và số lượng từ trong $H^{(0)}$ tùy thuộc vào kích thước thông điệp rút gọn.

Các cặp thuật toán SHA – 224 và SHA – 256; SHA – 384 và SHA – 512 có các thao tác thực hiện giống nhau, chỉ khác nhau về số lượng bits kết quả của thông điệp rút gọn. Nói cách khác, SHA -224 sử dụng 224 bits đầu tiên trong kết quả thông điệp rút gọn sau khi áp dụng thuật toán SHA – 256. Tương tự SHA – 384 và SHA – 512 sử dụng 384 bits/512 bits đầu tiên trong kết quả thông điệp rút gọn.

Khung thuật toán chung của các thuật toán SHA

Trong các hàm băm SHA, chúng ta cần sử dụng thao tác quay phải một từ, ký hiệu là ROTR, và thao tác dịch phải một từ, ký hiệu SHR.

Khung thuật toán chung cho các hàm băm SHA

For i = 1 to N

For t = 0 to 15

Wt = Mt(i)

End for

For t = 16 to scheduleRound

$W_t = \sigma_1(W_t\text{-}2) + W_t\text{-}7 + \sigma(W_t\text{-}15) + W_t\text{-}16$

End for

$a = H_0(i\text{-}1)$

$b = H_1(i\text{-}1)$

$c = H_2(i\text{-}1)$

$d = H_3(i\text{-}1)$

$e = H_4^{(i\text{-}1)}$

$f = H_5^{(i\text{-}1)}$

$g = H_6^{(i\text{-}1)}$

$h = H_7^{(i\text{-}1)}$

for t = 0 to 63

$T_1 = h + \sum_1(e) + Ch(e,f,g) + K_t + W_t$

$T_2 = \sum_{10}(a) + Maj(a,b,c)$

h = g

g = f

f = e

e = d + T_1

d = c

c = b

b = a

a = $T_1 + T_2$

end for

$H_0^{(i)} = a + H_0^{(i-1)}$

$H_1^{(i)} = b + H_1^{(i-1)}$

$H_2^{(i)} = c + H_2^{(i-1)}$

$H_3^{(i)} = d + H_3^{(i-1)}$

$H_4^{(i)} = e + H_4^{(i-1)}$

$H_5^{(i)} = f + H_5^{(i-1)}$

$H_6^{(i)} = g + H_6^{(i-1)}$

$H_7^{(i)} = h + H_7^{(i-1)}$

End for

Mỗi thuật toán có bảng hằng số phân bố thông điệp tương ứng. Kích thước bảng hằng số thông điệp của SHA – 224 và SHA -256 là 64 bits, kích thước bảng hằng số thông điệp của SHA- 384 và SHA – 512 là 80 bits.

Các hàm được sử dụng

- ***SHA1***

$$f_t(x,y,z) = \begin{cases} \text{ch}(x,y,z) = (x \cap y) \oplus (x \cap z) & 0 \le t \le 19 \\ \text{Parity}(x,y,z) = x \oplus y \oplus z & 20 \le t \le 39 \\ \text{Maj}(x,y,z) = (x \cap y) \oplus (x \cap z) \oplus (y \cap z) & 40 \le t \le 59 \\ \text{Parity}(x,y,z) = x \oplus y \oplus z & 60 \le t \le 79 \end{cases}$$

- ***SHA -224*** và ***SHA -256***, sử dụng các hàm sau:

$Ch(x,y,z) = (x \cap y) \otimes (\neg x \cap z)$

$Maj(x,y,z) = (x \cap y) \otimes (x \cap z) \otimes (y \cap z)$

$\sum_0(x) = ROTR^2(x) \otimes ROTR^{13}(x) \otimes ROTR^{22}(x)$

$\sum_1(x) = ROTR^6(x) \otimes ROTR^{11}(x) \otimes ROTR^{25}(x)$

$\delta_0(x) = ROTR^7(x) \otimes ROTR^{18}(x) \otimes SHR^3(x)$

$\delta_1(x) = ROTR^{17}(x) \otimes ROTR^{19}(x) \otimes SHR^{10}(x)$

- ***SHA -384*** và ***SHA -512***, chúng ta sử dụng các hàm sau:

$Ch(x,y,z) = (x \cap y) \otimes (\neg x \cap z)$

$Maj(x,y,z) = (x \cap y) \otimes (x \cap z) \otimes (y \cap z)$

$\sum_0(x) = ROTR^{28}(x) \otimes ROTR^{34}(x) \otimes ROTR^{29}(x)$

$\sum_1(x) = ROTR^{14}(x) \otimes ROTR^{18}(x) \otimes ROTR^{41}(x)$

$$\delta_0\,(x) = ROTR^1\,(x) \otimes ROTR^8\,(x) \otimes SHR^7(x)$$
$$\delta_1\,(x) = ROTR^{19}\,(x) \otimes ROTR^{61}\,(x) \otimes SHR^6(x)$$

Sự khác biệt chính của các thuật toán là số lượng bits bảo mật của dữ liệu được băm điều này ảnh hưởng trực tiếp đến chiều dài của thông điệp rút gọn. Khi một thuật toán băm được sử dụng kết hợp với thuật toán khác đòi hỏi phải cho kết quả số lượng bits tương ứng.

Ví dụ, nếu một thông điệp được ký với thuật toán chữ ký điện tử cung cấp 128 bits thì thuật toán chữ ký đó có thể đòi hỏi sử dụng một thuật toán băm an toàn cung cấp 128 bits như SHA – 256.

Bảng 2.1: Bảng so sánh giữa hàm băm SHA1 và các họ hàm băm SHA 2

Thuật toán	Kích thước (bit)				Độ an toàn (đơn vị: bits)
	Thông điệp	Khối	Từ	Thông điệp rút gọn	
SHA 1	$< 2^{64}$	512	32	160	80
SHA -224	$< 2^{64}$	512	32	224	112
SHA – 256	$< 2^{64}$	512	32	256	128
SHA – 384	$< 2^{128}$	1024	64	384	192
SHA -512	$< 2^{128}$	1024	64	512	256

2.4.3.2. Hàm băm SHA3

Trong tháng 11 năm 2007 Viện Tiêu Chuẩn và Công nghệ Quốc gia Mỹ (NIST) đã mở một cuộc thi để phát triển thuật toán hàm "băm" mới thay cho SHA2. Các thuật toán băm mới sẽ được gọi là Secure Hash Alorithm – 3 (SHA3) [10,11,12,13]. Có 56 trong đó 64 mẫu thiết kế đã tham gia cuộc thi SHA3, 51 mẫu đệ trình đã lọt qua vòng 1 và vào ngày 01 tháng 11 năm 2008, 14 mẫu đã lọt vào vòng 2. Chung kết thiết kế SHA -3 đã được công bố vào ngày 09 tháng 12 năm 2010. Các thuật toán cuối cùng được coi như một ứng cử viên thay thế cho SHA -3 lad BLAKE, Grostl, JH, Keccak và Skein. Các tiêu chí lựa chọn bao gồm việc thực thi trong cả phần mềm và phần cứng, dung lượng thực hiện phần cứng, phản ứng với những nguy cơ tấn công tốt nhất và đủ khác biệt với các ứng viên khác.

Trong tháng 10 năm 2012, Viện Tiêu Chuẩn và công nghệ (NIST) đã chọn các thuật toán Keccack như là tiêu chuẩn mới SHA – 3. Hàm băm được thiết kế bởi Guido Bertoni, Joan Daemen, Michael Peeters và Gilles van Assche. Các hoán vị cơ bản Keccak tạo điều kiện cho việc mở rộng các chức năng mã hóa hoán vị dựa trên hoán vị bổ sung.

Thuật toán SHA -3 bao gồm:

- Bốn dạng hàm băm mật mã là: SHA3-224, SHA3–256, SHA3–384, SHA3–512.

- Hai dạng hàm băm mở rộng là: SHAKE-128, SHAKE- 256.

1. Trạng thái Keccak

Trong phần này, các hoán vị Keccak – p được xác định với hai tham số:

- Độ dài cố định của chuỗi hoán vị được gọi là chiều rộng của hoán vị
- Số lần lặp lại của một chuyển đổi được gọi là một vòng.

SHA3 là tổ hợp các hàm sponge được đặc trưng bởi hai tham số, tốc độ r bits và cường độ an toàn c. Tổng, r+c xác định độ rộng của hàm băm SHA3. Phép hoán vị được sử dụng trong việc xây dựng Sponge và giới hạn giá tị cực đại là 1600.

Chiều rộng được biểu thị bởi b và số vòng được biểu thị bởi n_r. Các Keccak – p hoán vị với số vòng là n_r và chiều rộng b được ký hiệu Keccak – p[b n_r].

Mỗi hàm nén Keccak là duy nhất bao gồm 24 dạng viên đạn và mỗi vòng được chia thành năm bước là: $\boldsymbol{\theta(A)}$**, Rho (ρ) và Pi (π), Chi(X), Iota(i) (***sẽ tương ứng với 5 thuật toán sẽ trình bày bên dưới*)

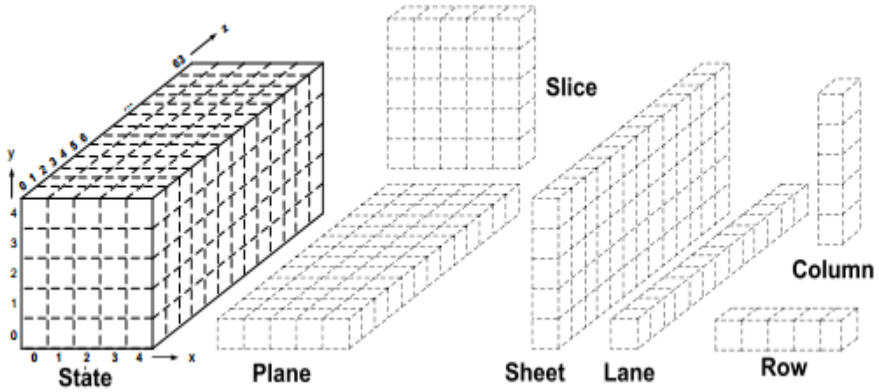
Keccak–p, ký hiệu bởi Keccak – f[b], trong $b \in \{25, 50, 100, 200, 400, 800, 1600\}$ là miền của phép hoán vị. Miền của phép hoán vị cũng là miền của trạng thái trong việc xây dựng sponge. Đặc tả này chứa hai số liên quan đến b: *b/25* và $log_2(b/25)$, kí hiệu bởi w và l.

Bảng 1.2: Keccak – p hoán vị chiều rộng và các số liệu liên quan.

B	25	50	100	200	400	800	1600
W	1	2	4	8	16	32	64
l	0	1	2	3	4	5	6

a. Thành phần của mảng trạng thái hàm băm SHA3

Trạng thái: mảng 5 x 5 x 2^l bits vào trong đó *l* dao động từ 0 đến 6. Trong đó 2^l bits là (1, 2, 4, 8, 16, 32 hoặc 64). (b = 25.2^l)



Hình 2.8 Qui ước đặt tên cho các trạng thái của keccak –p

b. Chuyển dạng chuỗi thành dạng mảng các trạng thái

Cho S biểu thị một chuỗi b- bits trạng thái đại diện các trạng thái cho

Keccak-p[b, n_r] phép hoán vị. Mảng trạng thái tương ứng biểu thị là ma trận A được định nghĩa như sau:

Với bộ (x,y,z) sao cho $0 \leq x < 5, 0 \leq y < 5, và\ 0 \leq z < w$; có 24 vòng nén (core) của SHA 3 và mỗi vòng gồm 5 bước Theta (θ), Rho (ρ), Pi (π), Chi (χ) và Iota (i)

$A[0,0,0] = s[0]$	A[1,0,0] = S[64]		A[4,0,0]= S[256]
A[0,0,1] = S[1]	A[1,0,1] = S[65]		A[4,0,1] = S[257]
A[0,0,2] = S[2]	A[1,0,2] = S[66]		A[4,0,2] = S[258]
⋮	⋮	…	⋮
A[0,0,61] = S[61]	A[1,0,61] = S[125]		A[4,0,61] = [317]
A[0,0,62] = S[62]	A[0,0,62] = S[126]		A[4,0,62]= S[318]
A[0,0,63] = S[63]	A[0,0,63] = S[127]		A[4,0,63]= S[319]

Và

A[0,1,0] = S[320]	A[1,1,0] = S[384]		A[4,1,0] = S[576]
A[0,1,1] = S[321]	A[1,1,1] = S[385]		A[4,1,1] = S[577]
A[0,1,2] = S[322]	A[1,1,2] = S[386]		A[4,1,2] = S[578]
⋮	⋮	…	⋮
A[0,1,61] = S[381]	A[1,1,61] = S[445]		A[4,1,61]= [637]
A[0,1,62] = S[382]	A[1,1,62] = S[446]		A[4,1,62]= S[638]
A[0,1,63] = S[383]	A[1,1,63] = S[447]		A[4,1,63]= S[639]

Và

A[0,2,0] = S[640]	A[1,2,0] = S[704]		A[4,2,0]= S[896]
A[0,2,1] = S[641]	A[1,2,1] = S[705]		A[4,2,1]= S[897]
A[0,2,63] = S[703]	A[1,2,63] = S[767]		A[4,2,63]= S[959]

c. Chuyển mảng trạng thái thành dạng chuỗi

A là mảng trạng thái, S biểu diễn chuỗi tương ứng được xây dựng từ *lanes và planes của mảng* A. Đối với mỗi cặp số nguyên (i, j) sao cho $0 \leq i < 5\ và\ 0 \leq j < 5$, xác định chuỗi Lane(i,j) như sau:

Lane(i,j) = $A[i, j, 0] \| A[i, j, 1] \| A[i, j, 2] \| \ldots \| A[i, j, w-2] \| A[i, j, w-1]$

Ví dụ: b = 1600 với w = 64

Lane (0,0) = A[0,0,0] || A[0,0,1] || A [0,0,2] || … || A[0,0,62] || A[0,0,63]

Lane (0,1) = A[1,0,0] || A[1,0,1] || A[1,0,2] || … || A[1,0,63] || A[1,0,63]

Lane (2,0) = A[2,0,0] || A[2,0,1] || A[2,0,2] || … || A[2,0,62] || A[2,0,63]

Với số nguyên j sao cho $0 \leq j < 5$, xác định chuỗi Plane(j).

Plane(j)= Lane(0, j) || Lane(1, j) || Lane(2, j) || Lane(3, j) || Lane(4, j)

Thì

S = Plane(0) || Plane(1) || Plane(2) || Plane(3) || Plane(4).

Ví dụ: b = 1600 ,w=64

S = A[0,0,0] || A[0,0,1] || A[0,0,2] || … || A[0,0,62] || A[0,0,63]
|| A[1,0,0] || A[1,0,1] || A[1,0,2] || … || A[1,0,62] || A[1,0,63]
|| A[2,0,0] || A[2,0,1] || A[2,0,2] || … || A[2,0,62] || A[2,0,63]
|| A[3,0,0] || A[3,0,1] || A[3,0,2] || … || A[3,0,62] || A[3,0,63]
|| A[4,0,0] || A[4,0,1] || A[4,0,2] || … || A[4,0,62] || A[4,0,63]

|| A[0,1,0] || A[0,1,1] || A[0,1,2] || … || A[0,1,62] || A[0,1,63]
|| A[1,1,0] || A[1,1,1] || A[1,1,2] || … || A[1,1,62] || A[1,1,63]
|| A[2,1,0] || A[2,1,1] || A[2,1,2] || … || A[2,1,62] || A[2,1,63]
|| A[3,1,0] || A[3,1,1] || A[3,1,2] || … || A[3,1,62] || A[3,1,63]
|| A[4,1,0] || A[4,1,1] || A[4,1,2] || … || A[4,1,62] || A[4,1,63]
⋮
|| A[0,4,0] || A[0,4,1] || A[0,4,2] || … || A[0,4,62] || A[0,4,63]
|| A[1,4,0] || A[1,4,1] || A[1,4,2] || … || A[1,4,62] || A[1,4,63]
|| A[2,4,0] || A[2,4,1] || A[2,4,2] || … || A[2,4,62] || A[2,4,63]
|| A[3,4,0] || A[3,4,1] || A[3,4,2] || … || A[3,4,62] || A[3,4,63]
|| A[4,4,0] || A[4,4,1] || A[4,4,2] || … || A[4,4,62] || A[4,4,63]

2. Đặc tả thuật toán chuyển trạng thái của Keccak –p[b,n_r]

a) Đặc tả thuật toán theta $\theta(A)$

Thuật toán 1: $\theta(A)$

Inphut:

State mảng A

Output:

State mảng **A′**

Các bước:

1. Đối với tất cả các cặp (x,z) sao cho $0 \leq x < 5$ và $0 \leq z < w$

$C[x,z] = A[x,0,z] \otimes A[x,1,z] \otimes A[x,2,z] \otimes A[x,3,z] \otimes A[x,4,z]$.

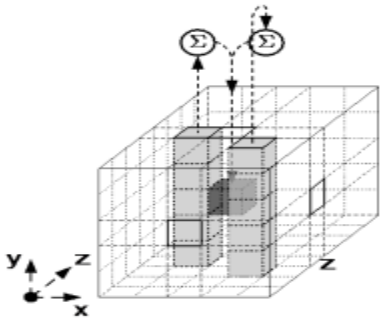
2. Đối với tất cả các cặp (x,z) sao cho $0 \leq x < 5$ và $0 \leq z < w$

$D[x,z] = C[(x-1) mod\ 5, z] \otimes C[(x+1) mod\ 5, (z-1) mod\ w]$.

3. Đối với tọa độ (x,y,z) sao cho $0 \leq x < 5, 0 \leq y < 5, 0 \leq z < w$

$A'[x,y,z] = a[x,y,z] \otimes D[x,z]$.

Hiệu quả của θ là để XOR các bits trạng thái chẵn lẽ với nhau giữa hai cột trong mảng. Đặc biệt với ma trận A[x_0,y_0,z_0] tọa độ của x là (x_0 -1) mod 5, với cùng tọa độ z, trong khi tọa độ x của một cột khác là (x_0 +1) mod 5, với z – (z_0 – 1) mod w. Hình minh họa bên dưới của θ , ∑ kí hiêu tổng, XOR tổng của tất cả các bít trong cột



Hình 2.9: Minh họa của θ áp dụng cho một bits đơn.

b) Đặc tả thuật toán 2 Rho ρ(A)

Input:

State mảng A

Output:

State mảng A′.

Các bước:

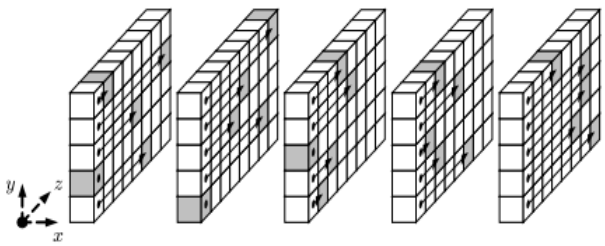
1. Với ∀ Z sao cho $0 \leq z < w, A'[0,0,z] = A[0,0,z]$
2. (x,y) = (1,0).
3. Cho t chạy từ {0,..,23}
   a. Với ∀ z sao cho $0 \leq z < w$, $A'[x,y,z] = A[x,y,\left(z - \frac{(t+1)(t+2)}{2}\right) mod\ w.$
   b. (x,y) = (y,(2x+3y) mod 5)
4. Quay lại A′.

Hiệu ứng của **ρ** để xoay các bit của mỗi *Lane* theo chiều dài là một khoảng chênh "called the offset" ngoài ra nó còn phụ thuộc tọa độ cố định x và y của *lane*. Đối với mỗi bít trong *lane* tọa độ z được sửa đổi bằng khoảng chênh "offset", *lane* chia thành các modulo.

Bảng 2.1: offset của ρ

	x =3	x=4	x =0	x=1	x =2
y=2	153	231	3	10	171
y=1	55	276	36	300	6
y=0	28	91	0	1	190
y=4	120	78	210	66	253
y=3	21	136	105	45	15

Một minh họa của ρ cho trường hợp w=8 (hình dưới). Qui ước dán nhãn cho tọa độ x và y trong hình 3.8, tương ứng với các hàng và cột trong bảng 3. Ví dụ: Lane A[0,0] đươc mô tả giữa bảng và lane A[2,3] được mô tả ở dưới cùng hầu hết bên phải bảng.



Hình 2.10: Hình minh họa của ρ với b = 200

Đối với mỗi lane trong hình 2.11, dấu chấm đen chỉ ra các tọa độ z bits là 0 và khối hình mờ chỉ ra vị trí của bits sau khi thực hiện ρ. Các bits khác của *lane* bằng khoảng chênh "offset". Ví dụ, khoảng chênh cho *lane* A[1,0] là 1, Nên bits cuối cùng của z là 7 khi dịch chuyển sang vị trí phía trước mà phối hợp với z là 0. Do đó, các "offset" có thể được giảm kích thước xuống là module.

c) Đặc tả thuật toán pi (π)

Thuật toán 3: π (A)

Input:

State mảng A.

Output:

State mảng A′.

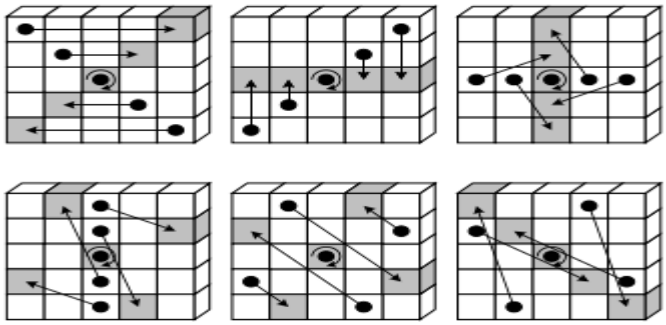
Các bước:

1. với ∀ (x,y,z) sao cho $0 \leq x < 5, v$à $0 \leq z < w$

$$A'[x, y, z] = [(x + 3y) mod\ 5, x, y].$$

2. Quay lại A′.

Hiệu ứng của π là sắp đặt lại vị trí của *lane*, được minh họa bằng lát cắt hình 2.8 Qui ước cho nhãn tọa độ minh họa ở *hình 2.10.* Ví dụ, các bits tọa độ x= y =0 được mô tả ở giữa *slice*.



Hình 2.11: Minh họa một lát cắt của π

d) Thuật toán 4 Chi(X);

Input:

State mảng A.

Output:

State mảng A′.

Các bước:

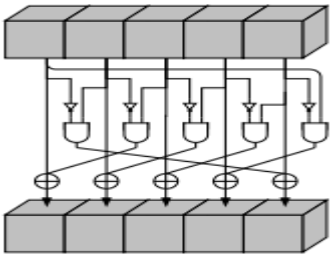
1. Với ∀ (x,y,z) soa cho $0 \leq x < 5, 0 \leq y < 5$ và $0 \leq z < w$

$$A'[x, y, z] = A[x, y, z] \otimes \big((A[(x + 1) mod\ 5, y, z] \otimes 1)\ .\ A[(x + 2) mod\ 5, y, z]\big).$$

2. Quay lại A′.

Ở bước 1 dấu chấm bên phải là phép nhân số nguyên, tương đương với biểu thức boolean "AND".

Hiệu ứng của X là để XOR từng bits với hàm phi tuyến tính "non –linear" của hai bits khác được minh họa trong hình 2.12 dưới đây.



Hình 2.12: Minh họa mô hình thuật toán Chi(X)

e) Thuật toán 5 (Iota): j(A,i_r)

Input

State mảng A

Round index i_r.

Output:

State mảng A′.

Các bước:

1. For all triples(x,y,z) sao cho $0 \le x < 5, 0 \le y < 5$ *v*à $0 \le z < w$, $A'[x,y,z] = A[x,y,z]$.

2. Let RC = 0^w.

3. For j from 0 to *l* RC[$2^j - 1$] = rc(j +7i_r).

4. For all z sao cho $0 \le z < w$, $A'[0,0,z] = A'[0,0,z] \otimes RC[z]$.

5. Return A′.

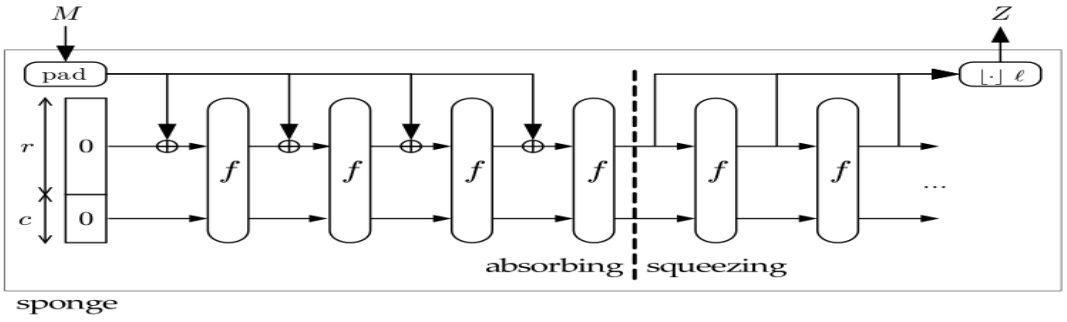
Hiệu ứng của thuật toán j là để sửa đổi một số bits của *lane (0,0)* theo chỉ mục tròn, 24 *lane* không bị ảnh hưởng bởi i.

3. Xây dựng Sponge.

Xây dựng sponge là một khuân khổ để xác định các hàm dạng nhị phân với độ dài đầu ra tùy ý. Việc xây dựng sử dụng ba thành phần sau:

- Hàm cơ bản về chuỗi có chiều dài cố định, kí hiệu là *f*.
- Một tham biến tốc độ, kí hiệu là *r*.
- Một quy tắc chêm/thêm, kí hiệu là *pad*.

Xây dựng sponge được minh họa trong hình 2.13.



Hình 2.13: Xây dựng sponge: Z=SPONGE[f,pad,r](M,d).

Hàm f ánh xạ một chuỗi có chiều dài cố định, kí hiệu b, đến chuỗi có chiều dài tương tự. Tỷ lệ r là số nguyên dương nhỏ hơn so với độ rộng b. Dung lượng kí hiệu là C, b- r số nguyên dương. Do đó r +c = b.

Qui tắc đệm "pad" nghĩa là chuỗi có độ dài thích hợp để gắn các chuỗi f. Cho x là số nguyên dương (*trong lý thuyết số)* và m số nguyên không âm (*trong lý thuyết tập hợp).* Đệm đầu ra pad(x,m) là chuỗi có các thuộc tính m+len(pad(x,m)) là một bội số của x. Trong việc xây dựng sponge x= r và m = len(N), sao cho chuỗi đệm đầu vào có thể được phân chia thành một chuỗi các r –bits.

Thuật toán: SPONGE[f,pad,r](N,d)

Input:

String N,

Nonnegative integer d.

Output:

String Z sao cho len(Z) = d.

Các bước:

1. Let $P = N \,||\, pad(r,len(N))$.
2. Let $n = len(P)/r$.
3. Let $c = b\text{-}r$.
4. Let P0,…, Pn-1 là chuỗi duy nhất có độ dài r sao cho P = P0 || … || Pn-1.
5. Let S=0b.
6. For i from 0 to n-1, let $S = f(S \otimes (P_i \,||\, 0^c))$.
7. Let Z chuỗi rỗng.
8. Let Z= Z || ztruncr(S).
9. if $d \leq |Z|$; return Truncd(Z); else continue.
10. Let S=f(S), tiếp tục với bước 8.

Đặc tả của Keccak[c]

Keccak là gia đình của hàm sponge với các hoán vị $\text{Keccak} - \text{p}[\text{b}, 12 + 2\text{l}]$ cộng với hàm cơ bản trong pad10*1 và quy tắc đệm. Keccak này được biểu diễn bằng bất kỳ tham số tốc độ r và dung lượng c sao cho r +c thuộc {25,50,100,200,400,800,1600}

Ví dụ b=1600

$$\text{KECCAK[c]} = \text{SPONGE[KECCAK} = \text{p[1600,24]}, \text{pad10} * 1,1600 - \text{c}].$$

Cho chuỗi bits đầu vào là N và chiều dài đầu ra là d:

$$\text{KECCAK[c]}(\text{N}, \text{d}) = \text{SPONGE[KECCAK} - \text{p[1600,24]}, \text{pad10} * 1,1600 - \text{c}](\text{N}, \text{d})$$

Bảng 3.2: So sánh giữa SHA1, SHA2 và SHA3

Function	Output Size	Security Strengths in Bits		
		Collision	Preimage	2nd Preimage
SHA1	160	<80	160	160 - L(M)
SHA-224	224	12	224	min(224,256 - L(M)
SHA-512/224	224	112	224	224
SHA-256	256	128	256	256-L(M)
SHA-512/256	256	128	256	256
SHA-384	384	192	384	384
SHA-512	512	256	512	512-L(M)
SHA3-224	224	112	224	224
SHA3-256	256	128	256	256
SHA3-384	384	192	384	384
SHA3-512	512	256	512	512
SHAKE128	d	min(d/2,128)	$\geq \min(d, 256)$	min(d/2,128)
SHAKE256	d	min(d/2,256)	$\geq \min(d, 128)$	min(d/2,256)

Bảng so sánh kích thước băm của SHA1, SH2, SHA3

	SHA1	**SHA2**				**SHA3**			
		224	**256**	**384**	**512**	**224**	**256**	**384**	**512**
Digest size	**160**	**224**	**256**	**384**	**512**	**224**	**256**	**384**	**512**
Message size	$2^{64} - 1$	$2^{64} - 1$	$2^{64} - 1$	$2^{128} - 1$	$2^{128} - 1$	∞	∞	∞	∞
Block size	**512**	**512**	**512**	**1024**	**1024**	**1152**	**1088**	**832**	**576**
Word size	**32**	**32**	**32**	**64**	**64**	**64**	**64**	**64**	**64**
No of rounds	**80**	**64**	**64**	**80**	**80**	**24**	**24**	**24**	**24**

Ví dụ băm của các hàm SHA

Hàm băm		Dữ liệu băm	Kết quả băm
SHA1		TRƯỜNG CÔNG NGHỆ	= ecb705fa6acba12ab59be790065631981ee63cfe
SHA2	**SHA2 – 224**		= 0238d556a16ad2bd1ebf8f1211477c099acbb4d4d6a34 feaeee0edc2
	SHA2 -256		=d27b8767bc2f8d69e9645c15c89bb91d193286c0a08841c1 1bdc4bbc239f5e51
	SHA2 – 384		=4152763b108e423f4ee9c8dc30e6f9b7c8c549016c0f1f9fac 43b9d5e6bf2259e5f2c2e3b4d9bc341f15f97f7b300d5a
	SHA2 – 512		=4697e3ed5c53a21334984072d559a6c38e3648038b82a96e 24281a7e11fade87a83c99b2eef58f0c80f460a8332bbb79aaa d2392519c507f0baee2144c4c5e42
SHA3	**SHA3 – 224**		=88-96-0D-03-20-7A-F9-31-CD-0C-7B-4E-C3-93-42-AE-74-6D-B7-31-17-99-81-B9-90-07-84-72
	SHA3 -256		=95-D0-AC-A4-27-6A-B2-76-EE-9E-88-8F-A2-3A-63-AA-46-D6-9E-86-40-43-B0-19-FF-4E-5C-70-57-E2-84-A1
	SHA3 - 384		=ED-BC-79-A7-C5-3C-CD-57-89-80-25-F2-9F-1D-84-38-D3-62-AF-8C-4C-CC-FF-08-CA-B8-FC-3E-36-89-1C-F3-D0-88-7E-B0-D7-75-B7-46-BC-00-EF-C7-51-C7-8B-52
	SHA3 - 512		=13-2E-73-5C-99-BD-A6-4A-1B-47-62-06-49-47-44-1B-D4-E4-A8-01-02-89-7C-A7-82-9B-A7-1B-8F-5A-D4-36-16-BE-B8-18-F1-A0-14-BA-83-36-5A-4B-1B-36-9A-7A-C4-82-DA-2C-FB-8C-78-1F-2F-33-28-5F-DB-5E-16-13

- Từ những năm 2005, SHA1 đã được chứng minh là khá yếu. Vào năm 2013, nhà nghiên cứu cứu Marc Stevens đã xuất bản một bài báo chứng minh về lý thuyết đã tìm ra một đụng độ của SHA-1. Tuy nhiên để tìm bằng chứng các đụng độ thì đến 2 năm sau các nhà nghiên cứu Google mới có thể tìm ra được phương pháp tạo các đụng độ. Các cuộc tấn công va chạm xuất hiện khi cùng một giá trị hàm băm được tạo ra cho hai thông điệp khác nhau. Điều này có thể được khai thác để giả mạo chữ ký điện tử, cho phép những kẻ tấn công phá vỡ các công cụ liên lạc được mã hóa bằng SHA 1.
- Thuật giải SHA 2 là kế nhiệm SHA1 có nhiều độ dài bits, phổ biến nhất là 256/512 bits. Tuy nhiên vì thuật toán SHA2 tương tự thuật toán SHA1 nên có thể có các cuộc tấn công bất ngờ, hiện tại SHA2 vẫn là hàm băm an toàn.

- Thuật toán SHA 3 bao gồm bốn hàm băm SHA3 -224, SHA3-256, SHA3-384, SHA3-512 và hai hàm mở rộng SHAKE128 và SHAKE256. SHA3 được thiết kế có hiệu quả cao về phần cứng nhưng tương đối chậm so với phần mềm. Và hàm băm SHA3 có thể được thực hiện trên chip mà không cần nhiều mạch bổ sung. Trong quá trình phân tích Keccak, các nhà thiết kế đã đưa vào một loạt các giả pháp cơ bản và sâu sắc. Trong đó kể đến là cấu trúc Sponege. Bên cạnh cấu trúc Sponge các tác giả còn thực hiện các biện pháp như bổ sung khóa mật vào đầu vào của keccak biến nó thành mã xác thực thông điệp, bổ sung khóa mật vào vector khởi đầu công khai và đưa vào chế độ gama có độ dài tùy ý biến cấu trúc Sponge thành mã dòng. Keccak coi tiêu chuẩn cần và đủ cho tính an toàn là sự không phân biệt được của hàm hash với Ramdom Oracle. Ramdom Oracle là một hàm lý tưởng mô tả công việc của máy với bộ nhớ vô hạn đáp ứng mọi yêu cầu đưa ra một số ngẫu nhiên lý tưởng. Bởi vậy, Keccak có nhiều ưu điểm vượt trội so với các hàm băm khác.

Thực tế, việc áp dụng rộng rãi thuật toán SHA3 sẽ thực hiện sau vài năm. Ưu tiên lớn hơn cho hầu hết các doanh nghiệp/cơ quan chuyển từ SHA1 sang SHA2. Theo quy định tại thông tư 06 của bộ TT&TT, thời gian tới, các nhà cung cấp dịch vụ chứng thực chữ ký số sẽ phải chuyển sang sử dụng hàm băm SHA2, thay cho hàm băm SHA1 đang sử dụng và hết hạn sử dụng sau ngày 1.1.2017.

2.5. CÁC PHƯƠNG PHÁP ĐẢM BẢO TÍNH TOÀN VẸN BẰNG MÃ XÁC THỰC

2.5.1 Xác thực thông điệp

Mã xác thực thông điệp là một đoạn mã cho phép xác định nguồn gốc của dữ liệu, thuyết phục với người dùng là dữ liệu này chưa bị sửa đổi hoặc giả mạo. Đây là là một cơ chế quan trọng để duy trì tính toàn vẹn và không thể từ chối dữ liệu.

2.5.2 Phân loại mã xác thực

Với các giao thức trực tuyến, mã xác thực thông báo mật mã (cryptographic Message Authentication Code –MAC) là rất quan trọng và có tính chất như bắt buộc để đảm bảo tính xác thực giữa các bên tham gia giao dịch[13]

Mục đích của hàm MAC là đảm bảo để hai (hay nhiều) bên tham gia giao dịch khi có chung khóa bí mật có thể giao dịch với nhau, kèm theo khả năng có thể phát hiện được thay đổi của thông báo trong quá trình vận chuyển, nhằm tránh các tấn công làm thay đổi thông báo.

Thuật toán MAC dựa trên thông báo đầu vào và khóa mật để tạo ra thẻ MAC (MAC tag). Thông điệp và MAC tag được gửi tới người nhận, người nhận tính lại giá trị MAC tag và so sánh nó với giá trị thẻ MAC nhận được. Nếu hai giá trị thẻ MAC trùng nhau thì coi như thông điệp chính xác, ngược lại thông điệp coi là đã bị thay đổi.

Đối với kẻ tấn công giả mạo thông điệp, nó phải phá được hàm MAC, việc này khó tương đương với việc phá khóa bảo vệ thông điệp. Trên thực tế, các giao thức thường chia thông điệp dài làm nhiều đoạn nhỏ và chúng được xác thực độc lập với nhau, dẫn đến phát sinh tấn công lặp (replay attack). Do vậy khi thiết kế giao thức có sử dụng MAC cần rất thận trọng.

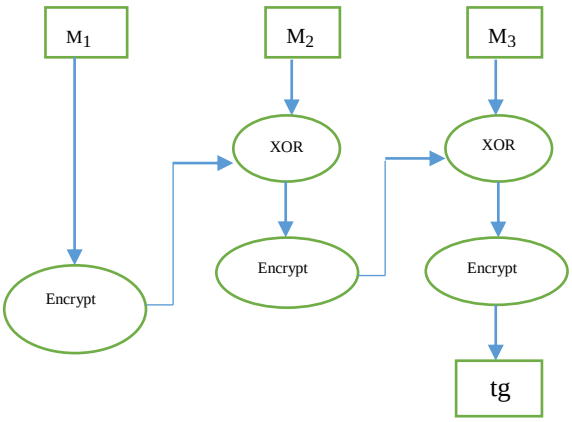
Để giúp các nhà phát triển ứng dụng tích hợp MAC vào sản phẩm khác nhau, viện tiêu chuẩn và công nghệ quốc gia của Mỹ (NIST) đưa ra hai chuẩn về hàm MAC. *Tiêu chuẩn thứ nhất là mã xác thực thông điệp sử dụng hàm một chiều có khóa HMAC* (Keyd-Hash Messasge Authentication Code). Chuẩn này mô tả phương pháp an toàn chuyển hàm một chiều kháng va chạm hash thành hàm MAC. *Chuẩn thứ hai NIST đưa ra là mã xác thực thông điệp mã hóa* (Cipher Message Authentication Code- CMAC). Không giống HMAC, CMAC sử dụng mã khối để thực hiện chức năng MAC, nó rất phù hợp với các ứng dụng bộ nhớ hạn chế chỉ đủ để dùng cho mã hóa dữ liệu.

Mục đích của hàm MAC khác với hàm một chiều hash, nó bảo đảm xác thực thông điệp chứ không phải toàn vẹn thông điệp, tuy nhiên cả hai dựa trên nguyên lý rất chung. Trong cả hai trường hợp chúng ta đều tìm cách xác định tính chính xác, cụ thể hơn là sự toàn vẹn của thông điệp. Tất nhiên mục tiêu của xác thực là tìm cách trả lại tính ban đầu của thông điệp.

Ví dụ nếu ta sử dụng SHA để tóm lược thông điệp để có 160 bits X và gửi cho người khác thì người đó chỉ có thể xác định được xem thông điệp đó là giữ nguyên như ban đầu hay không? Bằng cách tính giá trị tóm lược và so sánh chúng với nhau, nhưng cách này không thể xác định được ai tạo thông điệp đó. Giá trị tóm lược không cung cấp thông tin này. Giả sử trong giao dịch cả hai đều dùng chung khóa K, nếu ta gửi thông điệp và thẻ MAC với khóa K, thì người nhận có thể kiểm tra để xác định là thông điệp có gửi từ bạn hay không bằng cách kiểm tra thẻ MAC.

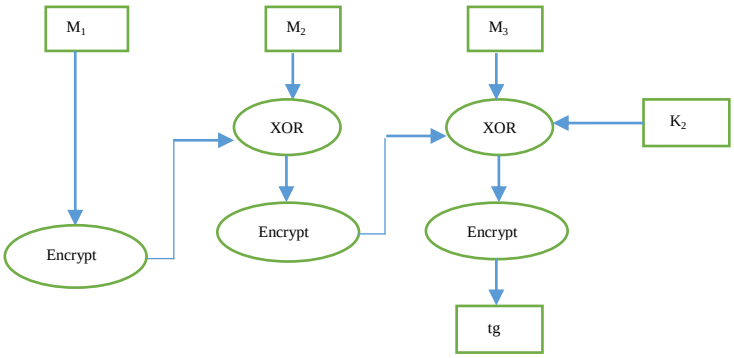
2.5.3 Mã xác thực thông điệp mã hóa (CMAC – CBC MAC)

CMAC được lấy từ dự thảo mã xác thực thông báo một khóa (One-Key Message Authentication Code – OMAC). Nó dựa trên mã xác thực thông báo chế độ mã CBC ba khóa (three-key cipher block chaining MAC). MAC mã hóa nguyên thủy của NIST là CBC – MAC. Trong đó người gửi chỉ cần chọn khóa độc lập với khóa mã và sử dụng mã theo chế độ CBC. Người gửi bỏ qua mọi bản mã giữa, chỉ quan tâm đến bản mã cuối cùng là thẻ MAC. Khóa dùng cho CBC – MAC không trùng với khóa mã của bản rõ, khi đó MAC là an toàn.

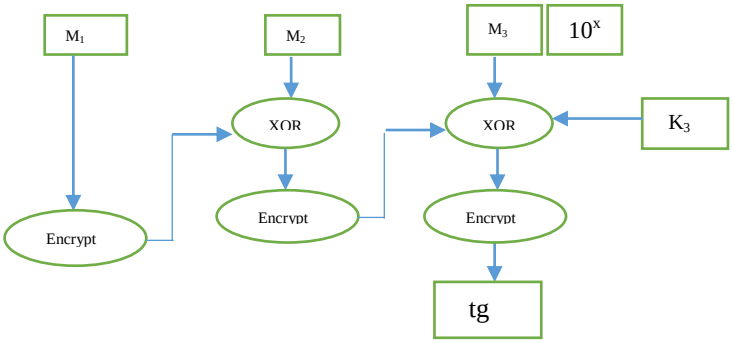


Hình 2.14: Sơ đồ CBC – MAC (nguyên thủy)

Như vậy nếu thông báo có thể chia thành các khối có độ dài bằng nhau (bằng độ dài của khối khóa) thì chỉ cần dùng một khóa. Khi thông báo gồm các khối có độ dài khác nhau (độ dài khối cuối nhỏ hơn độ dài các khối trước đó).



Hình 2.15: sơ đồ OMAC thông báo với các khối có cùng độ dài



Hình 2.16: Sơ đồ OMAC thông báo với các khối cuối ngắn hơn các khối trước

Để khắc phục, người ta đưa ra XCBC và sử dụng 3 khóa. Khóa đầu tiên sử dụng để mã hóa dữ liệu ở chế độ CBC –MAC. Hai khóa còn lại được áp dụng cho khối với phép hoặc loại trừ (XOR), bất kể khối cuối có cùng độ dài với các khối trước đó hay không? Đối với XCBC, ba khóa phải độc lập với nhau.

2.5.3.1 An toàn CMAC

Để thuận tiện, các khóa được tạo ra phụ thuộc vào nhau, Điều này phải trả giá là nếu kẻ tấn công biết được một khóa, nó có thể tìm ra khóa kia. Tấn công đối với hàm MAC là các tấn công trực tuyến. Vì tất cả dữ liệu cần xác thực, kẻ tấn công không dễ gì truy vấn tới thiết bị, nhưng nó có thể biết được dữ liệu nào đưa vào MAC. Trên thực tế kẻ tấn công có rất ít thời gian để thực hiện giả mạo mà không bị phát hiện, xác xuất thực hiện thành công sự giả mạo rất nhỏ. Ví dụ nếu sử dụng mã khối AES (độ dài khối n =128 bits), có thể sử dụng CMAC để xác thực 243 khối (1024 terabyte) mới thay đổi khóa mà vẫn đảm bảo xác xuất gói nhỏ hơn 2^{-40}. Trong trường hợp mã hóa, tấn công là offline (hoạt động độc lập, không trực tuyến). Kẻ tấn công có thể lặp lại các tính toán (thử giải mã với khóa ngẫu nhiên) không cần có sự tham gia của đối tượng bị tấn công. Như vậy đối với mã hóa số bits khóa cần lớn hơn nhiều.

2.5.3.2 Khởi tạo, mô tả cài đặt CMAC

Khởi tạo

CMAC nhận khóa mật K là đầu vào của quá trình khởi tạo. Nó sử dụng khóa để sinh hai khóa bổ sung K_1 và K_2.

Đầu vào

K: Secret key

Đầu ra

K_1, K_2: Khóa CMAC bổ sung

1. L = EncryptK(0) (* Hàm mã khối, với khóa K, khối vào là 0 *)

2. If MSB(L) = 0, then K1 = L << 1

else K1 = (L << 1) XOR Rb

3. If MSB(K1) = 0, then K2 = K1 << 1

else K2 = (K1 << 1) XOR Rb

4. Return K_1,

Các phép toán thực hiện trên xâu 64 hay 128 bits tùy thuộc vào độ dài khối được sử dụng trong mã khối. Giá trị Rb phụ thuộc vào độ dài khối. Nó là 0x87 khi khối là 128 bits và 0x1B khi khối là 64 bits. L là kết quả mã hóa của xâu toàn zero với khóa là K. Có K_1 và K_2, chúng ta tạo được MAC. K_1 và K_2 phải được giữ bí mật như khóa mã.

Cài đặt CMAC

Đầu vào

K: khóa bí mật ; K_1,K_2: khóa CMAC bổ sung

M: thông báo; L: số bit của thông báo

Tlen: độ dài thiết kế của thẻ MAC w: số bit/block

Đầu ra

T: thẻ MAC

1. Nếu L = 0, cho n = 1, ngược lại n = ceil(L/w)
2. Ký hiệu M_1,M_2,M_3,... M_n biểu diễn các khối của thông báo
3. Nếu L > 0 và L mod w = 0 thì

   Mn: Mn XOR K_1
4. Nếu L = 0 hoặc L mod w > 0 thì

   Gán vào một bit '1' và sau đó một số lượng các bit '0' để khối cuối đủ w bit

   Mn := Mn XOR K_2
5. C_0 =0
6. Cho i từ 1 tới n thực hiện

   C_i = EncryptK(Ci-1 XOR M_i)
7. T = MSBTlen(Cn) (* Lấy Tlen bits trái nhất của xâu Cn)
8. Return T

Các giá trị *Ci* nhận được như là bản mã của thông báo. Tuy nhiên không thể sử dụng nó như bản mã của thông báo được vì không có căn cứ chứng minh tính an toàn của nó đối với CMAC. Bản rõ cần được mã hóa với khóa khác để đảm bảo tính an toàn của CMAC.

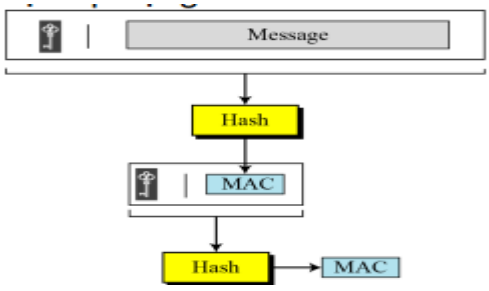
Nhìn chung tốc độ của CMAC phụ thuộc vào tốc độ mã hóa, với sự tối ưu hóa hàm tính toán, có thể sử dụng phép toán XOR đối với từ thay thế byte.

Mong muốn có một MAC dựa trên hàm băm

- Các hàm băm nhanh hơn mã khối đối xứng.
- Mã hàm băm có thể áp dụng một cách rộng rãi

Băm bao gồm một khóa (key) cùng với thông điệp

KeyedHash = Hash(key| Message)



Hình 2.17: Băm nhiều lần

2.5.4 Mã xác thực thông điệp sử dụng hàm một chiều

Tương tự CRC (*Cyclic Redundancy Check*), hàm băm chỉ có thể phát hiện các lỗi ngẫu nhiên bị nhiễu trong quá trình truyền. Hashed MAC kết hợp MAC và hàm băm để tăng cường an toàn cho hàm băm.

Hàm băm mật mã H, và khóa bí mật K

Đặc điểm:

- Dùng hàm băm nguyên mẫu.
- Cho phép thay thế dễ dàng hàm băm được nhúng vào trong trường hợp các hàm băm nhanh hơn hoặc nhiều bảo mật được tìm ra hoặc yêu cầu.
- Dùng và quản lý các khóa một cách dễ dàng.

KMAC sử dụng hàm hash một chiều có khóa mật và đưa nó thành thuật toán mã xác thực thông báo. Dưới đây trình bày các ứng dụng hash để tạo MAC.

HMAC thường được thiết kế dựa trên NMAC (Nested MAC), sử dụng hàm giả ngẫu nhiên – PRF (pseudo random function) để tạo hàm MAC với giới hạn an toàn chứng minh được.

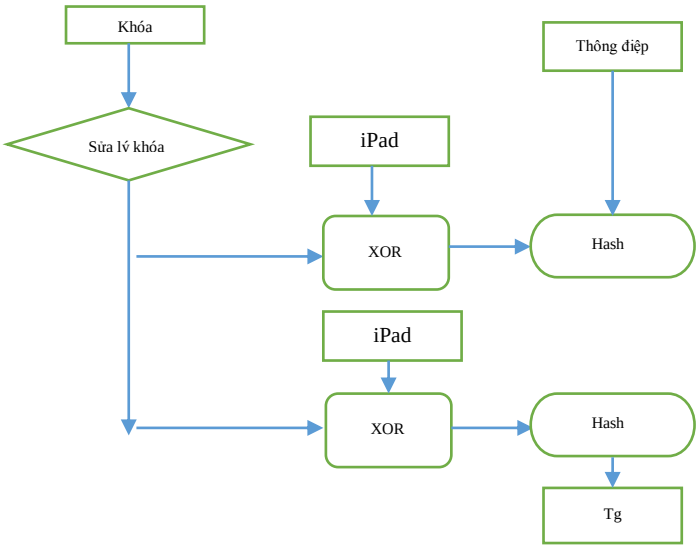
tag = hash(key1 || hash(key2 || message))

PRF ở đây có dạng Hash(key2||message). NMAC yêu cầu 2 khóa, một cho hash trong (inner) một cho hash ngoài (outer), hai khóa độc lập với nhau. HMAC dựa trên NMAC với sự khác biệt là hai khóa phụ thuộc tuyến tính, chỉ yêu cầu hàm PRF là an toàn, không yêu cầu nó kháng va chạm đối với một số hàm hash, nhưng nó cần chống được tấn công phương sai.

2.5.4.1 Thiết kế HMAC

Ban đầu HMAC được thiết kế để sử dụng với SHA1, trên thực tế có thể sử dụng nó an toàn với các hàm hash an toán khác. HMAC sinh hai khóa từ một khóa mật duy nhất

bằng cách XOR hai hằng số với nó. Nếu khóa lớn hơn kích thước khối nén, thì khóa trước hết phải được hash, đầu ra của nó coi như là khóa mật. Khóa mật được gán thêm các byte zero đảm bảo là độ dài của nó bằng độ dài khối nén. Khóa được sao thành hai bản, bản thứ nhất được XOR với 0x36 tạo khóa ngoài (outer), bản thứ hai XOR với 0x5C tạo khóa trong (inner)



Hình 2.18: Sơ đồ khối HMAC

Xâu gồm các byte 0x36 gọi là opad, xâu gồm các byte 0x5C là ipad. Khóa phải được bổ sung thêm (padded) để có độ dài bằng độ dài của khối nén để có thể thực hiện hash có khóa (keyed hash) thay cho hash thông qua việc tạo trạng thái hash khởi tạo phụ thuộc khóa. HMAC tương đương với việc lấy trạng thái khởi tạo hash một cách ngẫu nhiên trên khóa bí mật

2.5.4.2.Thuật toán

Đầu vào

K: khóa bí mật Message: thông báo chúng ta cần tính giá trị MAC cho nó.

W: độ dài khối nén Tlen: độ dài thẻ MAC theo thiết kế

Đầu ra

Tag: thẻ MAC

1. if length(K) > w then
   K = hash(K)
2. Làm đầy K bởi zeros cho đến khi nó có độ dài đúng w bytes
3. Tag = hash((opad XOR K) || hash((ipad XOR K) || message)).
4. Tag lấy Tlen bytes bằng cách chỉ giữ Tlen bytes đầu tiên.
5. Return Tag.

HMAC không quá phức tạp, trong thiết kế nó sử dụng hash như là hàm PRF, điểm hạn chế của nó là không song song được thuật toán, trong sơ đồ thuật toán chỉ có thể tính hash của (opad XOR K) khi tính hash trong.

2.5.5 Ứng dụng hàm MAC trên thực tế

MAC được thiết kế để đảm bảo xác thực giữa các bên trong kênh liên lạc, giúp việc gửi và nhận thông điệp được xác thực với nhau và khả năng bị kẻ tấn công giả mạo là rất thấp. Tuy vậy để sử dụng chúng an toàn, hiệu quả một số vấn đề sau đây cần đặc biệt lưu ý. Trên thực tế, MAC được triển khai sử dụng trong các giao thức SSL và TLS trên internet.

Ngày nay, MAC đươc sử dụng rộng rãi và phổ biến trong nhiều ứng dụng yêu cầu bảo mật cao, phòng tránh những tấn công về giả mạo, sửa đổi thông điệp.

2.5.5.1 Chống tấn công lặp

Trao đổi thông tin, các gói dữ liệu được chia thành các khối do vậy tấn công lặp rất dễ xảy ra. Khi đã bị tấn công các gói lặp vẫn giữ nguyên và việc giải mã vẫn bình thường nên điều nguy hiểm là không dễ phát hiện được sự tấn công. Có hai giải pháp cơ bản để chống tấn công lặp là sử dụng tem thời gian và con đếm. Nó gắn gói tin vào ngữ cảnh (context), cho phép người nhận xác định được vị trí gói tin trong luồng dữ liệu.

Sử sụng Tem thời gian

Mục đích của sử dụng tem thời gian là đảm bảo cho thông báo có hiệu lực chỉ trong khoảng thời gian nào đó. Ví dụ hệ thống chỉ cho phép gói dữ liệu có hiệu lực trong 30 giây từ khi tem thời gian được dán vào. Tuy vậy, từ góc độ an toàn, thật khó đặt tem thời gian đủ nhỏ để tránh tấn công lặp. Ví dụ ta đặt cửa sổ hiệu lực là 5 phút thì kẻ tấn công dễ dàng thực hiện tấn công lặp trong 5 phút. Nhưng nếu đặt cửa sổ hiệu lực là 3 giây, thì có khó khăn là trong khoảng thời gian đó chưa chắc đã truyền được gói dữ liệu. Mặt khác vấn đề đồng bộ thời gian giữa các nút của hệ thống là tương đối và có thể khi đó hệ thống nói chung không sử dụng được.

Sử dụng con đếm

Con đếm là giải pháp lý tưởng chống tấn công lặp. Ở mức nền tảng cần giải quyết câu hỏi giá trị tiếp theo của con đếm là gì? Ví dụ ta lưu trạng thái thanh ghi dịch phản hồi tuyến tính – LFSR (Linear Feedback Shift Register) trong một gói và hy vọng trạng thái tiếp theo của LFSR được điểm (clocked) ở gói tiếp theo. Nếu trạng thái của LFSR không đúng, ta có thể coi là không nhận được đúng gói dữ liệu. Một cách tiếp cận đơn giản mà thuận tiện là sử dụng giá trị con đếm là số tự nhiên. Chỉ cần giá trị con đếm không lặp lại trong cùng một phiên liên lạc. Giá trị con đếm được thỏa thuận giữa hai bên không cần phải ngẫu nhiên, và cũng không cần duy nhất nếu khóa MAC mới được sử dụng trong phiên. Khi mỗi gói gửi đi giá trị con đếm được tăng lên, bản thân giá trị con đếm được chứa trong thông báo ở phần đầu vào hàm MAC. Tức là tMAC sẽ là có giá trị con đếm và văn bản mã. Có thể mã hóa cả giá trị con đếm, nhưng điều đó không quan trọng vì độ an toàn tăng lên là "không đáng kể". Người nhận kiểm tra giá trị con đếm trước khi tiến hành bất kỳ động tác khác nào. Nếu giá trị con đếm nhỏ hơn hoặc bằng giá trị con đếm gói mới nhất, thì đó là gói lặp. Kích thước của con đếm có cỡ lớn nhất bằng kích thước gói dữ liệu cần gửi.

2.5.5.2 Sử dụng MAC

Việc sử dụng hàm MAC không nên quá lạm dụng trong các ứng dụng yêu cầu bảo mật, ít nhất không sử dụng nó như hàm hash hay hàm mã hóa

Một câu hỏi đặt ra là nên sử dụng CMAC hay HMAC? Dùng CMAC hay HMAC phải dựa trên vấn đề mà ta cần giải quyết. CMAC là hợp lý nếu ta đã có cơ sở mã hóa đảm bảo. Chúng ta có thể sử dụng lại chính mã chương trình hay thiết bị phần cứng đó để thực hiện xác thực. CMAC dựa trên mã khối nhưng với đầu vào nhỏ (so với hash) và đầu ra ngắn gọn, thời gian trễ cho tính toán nhỏ. Ví dụ với thông điệp ngắn khoảng 16 hay 32 byte, CMAC sử dụng AES cho kết quả nhanh, ít trễ hơn HMAC. HMAC thắng thế khi áp dụng cho thông điệp kích thước lớn. Hash xử lý dữ liệu với số vòng/byte ít hơn mã hóa. Hash cũng thường tạo ra dữ liệu thẻ MAC dài theo yêu cầu cụ thể. Nhưng HMAC yêu cầu cài đặt hàm hash.

2.5.5.3 Thứ tự thực hiện mã hóa và MAC

Thứ tự của mã hóa và MAC, nói chung không ảnh hưởng đến yêu cầu về độ an toàn, nhưng nó cũng có sự khác biệt nhỏ. Quan trọng nhất là khóa mã và khóa MAC không có liên hệ với nhau. Đơn giản nhất là sử dụng hàm tạo giữa khóa phải tách biệt khóa dành cho mã hóa và khóa dành cho MAC.

- Mã hoá trước, MAC sau

Ở chế độ này mã hóa bản rõ trước sau đó MAC bản mã (cùng với con đếm hay tem thời gian). Vì mã hóa là ánh xạ một cách ngẫu nhiên hoàn toàn bản rõ (đòi hỏi lựa chọn IV "Initial Vector" thích hợp) thông báo ngẫu nhiên. Như vậy MAC không lộ thông tin về bản rõ, và người nhận không cần phải giải mã để từ chối gói dữ liệu không hợp lệ.

- MAC trước, mã hoá sau

Ở đây trước tiên ta MAC bản rõ (cùng với con đếm hoặc tem thời gian) và sau đó mã bản rõ. Đầu vào của MAC không phải ngẫu nhiên và phần lớn thuật toán MAC (ít nhất là CMAC và HMAC) không sử dụng vectơ khởi tạo IV(Initial Vector), do vậy đầu ra của MAC cũng không ngẫu nhiên (nếu chúng ta không sử dụng chống tấn công lặp). Như vậy chúng ta đưa cho kẻ tấn công bản mã và thẻ MAC của bản rõ. Nếu bản rõ không đổi, bản mã có thể thay đổi (do lựa chọn IV tốt), nhưng thẻ MAC vẫn như cũ. Tuy nhiên thường ta có tem thời gian hoặc con đếm là một phần đầu vào hàm MAC. Và vì MAC là PRF, nên nó chống được tấn công ngay cả khi bản rõ không đổi. Tốt nhất là áp dụng mã hóa đối với cả thẻ MAC cùng như bản rõ, như vậy giá trị thẻ MAC không bị kẻ tấn công tiếp cận. Hạn chế ở đây là người nhận phải giải mã để so sánh thẻ MAC.

2.6 CHỮ KÝ SỐ

Các cơ sở toán học, các hệ mã hóa khóa bí mật, khóa công khai, một số mã hóa khóa công khai, hàm băm SHA và mã xác thực đã trình bày trên là nền tảng cho việc sử dụng chữ ký số vào việc bảo mật, xác thực thông tin. Đã có rất nhiều thuật toán chữ ký số khác nhau, luận văn này giới thiệu thuật toán chữ ký số thông dụng RSA.

2.6.1 Chữ ký điện tử

Chữ ký số là một cơ chế xác thực cho phép người tạo thông tin dùng khóa riêng của mình để xử lý khối thông tin theo một thuật toán nào đó giúp người nhận thông tin kiểm chứng được tính toàn vẹn về nội dung và nguồn gốc thông tin.

Một trong những ứng dụng của kỹ thuật mã hóa khóa công khai là chữ ký số. Chữ ký số được phát triển và ứng dụng rộng rãi hiện nay dựa trên một số thuật toán như RSA, DSA, ElGamal, Elliptic …là cơ sở quan trọng để hình thành hạ tầng khóa công khai cho phép người sử dụng mạng công cộng internet để trao đổi dữ liệu và tìm một cách an toàn, thông qua việc sử dụng một cặp khóa công khai và bí mật được cấp phát.

Chữ ký điện tử (electronic signature) được tạo lập dưới dạng từ, chữ, số, ký hiệu, âm thanh hoặc các hình thức khác bằng phương tiện số, gắn liền hoặc kết hợp một cách logic với thông điệp số, có khả năng xác nhận người ký thông điệp dữ liệu đã ký. *Chữ ký điện tử* cũng giống như chữ viết tay, tức là chữ ký điện tử được dùng để xác nhận lời hứa hay cam kết của người nào đó và sau đó không thể chối bỏ được. Chữ ký điện tử không đòi hỏi phải sử dụng giấy mực mà nó gắn đặc điểm nhận dạng của người ký vào cam kết nào đó. Như vậy, chữ ký điện tử sẽ chứng thực được định danh người gửi và bảo vệ sự toàn vẹn dữ liệu.

Chữ ký điện tử được sử dụng trong các cơ quan nhà nước khi gửi/nhận Email và truyền tải văn bản dạng số trên mạng internet. Xuất phát từ thực tế, chữ ký điện tử cần đảm bảo các chức năng: xác định được người chủ của một dữ liệu nào đó và xác thực dữ liệu đó có bị thay đổi hay không?

Chữ ký điện tử bao gồm ba thành phần: thuật toán tạo ra khóa, hàm tạo chữ ký và hàm kiểm tra chữ ký. Hàm tạo ra chữ ký là hàm tính toán chữ ký trên cơ sở khóa mật và dữ liệu cần ký. Hàm kiểm tra chữ ký là hàm kiểm tra xem chữ ký đã cho có đúng với khóa công khai không? Khóa này mọi người có quyền truy cập cho nên mọi người đều có thể kiểm tra được chữ ký.

2.6.2 Chữ ký số

Chữ ký số "digital signature" là một dạng chữ ký điện tử được tạo ra bằng sự biến đổi một thông điệp có sử dụng hệ mật mã khóa công khai, theo đó người có thông điệp ban đầu và khóa công khai của người ký có thể xác thực được chữ ký số vừa ký

Chữ ký của một người trên tài liệu thường đặt ở cuối văn bản để xác nhận nguồn gốc hay trách nhiệm của người ký lên tài liệu đó.Với tài liệu đã được "số hóa" nếu chữ ký đặt ở cuối văn bản thì việc sao chép "chữ ký số" là dễ dàng và không thể phân biệt bản gốc với bản sao vì chữ ký số là các số 0,1. Vì vậy một chữ số đặt ở cuối "tài liệu số" không thể chịu trách nhiệm đối với toàn bộ nội dung văn bản, mà chữ ký số phải được ký trên từng bit của dữ liệu đó. Những chữ ký số cũng không thể ký trên bất kỳ tài liệu nào với độ dài tùy ý, vì như vậy chữ ký số sẽ có độ dài rất lớn. Với tài liệu dài thì ký trên văn bản đại diện của nó. Văn bản đại diện của tài liệu được tạo ra bởi hàm băm.

Với chữ ký thông thường thì nó là một phần của tài liệu, nhưng chữ ký số không gắn theo kiểu vật lý vào thông điệp. Đối với chữ ký thông thường người ta kiểm tra bằng cách so sánh với chữ ký đúng nhưng như vậy cũng không phải là an toàn, nó có thể giả mạo. Đối với chữ ký số, người ta có thể kiểm tra thông qua thuật toán kiểm tra công khai. Bởi vì chữ ký số là một chuỗi số liên quan đến thông điệp và do vậy khi thông điệp thay đổi thì chữ ký số cũng thay đổi, chính vì vậy chữ ký số đảm bảo tính toàn vẹn của thông điệp, chữ ký số không thể sử dụng lại và cũng không thể giả mạo được. Hai thuộc tính không thể làm giả được và xác thực không chối bỏ của người ký chữ ký số là nguyên tắc để đảm bảo an toàn cho các hệ thống sử dụng chữ ký số trong truyền tải thông tin và dữ liệu qua mạng

Định nghĩa: [5]Sơ đồ chữ ký bao gồm các thành phần sau

1. Không gian bản rõ M.
2. Không gian chữ ký S.
3. Không gian khóa K để tạo nên chữ ký, không gian khóa K' để kiểm tra chữ ký.
4. Thuật toán hiệu quả để tạo nên khóa Gen: $N \rightarrow K \times K'$, ở đây K và K' tương ứng với không gian khóa mật và khóa công khai
5. Thuật toán tạo chữ ký Sign: $M \times K \rightarrow S$.
6. Thuật toán kiểm tra chữ ký Verfy: $M \times K \times K' \rightarrow \{True, False\}$.

Đối với bất kỳ khóa tạo chữ ký $sk \in K$ và bất kỳ bản tin $m \in M$ chữ ký bức điện được ký hiệu:

$$s \leftarrow Sign_{sk}(m).$$

Đối với bất kỳ khóa mật của chữ ký $k \in K$, tương ứng với khóa công khai để kiểm tra chữ ký $sk \in K'$, bất kỳ bản tin $m \in M$ và chữ ký $s \in S$ cần thõa mãn điều kiện sau:

$$\mathbf{Ver}_{\mathbf{pk}}(\mathbf{m}, \mathbf{s}) = \begin{cases} \text{Đúng, nếu } \mathbf{s} = \mathbf{Sig}_{\mathbf{pk}}(\mathbf{m}) \\ \text{Sai, nếu } \mathbf{s} \neq \mathbf{Sig}_{\mathbf{pk}}(\mathbf{m}) \end{cases}$$

Một biện pháp để ký là chia tài liệu ra thành các đoạn nhỏ và sau đó ký lên từng đoạn và ghép lại. Những phương pháp có nhược điểm là chữ ký lớn, thứ hai là ký chậm vì hàm ký là các hàm mũ, thứ ba là chữ ký có thể bị đảo loạn các vị trí không đảm bảo tính nguyên vẹn của tài liệu. Chính vì điều đó mà khi ký thông điệp thì chúng ta ký lên giá trị hàm băm.Vì giá trị đầu ra hàm băm luôn cho chiều dài là xác định.

Chức năng của chữ ký điện tử:

Xác thực được nguồn gốc tài liệu: tùy thuộc vào từng bản tin mà có thể thêm các thông tin nhận dạng, như tên tác giả, nhãn thời gian,…

Tính toán vẹn dữ liệu: vì khi có một sự thay đổi bất kỳ vô tình hay cố ý lên bức điện thì giá trị hàm hash sẽ bị thay đổi và kết quả kiểm tra bức điện sẽ không đúng.

*Chống từ chối bức điện***:** vì chỉ có chủ của bức điện mới có khóa mật để ký bức điện.

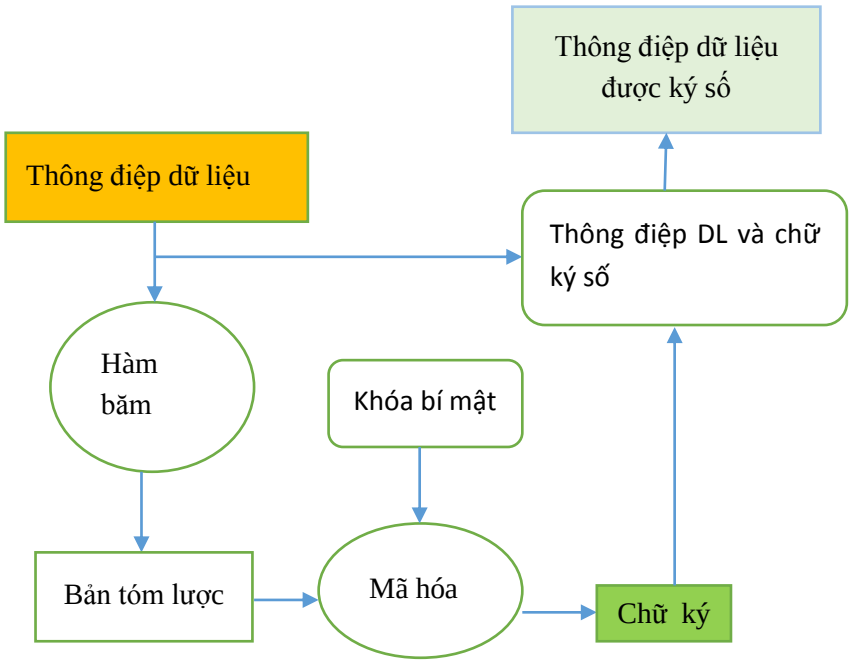
Các chức năng tấn công đối với chữ ký điện tử:

1. Tội phạm có thể giả mạo chữ ký tương ứng với văn bản đã chọn.
2. Tội phạm thử chọn bức điện tương ứng với chữ ký đã cho.
3. Tội phạm có thể ăn trộm khóa mật và có thể ký bất kỳ một bức điện nào nó muốn giống như chủ của khóa mật.
4. Tội phạm có thể giã mạo ông chủ ký một bức điện nào đó.
5. Tội phạm có thể đổi khóa công khai bởi khóa của mình.

2.6.3 Cách tạo chữ ký số

2.6.3.1 Quy trình tạo chữ ký số

Dùng giải thuật băm để tóm lược thông điệp cần truyền đi, kết quả ta được một message digest (MD), sử dụng khóa bí mật người gửi để mã hóa thông điệp thu được.



Hình 2.19: Qui trình tạo chữ ký

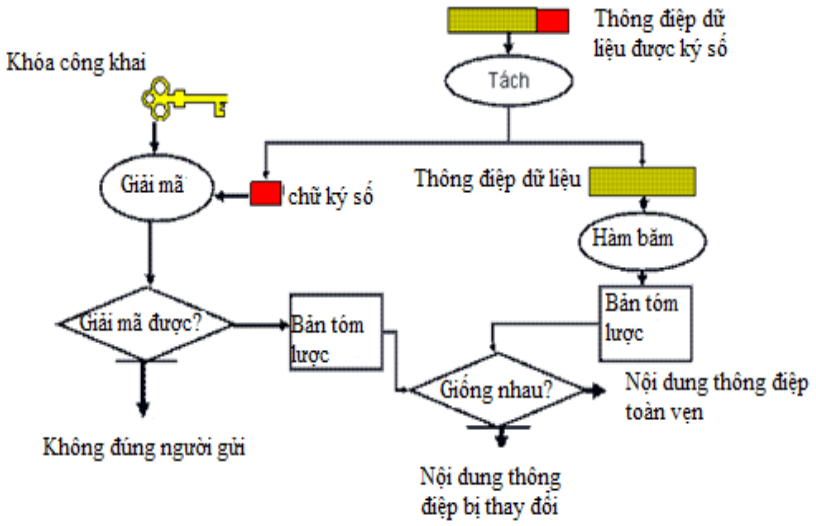
2.6.3.2 Quy trình kiểm tra chữ ký

Dùng public key của người gửi để giải mã chữ ký số của thông điệp.

Dùng giải thuật SHA băm thông điệp đính kèm.

So sánh kết quả thu được ở trên, nếu thấy trùng nhau ta kết luận thông điệp này không bị thay đổi trong quá trình truyền và thông điệp này là của người gửi.

Sơ đồ



Hình 2.20: Qui trình kiểm tra chữ ký số

Chữ ký số (digital signature) được tạo ra bằng sự biến đổi một thông điệp sử dụng hệ thống mật mã công khai, theo đó người có thông điệp dữ liệu ban đầu và khóa công khai của người ký có thể xác định được.

2.6.3.3 Thuật toán chữ ký số

Việc dùng mã xác thực chỉ có thể bảo vệ hai thực thể tham gia trong phiên giao dịch
đối với sự xâm phạm của thực thể thứ 3, tuy nhiên chưa thể bảo vệ giữa hai thực thể với nhau. Ví dụ: giả sử A gửi cho B một thông điệp đã được xác thực theo mô hình sử dụng mã xác thực thông điệp (MAC), A có thể tạo ra một thông điệp khác và tính toán MAC dựa vào key share giữa A và B, lúc này A nói thông điệp mà A tạo ra là do B gửi, B không có khả năng chứng minh điều này là sai.

Giải pháp để giải quyết vấn đề này là chữ ký số. Chữ ký số có khả năng:

- Xác minh người ký và thời điểm ký.
- Xác minh nội dung thông điệp tại thời điểm ký.
- Có thể xác minh bởi thực thể khác giải quyết việc tranh cải của hai thực thể này.

Có hai kỹ thuật tạo chữ ký:

- Ký trực tiếp (Direct Digital Signature): Hai thực thể tham gia truyền thông tin trực tiếp ký trên thông điệp, đòi hỏi đích phải biết public key của nguồn để xử lý trong quá trình xác minh. Chữ ký số này có thể là việc mã hóa cả thông điệp với private key hay việc mã hóa hash code của thông điệp với khóa riêng. Vấn đề bảo mật khóa riêng trong kỹ thuật ký trực tiếp.
- Ký qua trọng tài:

  + Trường hợp 1: sử dụng kỹ thuật mật mã đối xứng, trọng tài có thể xem nội dung thông tin.

  $X \rightarrow A$:

  $$M + E([ID_X + H(M), Kxa$$

  $A \rightarrow Y$:

  $$E([ID_x + E(M, K_{xy})], K_{ay}) + T], Kxa$$

  + Trường hợp 2: sử dụng mã đối xứng, trọng tài không thể xem nội dung thông tin.

  $X \rightarrow A$:

  $$ID_X + E(M, K_{xy}) + E([ID_X + H\left(E(M, K_{xy})\right)], Kxa$$

  $A \rightarrow Y$:

  $$E([ID_x + E(M, K_{xy})], K_{ay}) + E([ID_X + H\left(E(M, K_{xy})\right) + T], Kxa$$

  + Trường hợp 3: sử dụng mã bất đối xứng, trọng tài không thể xem nội dung thông tin.

  $X \rightarrow A$:

  $$ID_X + E([ID_X + E(E(M, PR_x), PU_y)], PR_x$$

  $A \rightarrow Y$:

  $$E + E(E(M, PR_x), PU_y) + T], KRa$$

Mô hình chữ ký số cũng giống mô hình mã hóa công khai do vậy thuật toán RSA có thể áp dụng để xây dựng chữ ký số.

Yêu cầu chữ ký số:

- Chữ ký phải dựa vào thông điệp được ký.
- Chứa thông tin duy nhất của người gửi để tránh giả mạo.
- Dễ nhận diện và xác nhận chữ ký số.
- Khó khăn giả mạo chữ ký.

Để có thể sử dụng chữ ký số ta cần sử dụng một hàm băm để rút gọn thông điệp có chiều dài bất kỳ thành một giá trị băm có kích thước cố định. Giá trị này có thể được kiểm chứng tính toàn vẹn thông tin.

Đặc điểm của chữ ký số:

- Tính xác thực: bảo đảm người ký là người tạo ra nó.
- Tính an toàn: Không thể giả chữ ký số nếu như không biết thông tin bí mật tạo chữ ký.
- Không thể dùng lại: một chữ ký số không thể dùng cho một tài liệu khác.
- Tính hiệu quả: ký và xác minh nhanh chóng dễ dàng.

2.6.4 Sơ đồ chữ ký số RSA

Tạo khóa:

Quá trình tạo khóa cho sơ đồ chữ ký RSA giống như quá trình hình thành khóa của hệ mật RSA, tức là: người gửi chọn cặp số nguyên tố đủ lớn p và q, với N = p.q. Chọn số nguyên b thỏa mãn UCLN(b,$\varphi(N)) = 1$.Người gửi đi xác định số nguyên d thỏa mãn phương trình bd $\equiv 1\ (mod\ \varphi(N))$. Số d là khóa mật của người gửi.

Tạo chữ ký

Để tạo ra chữ ký số của bức điện $m\ \in Z_N^*$ người gửi tạo ra số

$$S = Sign_d\ (m) \leftarrow m^d\ (mod\ N).$$

Thẩm tra chữ ký:

Để thẩm tra chữ ký S, người nhận kiểm chứng bằng thủ tục

$$Verify_{(N,b)}(m,s) = true, nếu\ như\ m \equiv s^b\ (\text{mod N})$$

Quá trình tạo chữ ký và thẩm tra chữ ký giống với quá trình mã và giải mã, khác là quá trình tạo chữ ký là dùng khóa mật còn quá trình thẩm tra thì dùng khóa công khai.

2.6.4.1 Độ an toàn của sơ đồ chữ ký số RSA

Bài toán căn bản bảo đảm độ an toàn của Sơ đồ chữ ký RSA

Tách số nguyên n thành tích của 2 số nguyên tố: n = p*q. Vì nếu giải được bài toán này thì có thể tính được khóa mật a từ khóa công khai b và phần tử công khai n.

Người gửi G gửi tài liệu **x** cùng chữ ký **y** đến người nhận N, có 2 cách xử lý.

- ***Ký trước, Mã hóa sau:***

G ký trước vào x bằng chữ ký y, sau đó mã hoá x và y nhận được z. G gửi z cho N. Nhận được z, N giải mã z để được x, y. Tiếp theo kiểm tra chữ ký y đúng hay sai.

- ***Mã hóa trước, Ký sau:***

G mã hoá trước **x** bằng u, sau đó ký vào u bằng chữ ký v. G gửi (u, v) cho N. N Nhận được (u, v), N giải mã u được x. Tiếp theo kiểm tra chữ ký v đúng hay sai.
Giả sử H lấy trộm được thông tin trên đường truyền từ G đến N.
Trong trường hợp *ký trước, mã hoá sau* H lấy được z. Trong trường hợp *mã hoá trước, ký sau* H lấy được (u,v).
Để tấn công **x** trong hai trường hợp, H đều phải giải mã thông tin lấy được.
Để tấn công vào chữ ký, thay bằng chữ ký (giả mạo), thì xảy ra điều gì?

- Trường hợp *ký trước, mã hoá sau* để tấn công chữ ký **y**, H phải giải mã **z** mới nhận được **y**.
- Trường hợp *mã hoá trước, ký sau* để tấn công chữ ký **v**, H đã sẵn có **v**, H chỉ việc thay **v** bằng **v'**. H thay chữ ký **v** trên u, bằng chữ ký của H là **v'**, gửi (**u**,**v'**) đến N. Khi nhận được **v'**, N kiểm thử thấy sai, gửi phản hồi lại G. G có thể chứng minh chữ ký đó là giả mạo. G gửi chữ ký đúng **v** cho N, nhưng quá trình truyền tin sẽ bị chậm lại.

CHƯƠNG 3: ỨNG DỤNG CHỮ KÝ ĐIỆN TỬ ĐẢM BẢO TÌNH TOÀN VẸN DỮ LIỆU TRONG TRƯỜNG HỌC

3.1. Thực trạng quy trình ra đề thi và bảo mật thông tin đề thi các trường ĐH - CĐ.

a) Quy trình biên soạn, duyệt, quản lý, sao in và sử dụng đề thi

Quy trình biên soạn, duyệt, quản lý, sao in và sử dụng đề thi được thực hiện như sau:

1. Biên soạn đề thi và đáp án hoặc hướng dẫn chấm (gọi chung là đáp án);
2. Duyệt đề thi và bàn giao cho phòng Khảo thí và Kiểm định chất lượng giáo dục (KT&KĐCLGD);
3. Sao in đề thi và quản lý đề thi;
4. Bàn giao đề thi cho ban coi thi;
5. Sử dụng đề thi.

b) Biên soạn đề thi và đáp án

1. Giáo viên được phân công soạn đề thi, đáp án (chính thức và dự phòng), nộp cho trưởng bộ môn kiểm tra nội dung, thể thức đề thi đúng quy định và ký duyệt.

2. Khi biên soạn cần chú ý các điểm sau đây:

- Sử dụng máy tính để biên soạn cần khóa mã. Khi sao in, chụp xong phải kiểm tra lại và hủy ngay bản thảo và bản in thử, in hỏng, in thừa,...(*nếu không cần lưu*).
- Đề thi phải ghi rõ được hay không được sử dụng tài liệu, từ điển,…;
- Bài làm của thí sinh được thực hiện trên giấy thi hoặc phiếu trả lời do cán bộ coi thi phát cho thí sinh. Không có bất cứ phần nào làm trực tiếp trên đề thi. Những đề thi có cả phần trắc nghiệm và tự luận Giáo viên thực hiện theo mẫu phiếu trả lời trắc nghiệm và tự luận. Những đề thi có yêu cầu trả lời khác với phiếu trả lời thì Giáo viên thiết kế phiếu trả lời cho phù hợp yêu cầu của đề.
- Đóng gói đề thi phải ghi đầy đủ các thông tin bên ngoài túi đựng đề thi và niêm phong cẩn mật
- Giáo viên giao đề thi, đáp án (chính thức và dự phòng) một lần cho Thư ký khoa chậm nhất là một ngày sau khi kết thúc môn dạy trên lớp theo thời khóa biểu.
- Những môn có thí sinh thi lần 2 thì có thể sử dụng đề dự phòng.
- Khi giao đề thi, đáp án phải có ký giao với người có trách nhiệm nhận đề thi và phải ghi rõ số lượng đề thi, đáp án, thời gian giao. Riêng đề thi hoặc đề kiểm tra môn Nghe của các lớp Cao đẳng, chứng chỉ Anh văn và thực hành Tin học phải chép trên đĩa CD.
- Đối với các lớp liên kết đào tạo: Thực hiện từ khoản 2 đến khoản 5 của Điều 2. Riêng đề và đáp án kiểm tra chứng chỉ có Quyết định phân công của Hiệu trưởng.

c) Duyệt đề thi và bàn giao cho phòng KT&KĐCLGD

- Người có trách nhiệm duyệt đề phải xem kỹ nội dung, thể thức của đề thi có đảm bảo yêu cầu không, nếu có chỉnh sửa thì phải trao đổi với người biên soạn đề thi để hoàn thiện theo quy định.
- Đề thi đã được duyệt, phải niêm phong lại theo chế độ bảo mật và có ký xác nhận của Trường ĐH, CĐ
- Thư ký các khoa giao đề thi cho Phòng KT&KĐCLGD đúng thời gian quy định, (trước ngày thi đầu tiên của kỳ thi ít nhất 01 tuần). Riêng đối với đề thi các lớp Đại học ngoài chính quy thì thư ký phòng Đào tạo phải nộp trước ngày thi đầu tiên của kỳ thi ít nhất 02 ngày theo lịch thi, không tính các ngày nghỉ, lễ, tết. Có ký sổ biên bản theo dõi giao, nhận đề thi.

d) Sao in đề thi và quản lý đề thi

- Việc sao, in đề thi do bộ phận sao in đề thực hiện. Bộ phận sao in đề thi thực hiện quy trình sao in đề theo Nội quy phòng sao in đề thi.
- Người trực tiếp sao in đề thi là những người không có người thân (vợ, chồng, con, anh, chị, em ruột) dự thi.
- Trước khi sao in, Trưởng phòng KT&KĐCLGD phân công cụ thể cán bộ, nhân viên thực hiện sao in đề cho từng đợt thi, buổi thi, môn thi, số lượng đề cần được nhân bản (có sổ biên bản).
- Đề thi được phép mở niêm phong để thực hiện việc sao in khi có mặt đủ 2 thành viên trong bộ phận sao in đề thi; khi mở đề thi phải làm biên bản tổng hợp ghi lại hiện trạng bảo mật của từng đề thi và các hiện tượng bất thường khác (nếu có).
- Đề thi phải được đánh máy và in thử rõ ràng, chính xác, sạch, đẹp, đúng quy cách. Nếu do sơ xuất khi ra đề. Những đề thi in mờ, không rõ chữ hoặc lỡ viết tay hay cần tiết kiệm giấy in, bộ phận sao in đề thi có thể đánh máy lại, khi được sự chấp thuận của Trưởng phòng KT&KĐCLGD hoặc Chủ tịch Hội đồng kiểm tra thực hiện việc xóa file trên máy vi tính sau khi có bản in đạt yêu cầu. Đề sao in được che phần tên và chữ ký của người ra đề. Các giấy tờ đánh máy hoặc in hỏng đề thi không được cho vào sọt rác mà phải thực hiện chế độ niêm phong và được bảo quản như đề thi.
- Khi in xong đề thi, cán bộ sao in đóng gói từng túi đựng đề thi và phải ghi đầy đủ các thông tin bên ngoài túi đựng đề thi, thực hiện chế độ niêm phong túi đựng đề thi và được bảo quản như đề thi chưa sao in, có chữ ký niêm phong của hai cán bộ sao in. Cán bộ sao in đề thi phải ký tên và ghi rõ họ tên vào biên bản sao in đề thi.
- Người đóng gói đề thi phải làm đúng quy cách thủ tục, bảo đảm đúng số lượng đề thi, đúng môn thi ghi ở phong bì đề thi, đủ số lượng đề thi cho từng điểm thi,

phòng thi, không có tờ trắng, tờ hỏng và ghi đầy đủ thông tin in trên túi đựng đề thi.

- Máy vi tính, máy in, máy photo không được mang ra khỏi phòng sao in đề thi.
- Túi đựng đề thi làm bằng giấy đủ kín, tối, được dán chặt, không bong mép, đóng dấu niêm phong và có chữ ký của hai cán bộ sao in đề.
- Đề thi được bảo quản theo chế độ quản lý tài liệu mật, không được phép mang ra khỏi cơ quan, phòng làm việc, đề thi phải được cất giữ vào tủ, hòm hoặc két có khoá an toàn.
- Riêng đề kiểm tra chứng chỉ Anh văn - Tin học, thực hiện sao in theo phân công của Chủ tịch Hội đồng kiểm tra có dán nhãn niêm phong và chữ ký của cán bộ sao in trên nhãn niêm phong.

e) Bàn giao đề thi cho Ban coi thi

Bộ phận sao in đề thi quản lý, bảo mật đề thi đến khi giao cho phòng Đào tạo (có biên bản giao, nhận chi tiết). Phòng Đào tạo bảo mật đề thi cho đến khi giao cho Ban coi thi vào đầu các buổi thi.

f) Sử dụng đề thi

Ban coi thi quản lý, bảo mật đề thi cho đến khi giao cho cán bộ coi thi trước mỗi buổi thi.

Cán bộ coi thi quản lý, bảo mật đề thi và giao cho thí sinh theo quy định coi thi.

Khi đề thi trao đến tay cán bộ coi thi, thì nhiệm vụ bảo mật do cán bộ coi thi chịu trách nhiệm. Đề thi được mở để phát cho thí sinh theo đúng số lượng, thời gian qui định của mỗi kỳ thi, nếu phát hiện có gì sai sót trong khi thi thì cán bộ coi thi phản ánh ngay với Hội đồng kiểm tra hoặc Ban chỉ đạo thi hết môn để kịp thời giải quyết.

g) Điều khoản thi hành

Mọi bộ phận, cá nhân khi được giao nhiệm vụ liên quan đến đề thi phải tuân thủ thực hiện nhiệm vụ theo qui định này và thực hiện nghiêm túc chế độ bảo mật cho đến khi đề thi đã được sử dụng xong.

3.2. Yêu cầu giải pháp quản lý đề thi theo phương pháp hiện đại.

Trong một trường học có thể có rất nhiều các loại đề thi, nếu không có cách xử lý hiệu quả và khoa học thì sẽ rất khó trong công tác quản lý, lưu trữ và tìm kiếm sau này. Từ mô hình quản lý nêu trên, giúp ta thấy được việc xử lý đề thi phần lớn trên giấy tờ làm mất nhiều thời gian và công sức. Do đó ngoài những cách sắp xếp, lưu trữ truyền thống như lưu sổ, phân loại, lưu kho thì việc áp dụng một phần mềm quản lý đề thi là cần thiết ở thời điểm hiện nay. Áp dụng phần mềm yêu cầu giải quyết các vấn đề:

- Đảm bảo đề thi được xử lý chính xác, đúng thời hạn, an toàn và hiệu quả. Việc lưu trữ công văn và tìm kiếm nhanh chóng và thuận tiện .
- Nâng cao hiệu quả xử lý đề thi và lập hồ sơ công việc.

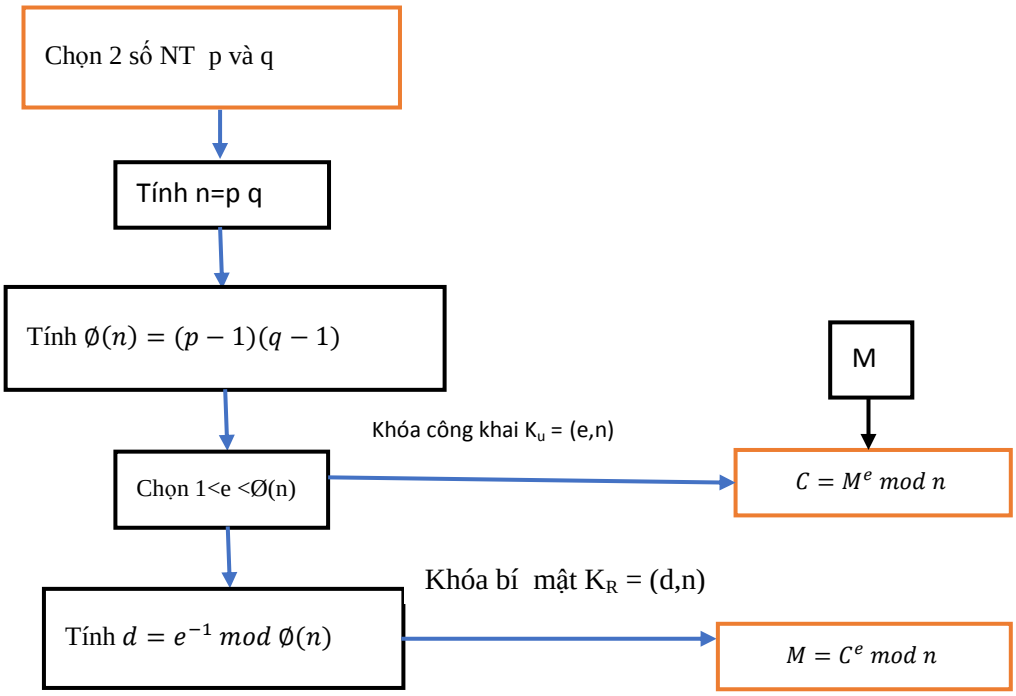
- Trao đổi thông tin, tiếp nhận ý kiến chỉ đạo từ cấp trên một cách kịp thời để thực hiện tốt nhiệm vụ được giao.
- Tiết kiệm thời gian, công sức, chi phí khi quản lý, lưu trữ và tìm kiếm đề thi.
- Việc quản lý đề thi thông qua một phần mềm, giải pháp quản lý đề thi chuyên dụng được cho là phương pháp tối ưu nhất hiện nay, vì nó vừa mang lại hiệu quả cao trong công tác quản lý đề thi và cũng như phù hợp với xu thế đổi mới công tác quản lý trường học thông minh.

3.3. Quá trình ký và xác thực ký số

3.3.1. Tạo và trao đổi khóa

Hệ thống sinh 2 khóa cho 2 quá trình mã hóa $K_u(e,n)$ và xác thực thông điệp $K_R(d,n)$. Chọn 2 số nguyên tố p & q đủ lớn, "hai nguyên tố này phải giữ bí mật". Tuy nhiên, khi công khai các khóa dùng đê mã hóa lại nảy sinh vẫn đề một người nào đó giả danh sử dụng để mã hóa các thông báo gửi đến bên nhận làm họ khong thể phân biệt được thông báo đó hợp lệ hay không. Có một phương pháp giải quyết vấn đề này mà điển hình là cơ chế chữ ký số.

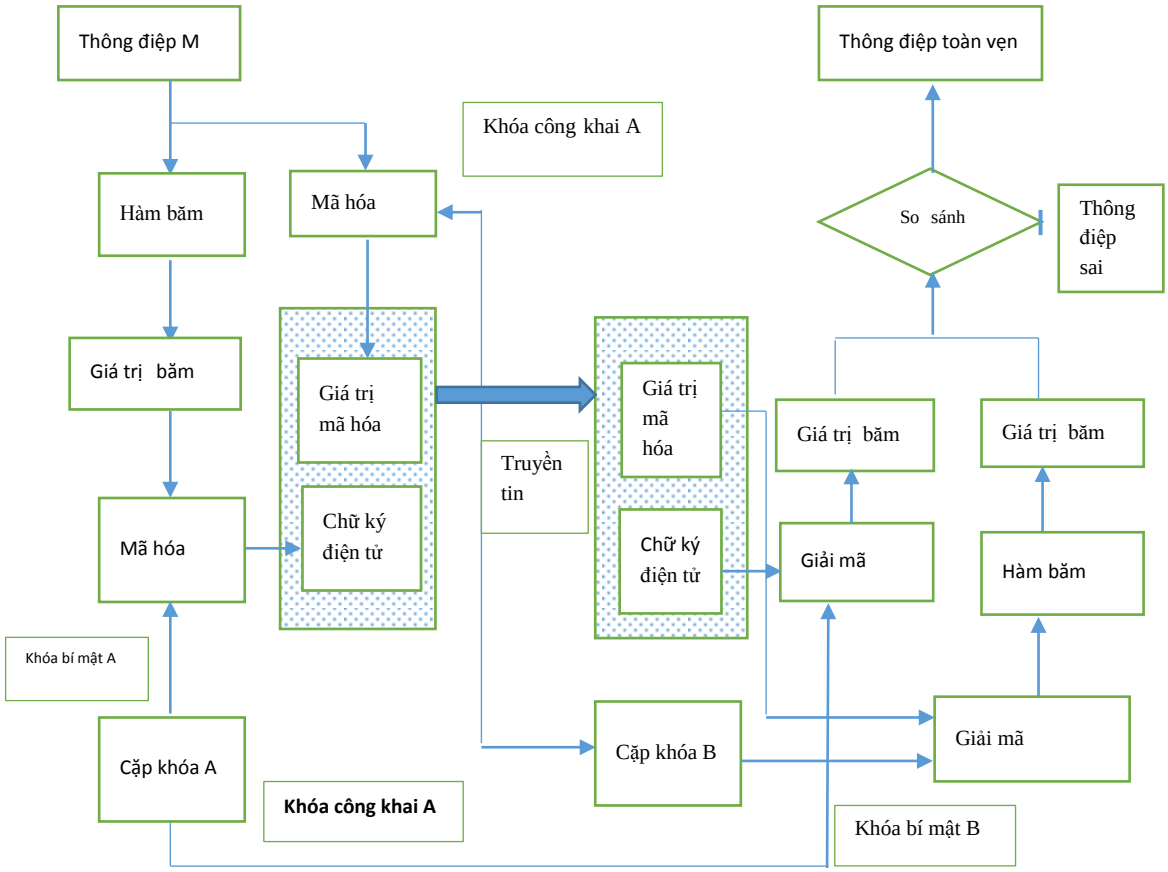
- Tính module n
- Sinh khóa e
- Tính khóa d



Hình 3.1: Mô hình tính 2 khóa

Các thành phần của chương trình:

1. Thông điệp ban đầu: P
2. Khóa bí mật K_R dùng trong quá trình giải hóa
3. Khóa công khai K_u dùng trong quá trình mã hóa.
4. Bản mã C và được trao đổi thông qua ường truyền tin



Hình 3.2 Mô hình quy trình tạo chữ ký và thẩm định chữ ký

3.3.2. Quá trình tạo chữ ký số

Giả sử A muốn gửi một thông điệp điện tử bí mật cho B và giả sử A đã có được khóa công cộng của B.

Số hóa thông điệp:

Bước 1: Dữ liệu ban đầu sử dụng hàm băm để nén dữ liệu thành "giá trị băm" để truyền đi.

Sử dụng khóa private key (khóa bí mật) của người gửi để mã hóa "giá trị băm" thu được ở bước 1. Kết quả thu được gọi là chữ ký điện tử của thông điệp ban đầu.

Bước 2: Dữ liệu ban đầu cùng với khóa công khai của người gửi để mã hóa thu được "giá trị mã hóa".

Bước 3: Gộp "chữ ký điện tử" thu được bước 1 vào "giá trị mã hóa" thu được bước 2. Công việc này gọi là "ký nhận" vào dữ liệu. Mọi sự phát hiện trong giai đoạn kiểm tra.

Thông điệp đã được ký *tạo và lưu chữ ký số*. Chọn nút "lưu chữ ký vào file" để thực hiện việc lưu chữ ký. Nội dung file được lưu bao gồm: "nội dung chữ ký, module n và khóa e".

3.3.3. Quá trình xác thực chữ ký

Dữ liệu đã được ký nhận đến B tách làm hai phần là: "chữ ký điện tử" và "giá trị mã hóa" để xử lý riêng.

Bước 1: Sử dụng khóa công khai của A giải mã "chữ ký điện tử" thu được "giá trị băm".

Bước 2: Sử dụng khóa bí mật B để giải mã "giá trị mã hóa", sau đó sử dụng hàm băm để tính toán chuỗi đại diện thu được "giá trị băm".

Bước 3: So sánh hai "giá trị băm" của bước 1 và bước 2 trên xem thông điệp có toàn vẹn hay không.

- Nếu toàn vẹn, người nhận B chấp nhận thông điệp. Chữ ký thành công
- Nếu không toàn vẹn, người nhận B có thể bỏ qua thông điệp

3.4. Chương trình demo

3.4.1. Giới thiệu chương trình

Chương trình xây dựng gồm 2 modul chính:

Demo thuật toán băm SHA3 (SHA3- 224; SHA3 – 256; SHA3 -384; SHA3- 512)

Demo chữ ký số lên đề thi

Cách sử dụng chương trình :

Ký lên đề thi : người gửi cần làm các bước sau để ký lên đề thi

Bước 1: Mở chương trình lên ta nhìn thấy là form thông tin chương trình ứng dụng gồm các thông tin về đề bài, tên giáo viên, tên người thực hiện và ngày tháng năm xây dựng chương trình .

Bước 2: Chuyển sang tab chương trình chính — ký và xác nhận để ký lên đề thi ta làm một số thao tác sau :

- Load file đề thi (*.doc) cần ký bằng cách nhấn vào nút —Browse. Nội dung đề thi được hiển thị trong textbox —Nội dung file.
- Nhấn nút tạo khóa hoặc load key từ file có sẵn trong máy tính
- Nhấn botton —Ký nhận để tạo chữ ký điện tử, chữ ký này được gắn vào cuối nội dung đề thi. Lưu lại đề thi này để gửi đi.

Bước 3: Gửi đi đề thi đã ký nhận và public key cho người nhận

Xác thực chữ ký: Người nhận sau khi nhận được đề thi đã ký, để xác thực cần làm các bước sau:

Bước 1: Mở chương trình và Load file đề thi đã ký nhận lên bằng nhấn botton – Browse.

Bước 2: Load public key mà đã nhận được từ người gửi.

Bước 3: Xác thực bằng cách nhấn botton - Kiểm tra. Nếu đề thi và chữ ký đúng của người gửi thì sẽ hiện thông báo nội dung đề thi không bị thay đổi và chữ ký chính xác là của người gửi.

3.4.3. Hình ảnh Demo chữ ký số áp dụng trong quản lý đề thi

Đề Thi:

ĐỀ SỐ 04	**ĐỀ KIỂM TRA HỌC KỲ 2** **MÔN: HÓA HỌC 9** Thời gian: 45 phút Trường THCS Lê Lợi

I/ Trắc nghiệm : Khoanh tròn chữ cái đầu câu đáp án đúng

Câu 1 :Dãy chất nào sau đây thuộc oxit bazơ

A/ CaO, SO_2, BaO; **B/ CaO, BaO, Na_2O;**

C/ CO_2, SiO_2; **D/ K_2O, Na_2O, SO_3**

Câu 2 :Trong các cặp chất sau, cặp chất nào đều được dung để điều chế oxi trong phòng thí nghiệm:

A/ $CuSO_4$, HgO; B/ K_2SO_4, $KMnO_4$;

C/ H_2O, $KClO_3$; D/ $KClO_3$, $KMnO_4$

Câu 3 : Dung dịch làm cho quỳ tím chuyển thành màu xanh:

A/ NaCl; B/ H_2SO_4 ; C/ NaOH ; D/ $CuSO_4$

Câu 4 : *25 gam dung dịch đường nồng độ 10% có chứa một lượng đường là:*

A/ 1,2g B/ 2,5g C/ 1,5g D/ 3,5g

Câu 5: Công thức hoá học của muối Canxidihidrophotphat là

A/ $Ca(H_2PO_4)_2$, B/ CaH_2PO_4,

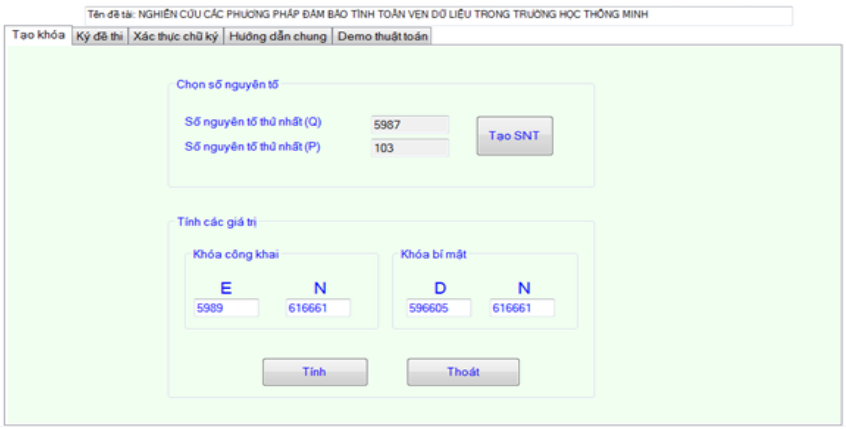
C/ $Ca(HPO_4)_2$, D/ Ca_2HPO_4

Câu 6: Ở 25°C cứ 204g đường hòa tan hết vào 100g nước tạo thành dd bão hòa . Độ tan của đường ở 25°C là:

A/ 408g ; B/ 204g ; C/ 100g ; D/ 102g.

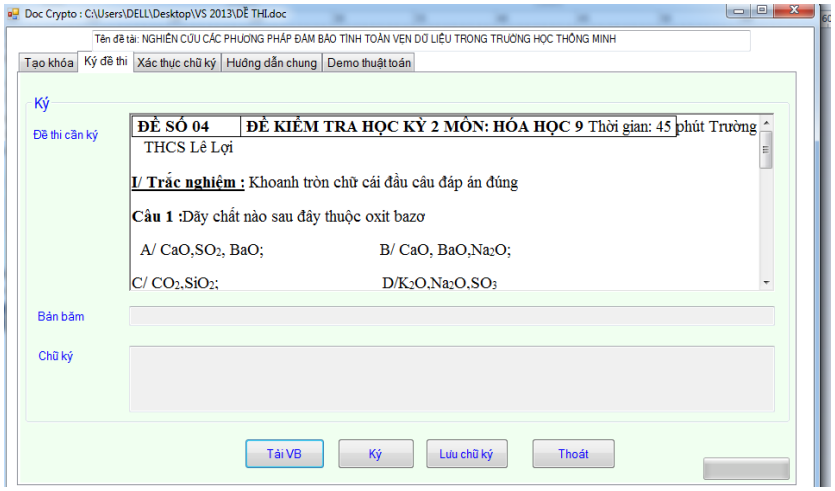
A. Tạo khóa K

- Chọn 2 số nguyên tố đủ lớn
- tính được khóa công khai và khóa bí mật

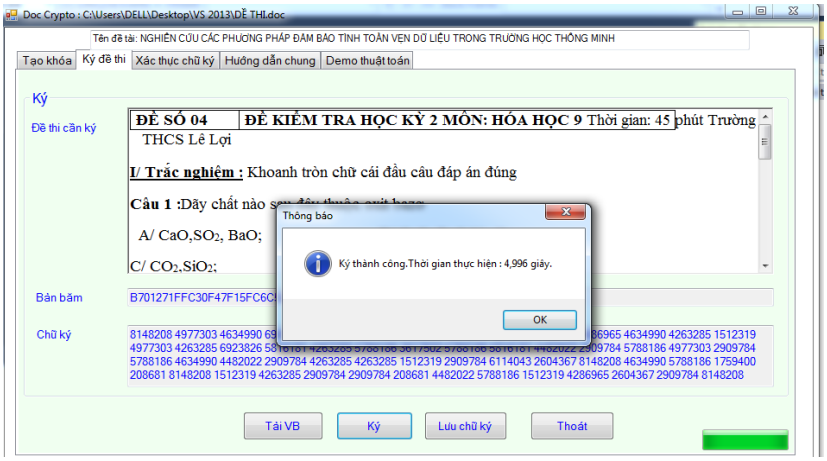


B. Thủ tục ký văn bản:

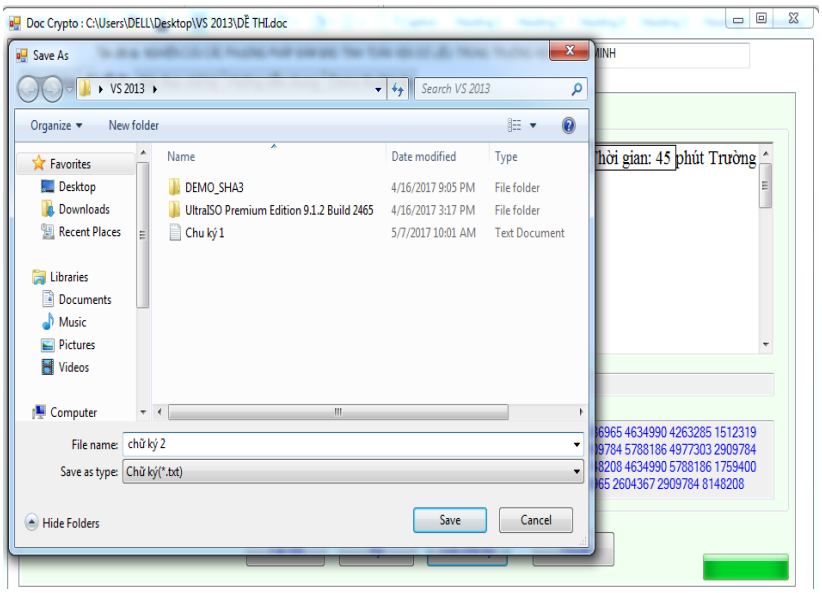
1. Tải văn bản



2. Ký



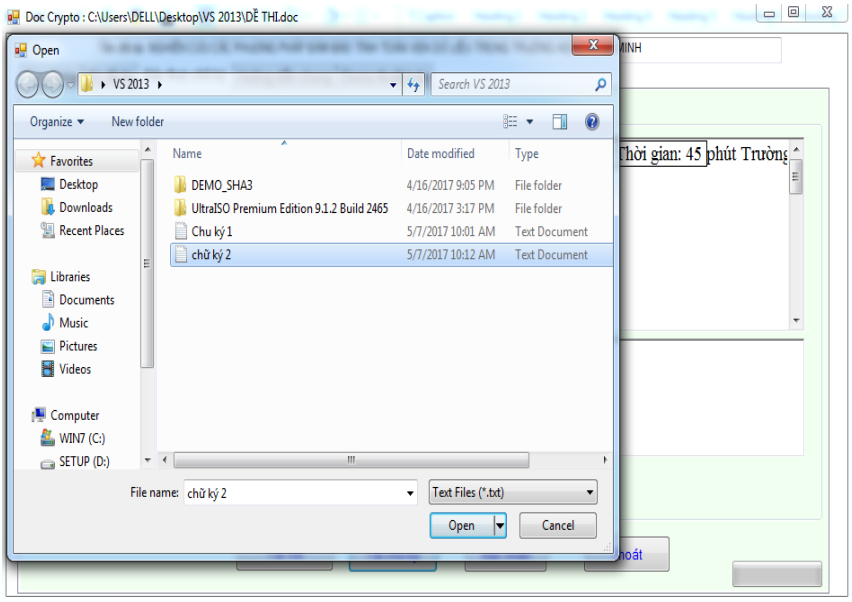
3. Lưu chữ ký.

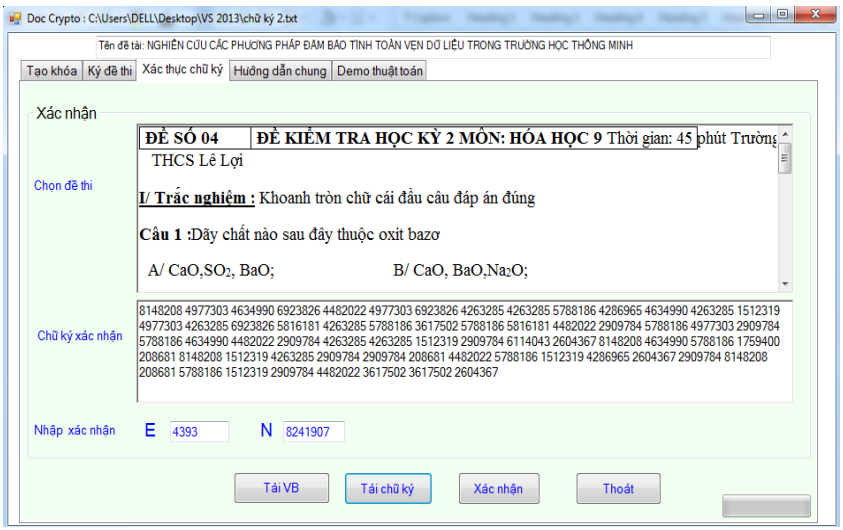


C. Xác thực chữ ký

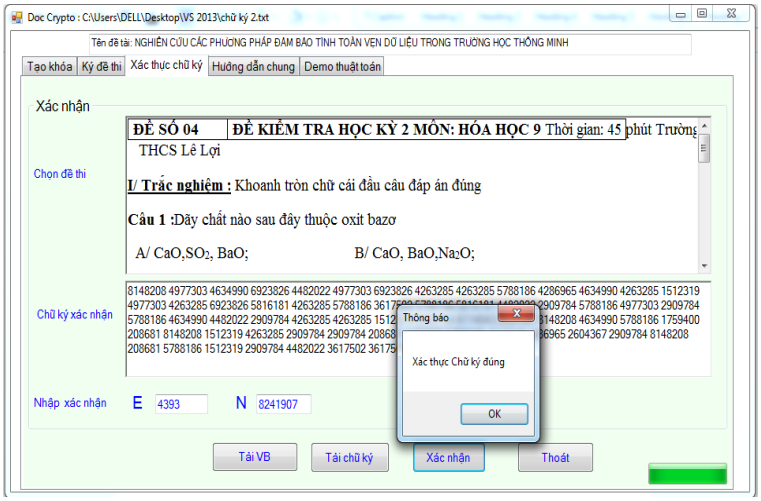
1. Tải văn bản

2. Tải chữ ký

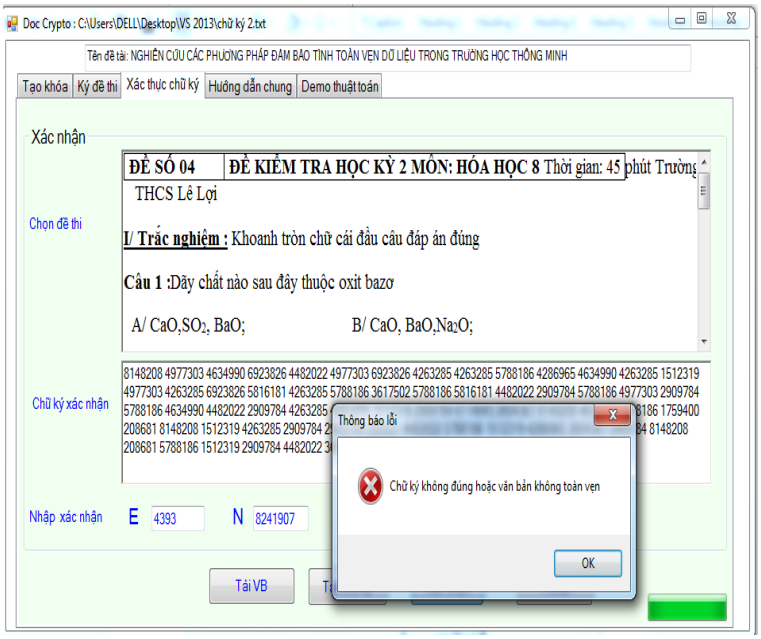




3. Xác nhận chữ ký



4. Báo lỗi khi chữ ký không đúng/ hoặc văn bản sai.



KẾT LUẬN

1. Các kết quả đạt được:

Để đảm bảo tính toàn vẹn dữ liệu là bài toán lớn trong trường học nói chung và trường học thông minh nói riêng. Việc sử dụng rộng rãi văn bản, tài liệu điện tử cùng với các đặc điểm của nó như: dễ dàng thay đổi nội dung thông tin trong tài liệu mà không để lại dấu vết, vấn đề xác định người gửi văn bản điện tử v.v... đã dẫn đến sự cần thiết phải tìm giải pháp cho các vấn đề trên. Luận văn nêu ra một số biện pháp để đảm bảo tính toàn vẹn của văn bản điện tử như: *Thứ nhất,* đảm bảo toàn vẹn dữ liệu bằng thuật toán băm SHA; *thứ hai,* đảm bảo toàn vẹn dữ liệu bằng một số phương pháp mã xác thực; *Thứ ba,* đảm bảo toàn vẹn dữ liệu bằng chữ ký số. Việc ứng dụng chữ ký số vào cơ quan trường học sẽ giúp quá trình luân chuyển văn bản được nhanh chóng, chính xác, kịp thời, không chối bỏ, quá trình xử lý và triển khai công việc không bị gián đoạn, giảm thiểu thời gian giải quyết công việc.

a. Lý thuyết:

- Để đảm bảo toàn vẹn dữ liệu cho các cơ quan nhà nước, trường học, các doanh nghiệp,... luận văn đã nghiên cứu các phương pháp toàn vẹn dữ liệu như:

+ Nghiên cứu các họ hàm băm mật mã SHA (SHA1,SHA2,SHA3). Trong đó nghiên cứu hàm băm SHA3 – Keccak do nhóm các nhà mật mã người Bỉ đứng đầu là Daemen (người đồng tác giả của thuật toán AES) thiết kế. Keccak có số vòng lặp là 18 vòng và kích thước trạng thái thay đổi lần lượt là 25, 50, 100, 200, 400, 800, 1600.

Nhóm thiết kế Keccak đã đưa vào cấu trúc Sponge, bên cạnh cấu trúc Sponge nhóm tác giả còn thực hiện các biện pháp như bổ sung khóa mật vào đầu vào của keccak biến nó thành mã xác thực thông báo.

SHA3 bao gồm bốn hàm băm SHA3 – 224, SHA3-256, SHA3-384, SHA3-512 và hai hàm mở rộng SHAKE128 và SHAKE256.

+Nghiên cứu các phương pháp đảm bảo tính toàn vẹn bằng mã xác thực. Có hai dạng chuẩn mà NIST đưa ra "mã xác thực thông điệp sử dụng hàm một chiều có khóa HMAC" và " mã xác thực thông điệp mã hóa". Ứng dụng MAC là đảm bảo tính xác thực giữa các bên trong kênh liên lạc có thể gửi và nhận thông điệp được xác thực với nhau và khả năng bị kẻ tấn công giả mạo là rất thấp.

+ Nghiên cứu về chữ ký số có nhiều loại sơ đồ chữ ký số khác nhau, trong luận văn này tôi sử dụng sơ đồ chữ ký thông dụng là RSA. Chữ ký số chức năng là xác thực được nguồn gốc tài liệu, tính toàn vẹn dữ liệu và tính chống từ chối bức thông điệp.

b. Thực nghiệm:

- Chương trình xây dựng gồm 2 modul chính: "demo thuật toán SHA3 và demo chữ ký số".

+ Demo các thuật toán băm SHA3 (SHA3-224, SHA3-256, SHA3-384, SHA3-512)

+ Demo chữ ký số lên đề thi.

2. Hướng nghiên cứu tiếp theo

Học viên sẽ phát triển ứng dụng các phương pháp mật mã vào các thuật toán mới và đảm bảo an toàn dữ liệu nói chung toàn vẹn dữ liệu nói riêng.

Tài liệu tham khảo

Tài liệu tiếng Việt

[1]. Phan Đình Diệu, "Lý thuyết mật mã và an toàn thông tin", Đại Học Quốc Gia Hà Nội, năm 2002.

[2]. Phạm Huy Điển, Hà Huy Khoái (2004), Mã hóa thông tin cơ sở toán học và ứng dụng, Viện toán học.

[3]. Trịnh Nhật Tiến (2009), "Bài giảng về mật mã và An toàn dữ liệu", Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.

[4]. TCVN 7635:2007, *Chữ ký số, kỹ thuật Mật mã*, 2007

[5]. Hàm băm an toàn và ứng dụng, *Luận văn ths Nguyễn Thanh Hưng, Đại học Quốc gia Hà Nội*

[6]. Giáo trình mã hóa và ứng dụng của nhóm tác giả TS. Dương Anh Đức – Ths Trần Minh Triết cùng cới nhóm SV, trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP Hồ Chí Minh.

Tài liệu tiếng Anh

[7] Design of SHA-3 Algorithm using Compression Box (3200 bit) for Digital Signature Applications

[8]. Secure Hash Algorithm-3(SHA-3) implementation on Xilinx FPGAs, Suitable for IoT Applications. Group of author Muzaffar Rao, Thomas Newe and Ian Grout University of Limerick, Ireland muhammad.rao @ ul.ie, thomas.newe @ ul.ie, Ian.grout @ ul.ie

[9] R.L. Rivest, A. Shamir, and L. Adleman, A Method for Obtaining Digital Signatures and Public-Key Cryptosystems, Communications of the ACM, 21 (2), trang 120-126, Feb 1978.

[10]. Keccak-reference-3.0. Guido Bertoni[1],Joan Daemen[1], Michael Peters[2], Gilles Van Assche[1].

[11]. Introduction to SHA-3 and Keccak, Joan Daemen STMicroelectronics and Radboud University **,**Crypto summer school 2015, Šibenik, Croatia, May 31 - June 5, 2015

[12]. N. F. Pub, "FIPS PUB 202. SHA-3 Standard: Permutation Based Hash and Extendable-Output Functions," *Federal Information Processing Standards Publication*, 2015.

[13] Introduction to Network Security Missouri S&T University CPE 5420 Data Integrity Algorithms.

[14]Design of SHA-3 Algorithm using Compression Box (3200 bit) for Digital Signature Applications

[15] Cryptography and Network Security, Fourth Edition – William Stallings

Nguồn Internet:

[16]. http://dx.doi.org/10.6028/NIST.FIPS.202.

[17] http://attt.vn