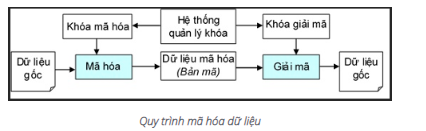
[Bài đọc] Giải thuật mã hóa công khai (RSA)

Tiêu chuẩn Rivest-Shamir-Adleman (RSA) - Giải thuật mã hóa công khai RSA là một tiêu chuẩn được các tác giả Ronal Rivest, Adi Shamir và Leonard Adleman phát triển tại Học viện Công nghệ Massachusetts (MIT) vào năm 1977, tên tiêu chuẩn được lấy từ 3 chữ cái đầu của tên 3 tác giả, hiện tiêu chuẩn được các tổ chức Viện Tiêu chuẩn Quốc gia Hoa Kỳ (American National Standards Institute – ANSI), Viện Kỹ nghệ Điện và Điện tử (Institute of Electrical and Electronics Engineers – IEEE) và Phòng thí nghiệm RSA công nhận (RSA Laboratories là một bộ phận của Tập đoàn EMC). Trước đó vài năm, Clifford Cox, một chuyên gia mã hoá người Anh đã phát triển riêng một biến thể của RSA. Tuy nhiên, Chính phủ Anh xem đây là vấn đề mật và đã không công bố. Khi Rivest, Shamir và Adleman công bố RSA trong ấn phẩm Scientific American tháng 9/1977, Cơ quan An ninh quốc gia Hoa Kỳ (NSA) không đồng ý về việc phổ biến rộng rãi RSA và ra lệnh cấm, tuy nhiên lệnh cấm này không có cơ sở pháp lý. Năm 1978, các tác giả đã công bố thuật toán trên Tạp chí của Hiệp hội Kỹ thuật Tính toán Hoa Kỳ (Communications of the Association for Computing Machinery - ACM). Hiện nay, có thể tham khảo [đặc tả của RSA](http://www.emc.com/collateral/white-papers/h11300-pkcs-1v2-2-rsa-cryptography-standard-wp.pdf) trên trang thông tin của Tập đoàn EMC.

Một số khái niệm

Mã hóa dữ liệu là sử dụng một phương pháp biến đổi dữ liệu từ dạng bình thường sang một dạng khác, mà một người không có thẩm quyền, không có phương tiện giải mã thì không thể đọc hiểu được.

Giải mã dữ liệu là quá trình ngược lại, là việc sử dụng một phương pháp biến đổi dữ liệu đã được mã hóa về dạng thông tin ban đầu. Có thể mô tả quy trình thực hiện mã hóa dữ liệu và giải mã dữ liệu như sau:



Sau đây là một số khái niệm và kí hiệu liên quan về vấn đề mã hóa dữ liệu :

- E (Encryption - Mã hóa): Là quá trình chuyển đổi dữ liệu gốc thành dữ liệu được mã hóa sao người khác không thể đọc hiểu.

-D (Decryption - Giải mã): Là quá trình ngược lại của mã hóa, biến đổi dữ liệu đã được mã hóa thành dạng gốc ban đầu.

- M (Message - Thông điệp), bản gốc hay bản rõ (Plaintext): Là tệp dữ liệu chưa được mã hóa hoặc đã được giải mã.

- C (Ciphertext - Bản mã): Tệp dữ liệu đã được mã hóa.

- K (Key - Khóa): Là dãy các bít 0, 1 thường được đưa ra dạng xâu ký tự, số...

- KE: Là khóa dùng để mã hóa.

- KD: Là khóa dùng để giải mã.

Theo quy ước, khi mã hóa thì C = E(M) và khi giải mã thì M = D(C) = D(E(M)).

Theo phương pháp truyền thống, người ta thường dùng cùng một khóa để mã hóa và giải mã. Lúc đó khóa phải được giữ bí mật tuyệt đối. Người ta gọi đây là hệ thống mã hóa cổ điển (các tên gọi khác: đối xứng, một khóa, khóa bí mật...). Một phương pháp khác, sử dụng khóa công khai (còn gọi là phương pháp mã hóa bất đối xứng, hay hệ thống hai khóa) trong đó khóa để mã hóa và khóa để giải mã là khác nhau. Các khóa này tạo thành một cặp chuyển đổi ngược nhau và không khóa nào có thể suy ra được từ khóa kia. Quy trình mã hóa khóa công khai gồm có các bước cơ bản như sau:

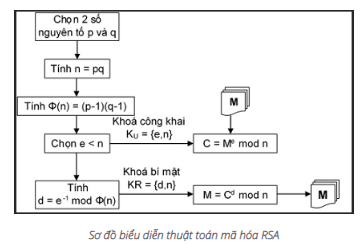
- Mỗi một hệ thống đầu cuối tạo một cặp khóa dùng cho quá trình mã hóa và giải mã mà hệ thống đó sẽ nhận.

- Mỗi hệ thống công bố khóa mã hóa của mình, gọi là khóa công khai, khóa còn lại gọi là khóa bí mật (khóa riêng) và phải được giữ an toàn.

- Nếu Bên gửi muốn gửi thông điệp cho Bên nhận, Bên gửi mã hóa thông điệp sử dụng khóa công khai của Bên nhận.

- Khi Bên nhận nhận thông điệp, Bên nhận giải mã bằng khóa riêng của Bên nhận.

Sơ đồ dưới đây minh họa các bước trong thuật toán RSA.



## Các đặc điểm chính trong thuật toán mã hóa RSA

Thuật toán RSA được thiết kế dựa trên độ khó của bài toán phân tích ra thừa số nguyên tố trên tập số nguyên *Zn*.

Cho số nguyên dương n = p \* q, với p, q là 2 số nguyên tố rất lớn (ít nhất 100 ký số). Khi biết *n*, muốn tìm *p, q*thì phải giải bài toán phân tích ra thừa số nguyên tố, công việc này đòi hỏi phải thực hiện một số lượng các phép tính vô cùng lớn.

### **Tạo khóa:**

- Tạo ngẫu nhiên 2 số nguyên tố *p, q* khác nhau và rất lớn (có số ký tự ít nhất là 100), sau đó tính: *n = p \* q*; *Ф(n) = (p -1) \* (q -1)*.

- Chọn ngẫu nhiên 1 số *e* sao cho *1 < e < Ф(n)*, với *e* là số nguyên tố  cùng nhau với *Ф(n)*.

- Tính số nghịch đảo *d* của *e* đối với *Ф(n)*: *1 < d < Ф(n), ed = 1(mod Ф(n))*.

Ở đây *d* là số mũ bí mật; *e* là số mũ công khai.

- Cặp *KU = {e, n}* được gọi là khoá công khai.

Cặp *KR = {d, n}* được gọi là khoá bí mật.

### **Mã hóa**

Khi Bên gửi muốn gửi thông điệp *M* cho Bên nhận với yêu cầu cần bảo mật thông tin. Bên gửi yêu cầu Bên nhận gửi khoá công khai *KU = {e, n}*. Bên gửi dùng khoá công khai *KU* này để mã hoá thông điệp *M* thành *C* theo công thức: *C = Me mod n*, sau đó Bên gửi gửi *C* cho Bên nhận.

### **Giải mã**

Để giải mã bản mã *C*, Bên nhận dùng khoá bí mật *KR = {d, n}* để có thể khôi phục lại dữ liệu gốc ban đầu do Bên gửi gửi đến thông qua phép toán *M = Cd mod n*.

## Các vấn đề quan tâm của RSA

Độ an toàn của RSA đã được các nhà khoa học trên thế giới trình bày trong nhiều nghiên cứu, sau đây sẽ xem xét một số các phương pháp tấn công RSA:

- Vét cạn khóa: cách tấn công này thử tất cả các khóa *d* có thể có để tìm ra bản giải mã có ý nghĩa, tương tự như cách thử khóa *K* của mã hóa đối xứng. Với *n* lớn, việc tấn công là rất khó thực hiện.

- Phân tích *n* thành các thừa số nguyên tố: Hiện nay, có nhiều thuật toán phân tích mới đã được đề xuất, cùng với tốc độ xử lý của máy tính ngày càng nhanh, đã làm cho việc phân tích*n* không còn quá khó khăn như trước đây. Năm 1977, các tác giả của RSA đã treo giải thưởng để tấn công RSA có kích thước của *n* vào khoảng 428 bít, tức 129 chữ số. Các tác giả này ước đoán phải mất 40 nghìn triệu triệu năm mới có thể  giải được. Tuy nhiên vào năm 1994, câu đố này đã được giải chỉ trong vòng 8 tháng. Một thiết bị lý thuyết có tên là TWIRL do Shamir và Tromer mô tả năm 2003 đã đặt ra câu hỏi về độ an toàn của khóa 1024 bít. Vì vậy, hiện nay người ta khuyến cáo sử dụng khóa có độ dài tối thiểu 2048 bít để bảo đảm an toàn. Để tránh gặp phải những vấn đề trên, RSA trên thực tế thường bao gồm một hình thức chuyển đổi ngẫu nhiên *M* trước khi mã hóa. Quá trình chuyển đổi này phải đảm bảo rằng *M* không rơi vào các giá trị không an toàn. Sau khi chuyển đổi, mỗi bản rõ khi mã hóa sẽ cho ra một trong số khả năng trong tập hợp bản mã. Điều này làm giảm tính khả thi của phương pháp tấn công lựa chọn bản rõ (một bản rõ sẽ có thể tương ứng với nhiều bản mã tuỳ thuộc vào cách chuyển đổi).

- Đo thời gian: Đây là một phương pháp phá mã không dựa vào mặt toán học của thuật toán RSA, mà dựa vào một “hiệu ứng lề” sinh ra bởi quá trình giải mã RSA. Hiệu  ứng lề đó là thời gian thực hiện giải mã. Giả sử  người phá mã có thể đo được thời gian giải mã *M = Cd mod n* dùng thuật toán bình phương liên tiếp. Trong thuật toán bình phương liên tiếp, nếu một bít của *d* là 1 thì xảy ra hai phép mô-đun, nếu bít đó là 0 thì chỉ có một phép mô-đun, do đó thời gian thực hiện giải mã là khác nhau. Bằng một số  phép thử bản rõ chọn trước, người phá mã có thể biết được các bít của *d* là 0 hay 1 và từ đó biết được *d*. Phương pháp phá mã này là một ví dụ cho thấy việc thiết kế một hệ mã an toàn rất phức tạp. Người thiết kế phải lường trước được hết các tình huống có thể xảy ra.

## Ứng dụng

Tiêu chuẩn RSA được ứng dụng rộng rãi trong nhiều lĩnh vực như chữ ký số, thương mại điện tử, bảo mật, xác thực…