*CHƯƠNG IV*

# Các hệ mật mã khoá công khai

## 4.1. Giới thiệu mở đầu.

### 4.1.1. Sự ra đời của mật mã khoá công khai.

Trong chương I ta đã giới thiệu qua định nghĩa của các khái niệm hệ mật mã khoá đối xứng và hệ mật mã khoá công khai. Sự ra đời của khái niệm hệ mật mã khoá công khai là một tiến bộ có tính chất bước ngoặt trong lịch sử mật mã nói chung, gắn liền với sự phát triển của khoa học tính toán hiện đại. Người ta có thể xem thời điểm khởi đầu của bước ngoặt đó là sự xuất hiện ý tưởng của W. Diffie và M.E. Hellman được trình bày vào tháng sáu năm 1976 tại Hội nghị quốc gia hàng năm của AFIPS (Hoa kỳ) trong bài *Multiuser cryptographic techniques*. Trong bài đó, cùng với ý tưởng chung, hai tác giả cũng đã đưa ra những thí dụ cụ thể để thực hiện ý tưởng đó, và mặc dù các thí dụ chưa có ý nghĩa thuyết phục ngay đối với tác giả, thì ý tưởng về các hệ mật mã khoá công khai cũng đã rất rõ ràng và có sức hấp dẫn đối với nhiều người. Và ngay sau đó, công việc tìm kiếm những thể hiện cụ thể có khả năng ứng dụng trong thực tế đã bắt đầu thu hút sự quan tâm của nhiều chuyên gia. Một năm sau, năm 1977, R.L. Rivest, A. Shamir và L.M. Adleman đề xuất một hệ cụ thể về mật mã khoá công khai mà độ an toàn của hệ dựa vào bài toán khó "phân tích số nguyên thành thừa số nguyên tố", hệ này về sau trở thành một hệ nổi tiếng và mang tên là hệ RSA, được sử dụng rộng rãi trong thực tiễn bảo mật và an toàn thông tin. Cũng vào thời gian đó, M.O. Rabin cũng đề xuất một hệ mật mã khoá công khai dựa vào cùng bài toán số học khó nói trên. Liên tiếp sau đó, nhiều hệ mật mã khóa công khai được đề xuất, mà khá nổi tiếng và được quan tâm nhiều là các hệ: hệ McEliece được đưa ra năm 1978 dựa trên độ *NP*-khó của bài toán giải mã đối với các hệ mã cyclic tuyến tính, hệ Merkle-Hellman dựa trên tính *NP*- đầy đủ của bài toán xếp ba lô(knapsack problem), hệ mật mã nổi tiếng ElGamal dựa trên độ khó của bài toán lôgarit rời rạc, hệ này về sau được mở rộng để phát triển nhiều

hệ tương tự dựa trên độ khó của các bài toán tương tự lôgarit rời rạc trên các cấu trúc nhóm cyclic hữu hạn, nhóm các điểm nguyên trên đường cong eliptic, v.v... Để tăng độ bảo mật, hệ mật mã ElGamal còn dùng với tư cách đầu vào cho thuật toán lập mật mã của mình, ngoài khoá công khai và bản rõ, một yếu tố ngẫu nhiên được chọn tuỳ ý, điều đó làm cho hệ mật mã trở thành một hệ mật mã xác suất khoá công khai. Một số hệ mật mã xác suất khoá công khai cũng được phát triển sau đó bởi Goldwasser-Micali và Blum-Goldwasser. Tất cả các hệ mật mã khoá công khai kể trên sẽ được trình bày trong chương này cùng với một số tính chất liên quan của chúng.

### 4.1.2. Một số bài toán cơ bản.

Sau đây ta sẽ nhắc lại một số bài toán số học được sử dụng đến khi xây dựng các hệ mật mã khoá công khai như nói ở trên. Các bài toán này phần lớn đã được trình bày trong chương II, một số được phát triển thêm cho các ứng dụng trực tiếp khi xây dựng các hệ mã cụ thể, ta liệt kê dưới đây một lần để thuận tiện cho các chỉ dẫn về sau.

*Bài toán phân tích số nguyên (thành thừa số nguyên tố)*:

Cho số nguyên dương $n$, tìm tất cả các ước số nguyên tố của nó, hay là tìm dạng phân tích chính tắc của $n = p_1^{\alpha_1}.p_2^{\alpha_2}...p_k^{\alpha_k}$, trong đó $p_i$ là các số nguyên tố từng cặp khác nhau và các $\alpha_i \geq 1$.

Bài toán này có liên hệ mật thiết với các bài toán thử tính nguyên tố hay thử tính hợp số của một số nguyên, nhưng với những gì mà ta biết đến nay, nó dường như khó hơn nhiều so với hai bài toán thử tính nguyên tố và tính hợp số.

Trong lý thuyết mật mã, bài toán này thường được sử dụng với các dữ liệu $n$ là số nguyên Blum, tức các số nguyên dương có dạng tích của hai số nguyên tố lớn nào đó.

*Bài toán RSA (Rivest-Shamir-Adleman)* :

Cho số nguyên dương $n$ là tích của hai số nguyên tố lẻ khác nhau, một số nguyên dương $e$ sao cho gcd($e$,$\phi$ ($n$)) =1, và một số nguyên $c$; tìm một số nguyên $m$ sao cho $m^e \equiv c \pmod n$.

Điều kiện gcd($e$,$\phi$ ($n$)) =1 bảo đảm cho việc với mỗi số nguyên $c \in \{0,1,...,n-1\}$ có đúng một số $m \in \{0,1,...,n-1\}$ sao cho $m^e \equiv c \pmod n$.

Dễ thấy rằng nếu biết hai thừa số nguyên tố của $n$, tức là biết $n = p.q$ thì sẽ biết $\phi(n) = (p-1)(q-1)$, và từ đó, do gcd($e$,$\phi$ ($n$)) =1 sẽ

tìm được $d = e^{-1} \bmod \phi(n)$, và do đó sẽ tìm được $m = c^d \bmod n$. Như vậy, bài toán RSA có thể qui dẫn trong thời gian đa thức về bài toán phân tích số nguyên. Tuy rằng cho đến nay chưa có một chứng minh nào cho việc qui dẫn ngược lại nhưng nhiều người vẫn tin rằng hai bài toán đó là tương đương với nhau về độ phức tạp tính toán.

*Bài toán thặng dư bậc hai* :

Cho một số nguyên lẻ $n$ là hợp số, và một số nguyên $a \in J_n$, tập tất cả các số $a$ có ký hiệu Jacobi $\left(\frac{a}{n}\right)=1$. Hãy quyết định xem $a$ có là thặng dư bậc hai theo mod$n$ hay không?

Trong lý thuyết mật mã, bài toán này cũng thường được xét với trường hợp $n$ là số nguyên Blum, tức $n$ là tích của hai số nguyên tố $p$ và $q$, $n = p.q$. Ta chú ý rằng trong trường hợp này, nếu $a \in J_n$, thì $a$ là thặng dư bậc hai theo mod$n$ khi và chỉ khi $\left(\frac{a}{p}\right)=1$, điều kiện này có thể thử được dễ dàng vì nó tương đương với điều kiện $a^{(p-1)/2} \equiv 1 \pmod{p}$. Như vậy, trong trường hợp này, bài toán thặng dư bậc hai có thể qui dẫn trong thời gian đa thức về bài toán phân tích số nguyên. Mặt khác, nếu không biết cách phân tích $n$ thành thừa số nguyên tố thì cho đến nay, không có cách nào giải được bài toán thặng dư bậc hai trong thời gian đa thức. Điều đó củng cố thêm niềm tin rằng bài toán thặng dư bậc hai và bài toán phân tích số nguyên là có độ khó tương đương nhau.

*Bài toán tìm căn bậc hai mod$n$* :

Cho một số nguyên lẻ $n$ là hợp số Blum, và một số $a \in Q_n$, tức $a$ là một thặng dư bậc hai theo mod$n$ . Hãy tìm một căn bậc hai của $a$ theo mod$n$, tức tìm $x$ sao cho $x^2 \equiv a \pmod{n}$.

Nếu biết phân tích $n$ thành thừa số nguyên tố, $n = p.q$ , thì bằng cách giải các phương trình $x^2 \equiv a$ theo các mod$p$ và mod$q$, rồi sau đó kết hợp các nghiệm của chúng lại theo định lý số dư Trung quốc ta sẽ được nghiệm theo mod$n$ , tức là căn bậc hai của $a$ theo mod$n$ cần tìm. Vì mỗi phương trình $x^2 \equiv a$ theo mod$p$ và mod$q$ có hai nghiệm (tương ứng theo mod$p$ và mod$q$ ), nên kết hợp lại ta được bốn nghiệm, tức bốn căn bậc hai của $a$ theo mod$n$. Người ta đã tìm được một số thuật toán tương đối đơn giản (trong thời gian đa thức) giải phương trình $x^2 \equiv a \pmod{p}$ với $p$ là số nguyên tố.

Như vậy, bài toán tìm căn bậc hai mod$n$ có thể qui dẫn trong thời gian đa thức về bài toán phân tích số nguyên. Ngược lại, nếu có thuật toán $\triangleq$ giải bài toán tìm căn bậc hai mod$n$ thì cũng có thể xây dựng một thuật toán giải bài toán phân tích số nguyên như sau: Chọn ngẫu nhiên một số $x$ với gcd($x$,$n$) =1, và tính $a = x^2 \bmod n$. Dùng thuật toán $\triangleq$ cho $a$ để tìm một căn bậc hai mod$n$ của $a$. Gọi căn bậc hai tìm được đó là $y$. Nếu $y \equiv \pm x \pmod n$, thì phép thử coi như thất bại, và ta phải chọn tiếp một số $x$ khác. còn nếu $y \not\equiv \pm x \pmod n$, thì gcd($x$-$y$, $n$) chắc chắn là một ước số không tầm thường của $n$, cụ thể là $p$ hay là $q$. Vì $n$ có 4 căn bậc hai mod$n$ nên xác suất của thành công ở mỗi lần thử là 1/2, và do đó số trung bình (kỳ vọng toán học) các phép thử để thu được một thừa số $p$ hay $q$ của $n$ là 2, từ đó ta thu được một thuật toán giải bài toán phân tích số nguyên (Blum) với thời gian trung bình đa thức. Tóm lại, theo một nghĩa không chặt chẽ lắm, ta có thể xem hai bài toán phân tích số nguyên và tìm căn bậc hai mod$n$ là khó tương đương nhau.

*Bài toán lôgarit rời rạc* :

Cho số nguyên tố $p$, một phần tử nguyên thuỷ $\alpha$ theo mod$p$ (hay $\alpha$ là phần tử nguyên thuỷ của $Z_p^*$), và một phần tử $\beta \in Z_p^*$.Tìm số nguyên $x$ ($0 \le x \le p-2$) sao cho $\alpha^x \equiv \beta \pmod p$.

Trong mục 2.4.3 ta đã giới thiệu qua bài toán này, và biết rằng trong trường hợp chung, cho đến nay chưa có một thuật toán nào giải bài toán này trong thời gian đa thức.

Bài toán này cũng được suy rộng cho các nhóm cyclic hữu hạn như sau:

*Bài toán lôgarit rời rạc suy rộng* :

Cho một nhóm cyclic hữu hạn $G$ cấp $n$, một phần tử sinh (nguyên thuỷ) $\alpha$ của $G$, và một phần tử $\beta \in G$. Tìm số nguyên $x$ ($0 \le x \le n-1$) sao cho $\alpha^x = \beta$.

Các nhóm được quan tâm nhiều nhất trong lý thuyết mật mã là: nhóm nhân của trường hữu hạn $GF(p)$ - đẳng cấu với nhóm $Z_p^*$ của trường $Z_p$ ,nhóm nhân $\mathbb{F}_{2^m}^*$ của trường hữu hạn $GF(2^m)$, nhóm nhân $Z_n^* = \{a : 0 \le a \le n-1, \gcd(a,n) = 1\}$ của trường $Z_n$ với $n$ là hợp số, nhóm gồm các điểm trên một đường cong elliptic xác định trên một trường hữu hạn, v.v...

*Bài toán Diffie-Hellman* :

Cho số nguyên tố $p$, một phần tử nguyên thuỷ $\alpha$ theo mod$p$ (tức phần tử sinh của $Z_p^*$), và các phần tử $\alpha^a \bmod p$ và $\alpha^b \bmod p$.

Hãy tìm giá trị $\alpha^{ab} \bmod p$.

Có thể chứng minh được rằng bài toán Diffie-Hellman qui dẫn được về bài toán lôgarit rời rạc trong thời gian đa thức. Thực vậy, giả sử có thuật toán $\triangleq$ giải bài toán lôgarit rời rạc. Khi đó, cho một bộ dữ liệu vào của bài toán Diffie-Hellman gồm $p$, $\alpha$, $\alpha^a \bmod p$ và $\alpha^b \bmod p$; trước hết dùng thuật toán $\triangleq$ cho ($p$, $\alpha$, $\alpha^a \bmod p$) ta tìm được $a$, và sau đó tính được $\alpha^{ab} \bmod p = (\alpha^b)^a \bmod p$. Người ta cũng chứng minh được hai bài toán lôgarit rời rạc và Diffie-Hellman là tương đương về mặt tính toán trong một số trường hợp, ví dụ $p$-1 là $B$-mịn với $B = O((\ln p)^c)$, $c$ là hằng số.

Tương tự như với bài toán lôgarit rời rạc, ta cũng có thể định nghĩa các bài toán Diffie-Hellman suy rộng cho các nhóm cyclic hữu hạn khác.

*Bài toán tổng tập con (hay bài toán KNAPSACK)*:

Cho một tập các số nguyên dương $\{a_1, a_2, ..., a_n\}$ và một số nguyên dương $s$. Hãy xác định xem có hay không một tập con các $a_j$ mà tổng của chúng bằng $s$. Một cách tương đương, hãy xác định xem có hay không các $x_i \in \{0,1\}$ $(1 \le i \le n)$ sao cho $\sum_{i=1}^{n} a_i x_i = s$.

Bài toán này là một bài toán $\mathcal{NP}$- đầy đủ, tức là thuộc lớp những bài toán khó mà cho đến nay chưa tìm được thuật toán giải chúng trong thời gian đa thức !

*Bài toán giải mã đối với mã tuyến tính*:

Mã tuyến tính là một lớp mã truyền tin có tính chất tự sửa sai được sử dụng trong kỹ thuật truyền tin số hoá. Không đi vào chi tiết của lớp mã này, ta có thể phát biểu trực tiếp bài toán giải mã đối với mã tuyến tính như sau:

Cho một ma trận cấp $n \times m$ A=($a_{ij}$) gồm các thành phần là 0 hoặc 1, một vectơ $\boldsymbol{y}$=($y_1, y_2, ..., y_m$) các giá trị 0 và 1, và một số nguyên dương $K$. Hỏi: có hay không một vectơ $\boldsymbol{x}$=($x_1, x_2, ..., x_n$) gồm các số 0 hoặc 1 và có không nhiều hơn $K$ số 1 sao cho với mọi $j$ $(1 \le j \le m)$:

$$\sum_{i=1}^{n} x_i . a_{ij} \equiv y_j \pmod 2 \ ?$$

Chú ý rằng ở đây, $\boldsymbol{x}$ là vectơ thông tin, và $\boldsymbol{y}$ là vectơ mã, phép giải mã là tìm lại $\boldsymbol{x}$ khi nhận được $\boldsymbol{y}$, bài toán này tiếc thay lại là một bài toán khó; Berlekamp, McEliece và Tilborg năm 1978 đã chứng minh nó thuộc lớp các bài toán $\mathcal{NP}$- đầy đủ !

## 4.2. Hệ mật mã khoá công khai RSA.

### 4.2.1. Mô tả hệ mật mã RSA.

Sơ đồ chung của hệ mật mã khoá công khai được cho bởi

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

trong đó $\mathcal{P}$ là tập ký tự bản rõ, $\mathcal{C}$ là tập ký tự bản mã, $\mathcal{K}$ là tập các khoá $K$, mỗi khoá $K$ gồm có hai phần $K=(K',K'')$, $K'$ là khoá công khai dành cho việc lập mật mã, còn $K''$ là khoá bí mật dành cho việc giải mã. Với mỗi ký tự bản rõ $x \in \mathcal{P}$, thuật toán lập mã $\mathcal{E}$ cho ta ký tự mã tương ứng $y = \mathcal{E}(K', x) \in \mathcal{C}$, và với ký tự mã $y$ thuật toán giải mã $\mathcal{D}$ sẽ cho ta lại ký tự bản rõ $x$: $\mathcal{D}(K'', y) = \mathcal{D}(K'', \mathcal{E}(K', x)) = x$.

Để xây dựng một hệ mật mã khoá công khai RSA, ta chọn trước một số nguyên $n = p.q$ là tích của hai số nguyên tố lớn, chọn một số $e$ sao cho $\gcd(e, \phi(n)) = 1$, và tính số $d$ sao cho

$$e.d \equiv 1 (\mathrm{mod}\, \phi(n)).$$

Mỗi cặp $K=(K',K'')$, với $K' = (n,e)$ và $K'' = d$ sẽ là một cặp khoá của một hệ mật mã RSA cụ thể cho một người tham gia.

Như vậy, <u>sơ đồ chung của hệ mật mã RSA</u> được định nghĩa bởi danh sách (1), trong đó:

$\mathcal{P} = \mathcal{C} = Z_n$, trong đó $n$ là một số nguyên Blum, tức là tích của hai số nguyên tố;

$\mathcal{K} = \{K=(K',K''): K' = (n,e)$ và $K'' = d$, $\gcd(e, \phi(n)) = 1$, $e.d \equiv 1 (\mathrm{mod}\, \phi(n))\}$;

$\mathcal{E}$ và $\mathcal{D}$ được xác định bởi:

$$\mathcal{E}(K', x) = x^e \bmod n, \text{ với mọi } x \in \mathcal{P},$$
$$\mathcal{D}(K'', y) = y^d \bmod n, \text{ với mọi } y \in \mathcal{C}.$$

Để chứng tỏ định nghĩa trên là hợp thức, ta phải chứng minh rằng với mọi cặp khoá $K=(K',K'')$, và mọi $x \in \mathcal{P}$, ta đều có

$$\mathcal{D}(K'', \mathcal{E}(K', x)) = x .$$

Thực vậy, do $e.d \equiv 1 (\mathrm{mod}\, \phi(n))$ ta có thể viết $e.d = t.\phi(n) + 1$. Nếu $x$ nguyên tố với $n$, thì dùng định lý Euler (xem 2.1.3) ta có

$$\mathcal{D}(K'', \mathcal{E}(K', x)) = x^{ed} \equiv x^{t\phi(n)+1} \equiv x^{t\phi(n)}.x \,(\mathrm{mod}\, n) = x.$$

Nếu $x$ không nguyên tố với $n$, thì do $n = p.q$, hoặc $x$ chia hết cho $p$ và nguyên tố với $q$, hoặc $x$ chia hết cho $q$ và nguyên tố với $p$, và $\phi(n) = (p-1).(q-1)$, trong cả hai trường hợp ta đều có

$$x^{t\phi(n)+1} \equiv x \,(\mathrm{mod}\, p),$$
$$x^{t\phi(n)+1} \equiv x \,(\mathrm{mod}\, q);$$

từ đó suy ra $x^{t\phi(n)+1} \equiv x (\mathrm{mod}\, n)$, tức $\mathcal{D}(K'', \mathcal{E}(K', x)) = x$.
Thí dụ: Giả sử chọn $n = p.q$ = 2357.2551 = 6012707, ta sẽ có $\phi(n)$ = $(p-1).(q-1)$=2356.2550 = 6007800. Chọn $e$ = 3674911, và tính được $d$ = 422191 sao cho $e.d \equiv 1 (\mathrm{mod}\, \phi(n))$. Một người dùng A có thể chọn khoá công khai là $K'$=($n$ =6012707, $e$ = 3674911) và giữ khoá bí mật $K''$=$d$ =422191. Một đối tác B muốn gửi cho A một thông báo $x$ =5234673, sẽ dùng khoá công khai để tạo bản mật mã $y = x^e$ = $5234673^{3674911} \mathrm{mod} 6012707$ = 3650502. A nhận được $y$, giải mã sẽ được bản rõ $x = 3650502^{422191} \mathrm{mod}\ 6012707$ =5234673.

## 4.2.2. Thực hiện hệ mật mã RSA.

Để thực hiện hệ mật mã RSA cho một mạng truyền tin bảo mật, ngoài việc xây dựng các chương trình tính toán hàm $\mathcal{E}$ (với tham biến đầu vào là $n$ ,$e$ và $x$) và hàm $\mathcal{D}$ (với tham biến đầu vào là $n$ ,$d$ và $y$), ta còn phải chọn cho mỗi người tham gia một bộ ($n$,$e$,$d$) để tạo các khoá công khai $K'$ và khoá bí mật $K''$. Hệ mã của mỗi người tham gia chỉ có khả năng bảo mật khi $n = p.q$ là số nguyên rất lớn (và do đó $p$,$q$ cũng phải là những số nguyên tố rất lớn); rất lớn có nghĩa là $p$,$q$ phải có biểu diễn thập phân cỡ hơn 100 chữ số, do đó $n$ có cỡ hơn 200 chữ số thập phân, hay $n \geq 10^{200}$!

Tính toán các số $e$,$d$ , hay thực hiện các hàm $\mathcal{E}$, $\mathcal{D}$, đều chủ yếu là thực hiện các phép tính số học trên các số nguyên rất lớn; về vấn đề này trong mấy chục năm qua, khoa lập trình máy tính đã đề xuất nhiều chương trình máy tính làm việc rất có hiệu quả, ta có thể tham khảo để sử dụng khi thực thi các hệ mật mã RSA cũng như nhiều hệ mật mã khác.

## 4.2.3. Tính bảo mật của mật mã RSA.

Bài toán thám mã (khi chỉ biết bản mã) đối với mật mã RSA là: biết khoá công khai $K'$=($n$,$e$), biết bản mã $y = x^e \mathrm{mod} n$, tìm $x$. Bài toán này chính là bài toán RSA được trình bày trong mục 4.1.2. Trong mục đó ta đã chứng tỏ rằng nếu biết hai thừa số $p$,$q$ của $n$ thì dễ tìm được $x$ từ $y$, và nói chung có bằng chứng để coi rằng bài toán RSA (hay bài toán thám mã RSA) là có độ khó tương đương với bài toán phân tích số nguyên (Blum) thành thừa số nguyên tố. Do đó, giữ tuyệt mật khoá bí mật $d$ , hay giữ tuyệt mật các thừa số $p$,$q$ , là có ý nghĩa rất quyết định đến việc bảo vệ tính an toàn của hệ mật mã RSA.

Một mạng truyền tin bảo mật sử dụng sơ đồ các hệ mật mã RSA được xem là an toàn, nếu tuân thủ các điều kiện cơ bản: mỗi

người tham gia phải độc lập lựa chọn các tham số $n, e,d$ của riêng mình, chọn $n$ cũng có nghĩa là chọn các thừa số $p,q$ của $n$ ($n$ =$p.q$), và do có $p,q$ nên tính được $\phi$ ($n$) = ($p$ -1).($q$ -1), và từ đó tìm được $e,d$ tương đối dễ dàng; nhưng cũng chính vì vậy mà sau khi đã chọn thì mỗi người tham gia phải giữ tuyệt đối bí mật các giá trị $p,q,d$ , chỉ công bố khoá công khai ($n,e$) mà thôi.

Tuy nhiên, đó là điều kiện chung, còn trong thực tế vẫn có thể còn nhiều sơ hở mà người thám mã có thể lợi dụng để tấn công vào tính bảo mật của các hệ mã RSA khó mà lường trước hết được; sau đây là một số trường hợp đơn giản đã biết mà ta cần chú ý:

1.*Dùng môđuyn n chung*. Giả sử có hai người tham gia A và B cùng sử dụng một môđuyn chung $n$ trong khoá công khai của mình, chẳng hạn A chọn khoá công khai ($n,e$) và giữ khoá bí mật $d$, B chọn khoá công khai ($n,a$) và giữ khoá bí mật $b$. Một người tham gia thứ ba C gửi một văn bản cần bảo mật $x$ đến cả A và B thì dùng các khoá công khai nói trên để gửi đến A bản mật mã $y = x^e \bmod n$ và gửi đến B bản mật mã $z = x^a \bmod n$. Ta sẽ chứng tỏ rằng một người thám mã O có thể dựa vào những thông tin $n,e,a,y,z$ trên đường công khai mà phát hiện ra bản rõ $x$ như sau:

a. Tính $c = e^{-1} \bmod a$,
b. Sau đó tính $h = (ce - 1)/a$ ,
c. Và ta được $x = y^c (z^h)^{-1} \bmod n$.

Thực vậy, theo định nghĩa trên, $ce$ -1 chia hết cho $a$, và tiếp theo ta có: $y^c (z^h)^{-1} \bmod n = x^{ec} . (x^{a(ce-1)/a})^{-1} \bmod n = x^{ce} . (x^{ce-1})^{-1} \bmod n = x$ . Như vậy, trong trường hợp này việc truyền tin bảo mật không còn an toàn nữa. Vì vậy, ta cần nhớ khi dùng các hệ RSA để tổ chức mạng truyền tin bảo mật, cần tránh dùng môđuyn $n$ chung cho các người tham gia khác nhau!

2. *Dùng số mũ lập mã e bé.* Để cho việc tính toán hàm lập mã được hiệu quả, ta dễ có xu hướng chọn số mũ $e$ của hàm lập mã là một số nguyên bé, chẳng hạn $e$ =3. Tuy nhiên, nếu trong một mạng truyền tin bảo mật dùng các hệ mật mã RSA, nếu có nhiều người cùng chọn số mũ lập mã $e$ bé giống nhau thì sẽ có nguy cơ bị tấn công bởi việc thám mã như sau : Giả sử có ba người tham gia chọn ba khoá công khai là ($n_1$, $e$), ($n_2$, $e$), ($n_3$, $e$) với cùng số mũ $e$ =3. Một người tham gia A muốn gửi một thông báo $x$ cho cả ba người đó, và để bảo mật, gửi bản mã $c_i = x^3 \bmod n_i$ cho người thứ $i$. Ba môđuyn $n_i$ là khác nhau, và có phần chắc là từng cặp nguyên tố với nhau. Một người thám mã có thể dùng định lý số dư Trung quốc để tìm một số $m$ ($0 \le m \le n_1 n_2 n_3$) thoả mãn

$$\begin{cases} m \equiv c_1 \bmod n_1 \\ m \equiv c_2 \bmod n_2 \\ m \equiv c_3 \bmod n_3 \end{cases}$$

Vì $x \le n_i$, nên $x^3 \le n_1 n_2 n_3$, do đó ắt có $m = x^3$. Vậy là ta đã đưa được bài toán tìm căn bậc ba theo nghĩa đồng dư mod$n_i$ về bài toán tìm căn bậc ba theo nghĩa số học thông thường: tìm căn bậc ba của $m$ ta được $x$, tức được bản rõ!

Với những lý do khác, người ta đã có những bằng chứng để chứng tỏ rằng hệ RSA cũng không bảo đảm an toàn nếu ta dùng các khoá có số mũ giải mã $d$ là số nguyên bé, dù rằng khi đó thuật toán giải mã có làm việc hiệu quả hơn. Vì thế, khi sử dụng các hệ mật mã RSA, để bảo đảm an toàn ta nên chọn các số mũ $e$ và $d$ là những số nguyên lớn, có kích cỡ lớn gần như bản thân số $n$.

3. *Lợi dụng tính nhân của hàm lập mã*. Ta chú ý rằng hàm lập mã $f(x) = x^e \bmod n$ có tính nhân (multiplicative property), nghĩa là $f(x.y) = f(x).f(y)$. Dựa vào tính chất đó, ta thấy rằng nếu $c$ là mật mã của bản rõ $x$, thì $\overline{c} = c.u^e \bmod n$ sẽ là mật mã của bản rõ $xu$. Do đó, khi lấy được bản mật mã $c$, để phát hiện bản rõ $x$ người thám mã có thể chọn ngẫu nhiên một số $u$ rồi tạo ra bản mã $\overline{c}$, và nếu người thám mã có khả năng thám mã theo kiểu « có bản mã được chọn » (xem 1.5.1), tức có khả năng với $\overline{c}$ được chọn tìm ra bản rõ tương ứng là $\overline{x} = xu$, thì bản rõ gốc cần phát hiện sẽ là $x = \overline{x}.u^{-1} \bmod n$. Tất nhiên, khả năng người thám mã có năng lực giải quyết bài toán thám mã theo kiểu có bản mã được chọn là rất hiếm, nhưng dầu sao đấy cũng là một trường hợp mà vấn đề bảo mật dễ bị tấn công, ta không thể không tính đến để tìm cách tránh!

4. *Tấn công bằng cách lặp phép mã*. Ta cũng chú ý rằng hàm lập mã $f(x) = x^e \bmod n$ là một phép hoán vị trên tập $Z_n = \{0,1,...,n-1\}$, do đó với mọi $c \in Z_n$ nếu ta thực hiện lặp phép lập mã để được

$$c_0 = c, c_1 = c^e \bmod n, c_2 = c^{e^2} \bmod n, ..., c_i = c^{e^i} \bmod n, ...$$

ắt sẽ tìm được số $k \ge 1$ sao cho $c_k = c^{e^k} \bmod n = c$. Nếu $c$ là bản mã của một bản rõ $x$ nào đó, $c = x^e \bmod n$, thì người thám mã có thể xuất phát từ $c$ thực hiện lặp phép lập mã như trên sẽ tìm được số $k \ge 1$ bé nhất sao cho $c_k = c$. Và khi đó ta sẽ có số hạng trước đó $c_{k-1} = x$, là bản rõ cần phát hiện. Thuật toán về hình thức là khá đơn giản, nhưng hiệu quả thực hiện không đáng hy vọng lắm, vì số phép lặp cần thực hiện nói chung có thể là rất lớn, cỡ bằng số các phép hoán vị trên $Z_n$, tức là bằng $n!$, với số $n$ có khoảng 200 chữ số thập phân. Trên thực tế, phỏng theo thuật toán nói trên ta có thể dễ dàng có một thuật toán phân tích $n$ thành thừa số nguyên tố, mà một thuật

toán như vậy làm việc có hiệu quả thiết thực, như đã trình bày trong một phần trên, là chưa có! Vì vậy, nguy cơ bị thám mã bằng thuật toán đơn giản nói trên đối với tính an toàn của hệ mật mã RSA là không đáng ngại lắm.

5. *Về khả năng che giấu của bản mật mã*. Mật mã, sở dĩ nó giữ được bí mật, là do khả năng che giấu thông tin của nó, tức là biết bản mã $y$ khó lòng tìm được thông tin nào để phát hiện ra bản rõ $x$. Một cách thô thiển, ta nói bản rõ $x$ là *không che giấu được* qua phép lập mật mã RSA $e_K(x) = x^e \bmod n$, nếu $e_K(x) = x$. Nói cách khác, $x$ là không che giấu được nếu bản mã của $x$ cũng chính là $x$. Tiếc rằng với bất kỳ hệ mật mã RSA nào cũng có những bản rõ không che giấu được, đó là những bản rõ $x = -1, 0, 1 \bmod n$ (vì số mũ $e$ luôn luôn là số lẻ). Người ta chứng minh được rằng nếu $n = p.q$, thì số các bản rõ $x \in Z_n$ không che giấu được là bằng

$$(1+\gcd(e-1, p-1)).(1+\gcd(e-1, q-1)).$$

Vì $e-1$, $p-1$, $q-1$ là các số chẵn, nên số đó ít nhất là 9, nên mỗi hệ RSA có ít nhất 9 bản rõ không che giấu được. Tuy nhiên, thường $n$, và do đó cả $p$ và $q$, đều rất lớn, nên tỷ lệ các bản rõ không che giấu được nói chung là bé không đáng kể, và do đó khả năng gặp các bản rõ không che giấu được không tạo nên một nguy cơ đáng kể nào đối với việc dùng các hệ mật mã RSA.

## 4.3. Hệ mật mã khoá công khai Rabin.

### 4.3.1. Mô tả hệ mật mã Rabin.

Sơ đồ hệ mật mã khoá công khai Rabin được cho bởi

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

trong đó: $\mathcal{P} = \mathcal{C} = Z_n$, trong đó $n$ là một số nguyên Blum, $n = p.q$, với $p$ và $q$ là hai số nguyên tố có tính chất $p \equiv 3 (\text{mod} 4)$, $q \equiv 3 (\text{mod} 4)$,

$$\mathcal{K} = \{K = (K', K'') : K' = (n, B), K'' = (p, q), 0 \le B \le n-1\},$$

các thuật toán $\mathcal{E}$ và $\mathcal{D}$ được xác định bởi

$$\mathcal{E}(K', x) = x(x+B) \bmod n,$$

$$\mathcal{D}(K'', y) = \sqrt{\frac{B^2}{4} + y} - \frac{B}{2} \bmod n.$$

(ký hiệu căn bậc hai sẽ được giải thích sau).

Trong một mạng truyền tin bảo mật với sơ đồ mật mã Rabin, mỗi người tham gia chọn cho mình các yếu tố $n,B,p,q$ để lập nên khoá công khai và khoá bí mật của mình.

Ta chú ý rằng với mỗi bộ khoá $K$, các thuật toán $e_{K'} = \mathcal{E}(K',.)$ và $d_{K''} = \mathcal{D}(K'',.)$ không lập thành một cặp song ánh, cụ thể là $e_{K'}$ không phải là một đơn ánh, vì nếu $w$ là một căn bậc hai của 1 theo mod$n$ thì $e_{K'}(w(x+\frac{B}{2}) - \frac{B}{2}) = e_{K'}(x)$, mà ta có đến 4 căn bậc hai của 1 theo mod$n$ ,tức là ta có 4 giá trị khác nhau của đối số $x$ cho cùng một giá trị $e_{K'}(x)$.

Bây giờ nói đến thuật toán giải mã $d_{K''} = \mathcal{D}(K'',.)$. Đặt $C = B^2/4 + y$, ta có $d_{K''}(y) = \sqrt{C} - B/2 \bmod n$, do đó để có $d_{K''}(y)$, ta cần tính $\sqrt{C} \bmod n$, tức cần giải phương trình $z^2 \equiv C \bmod n$ . Phương trình đó tương đương với hệ thống gồm hai phương trình sau đây:

$$\begin{cases} z^2 \equiv C \bmod p, \\ z^2 \equiv C \bmod q. \end{cases} \qquad (2)$$

Vì $p$ và $q$ là các số nguyên tố nên ta có $C^{\frac{p-1}{2}} \equiv 1 \bmod p$, $C^{\frac{q-1}{2}} \equiv 1 \bmod q$. Theo giả thiết, $p \equiv 3 (\text{mod} 4)$ và $q \equiv 3 (\text{mod} 4)$, nên $\frac{p+1}{4}$ và $\frac{q+1}{4}$ là các số nguyên; và ta có

$$(\pm C^{\frac{p+1}{4}})^2 \equiv C (\text{mod}\, p),\ (\pm C^{\frac{q+1}{4}})^2 \equiv C (\text{mod}\, q).$$

Do đó,phương trình $z^2 \equiv C \bmod n$ , hay hệ phương trình (2), có 4 nghiệm theo mod$n$ , tương ứng với 4 hệ phương trình sau đây :

$$\begin{cases} z \equiv C^{(p+1)/4} (\text{mod}\, p) \\ z \equiv C^{(q+1)/4} (\text{mod}\, q) \end{cases} \quad \begin{cases} z \equiv C^{(p+1)/4} (\text{mod}\, p) \\ z \equiv -C^{(q+1)/4} (\text{mod}\, q) \end{cases}$$

$$\begin{cases} z \equiv -C^{(p+1)/4} (\text{mod}\, p) \\ z \equiv C^{(q+1)/4} (\text{mod}\, q) \end{cases} \quad \begin{cases} z \equiv -C^{(p+1)/4} (\text{mod}\, p) \\ z \equiv -C^{(q+1)/4} (\text{mod}\, q) \end{cases}$$

Cả 4 nghiệm của 4 hệ phương trình đó theo mod$n$ đều được viết chung dưới một ký hiệu là $\sqrt{C} \bmod n$, và vì vậy thuật toán giải mã $d_{K''}(y)$ thực tế sẽ cho ta 4 giá trị khác nhau theo mod$n$ mà bản rõ là một trong 4 giá trị đó. Việc chọn giá trị nào trong 4 giá trị tìm được làm bản rõ là tuỳ thuộc vào những đặc trưng khác của bản rõ mà người giải mã nhận biết (thí dụ bản rõ dưới dạng số phải có biểu diễn nhị phân là mã của một văn bản tiếng Anh thông thường).

Thí dụ : Giả sử $n$ =77 = 7.11, $B$ =9 (ở đây $p$ =7, $q$ =11). Ta có

$$e_{K'}(x) = x^2 + 9x \bmod 77,$$

$$d_{K''}(y) = \sqrt{1+y} - 43 \bmod 77,$$

vì $2^{-1}$=39mod77, $9.2^{-1}$ =9.39 =43mod77, $B^2$=4mod77, $B^2/4$ =1mod 77.

Với $x$ =44 ta có $e_{K'}(x)$ = $44^2$+9.44 =2332 =22mod77, bản mã tương ứng với $x$ là $y$ = 22. Bây giờ giải mã với bản mã $y$ =22, bằng thủ tục nói trên ta có thể tìm được 4 giá trị của $\sqrt{1+y} = \sqrt{1+22} = \sqrt{23}$ theo mod77 là 10,67,32,45, từ đó 4 giá trị có thể có của $d_{K''}(y)$ là

$$d_{K''}(y) = 44, 24, 66, 2.$$

Bản rõ nằm trong 4 giá trị đó, trong trường hợp này là 44.

### 4.3.2. Tính an toàn của hệ mật mã Rabin.

Trong định nghĩa của hệ mật mã Rabin, khoá công khai là ($n$,$B$), khoá bí mật là ($p$,$q$) tức là cặp thừa số nguyên tố của $n$ . Như vậy, tính an toàn của hệ mật mã nằm ở việc giữ bí mật các thừa số $p$ và $q$. Định nghĩa của phép giải mã cũng cho ta thấy rằng yếu tố có ý nghĩa quyết định trong phép giải mã là việc tính căn bậc hai của một số theo mod$n$. Trong mục 4.1.2 bài toán tìm căn bậc hai theo mod$n$ (với $n$ là hợp số Blum) đã được chứng tỏ là có độ khó tương đương với bài toán phâ n tích $n$ thành thừa số nguyên tố. Vì vậy, bài toán giải mã đối với hệ mật mã Rabin, cũng là bài toán giữ bí mật khoá bí mật ($p$,$q$), và bài toán phân tích số nguyên thành thừa số nguyên tố là có độ khó tương đương nhau. Và đó cũng là yếu tố bảo đảm tính an toàn của hệ mật mã Rabin !

## 4.4. Hệ mật mã khoá công khai ElGamal.

### 4.4.1. Mô tả hệ mật mã ElGamal.

Hệ mật mã ElGamal được T. ElGamal đề xuất năm 1985, dựa vào độ phức tạp của bài toán tính lôgarit rời rạc, và sau đó đã nhanh chóng được sử dụng rộng rãi không những trong vấn đề bảo mật truyền tin mà còn trong các vấn đề xác nhận và chữ ký điện tử.

Sơ đồ hệ mật mã khoá công khai ElGamal được cho bởi

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

trong đó: $\mathcal{P} = Z_p^*$, $\mathcal{C} = Z_p^* \times Z_p^*$, với $p$ là một số nguyên tố;

$$\mathcal{K} = \{K = (K', K'') : K' = (p, \alpha, \beta),\ K'' = a,\ \beta \equiv \alpha^a \bmod p\},$$

ở đây $\alpha$ là một phần tử nguyên thuỷ theo mod$p$, tức của $Z_p^*$.

Các thuật toán lập mã $e_{K'} = \mathcal{E}(K',.)$ và giải mã $d_{K''} = \mathcal{D}(K'',.)$ được xác định như sau: Với mỗi $x \in \mathcal{P} = Z_p^*$, để lập mật mã cho $x$ trước hết ta chọn thêm một số ngẫu nhiên $k \in Z_{p-1}$ rồi tính:

$$e_{K'}(x,k) = (y_1, y_2), \quad \text{với} \quad \begin{cases} y_1 = \alpha^k \bmod p, \\ y_2 = x.\beta^k \bmod p. \end{cases}$$

Với mọi số ngẫu nhiên $k$ bất kỳ, ta đều xem $e_{K'}(x,k)$ là mật mã của $x$. Và thuật toán giải mã được xác định bởi

$$d_{K''}(y_1, y_2) = y_2.(y_1^a)^{-1} \bmod p.$$

Các phép lập mật mã và giải mã được xác định như vậy là hợp thức, vì ta có với mọi $x \in \mathcal{P} = Z_p^*$ và mọi $k \in Z_{p-1}$:

$$d_{K''}(e_{K'}(x,k)) = x.\beta^k.(\alpha^{k.a})^{-1} \bmod p = x.\beta^k.\beta^{-k} \bmod p = x.$$

Ta chú ý rằng trong một mạng truyền thông bảo mật với việc dùng sơ đồ mật mã ElGamal, mỗi người tham gia tự chọn cho mình các tham số $p, \alpha, a,$ rồi tính $\beta$, sau đó lập và công bố khoá công khai $K' = (p, \alpha, \beta)$, nhưng phải giữ tuyệt mật khoá bí mật $K'' = a$. Bài toán biết khoá công khai tìm ra khoá bí mật chính là bài toán tính lôgarit rời rạc được kể đến trong mục 4.1.2, một bài toán khó cho đến nay chưa có một thuật toán nào làm việc trong thời gian đa thức giải được nó.

Thí dụ : Chọn $p$ = 2579, $\alpha$ =2, $a$ =765, ta tính được $\beta = 2^{765} = 949$ mod2579. Ta có khoá công khai (2579, 2, 949) và khoá bí mật 765. Giả sử để lập mật mã cho $x$ =1299, ta chọn ngẫu nhiên $k$ =853, sẽ có

$$e_{K'}(1299, 853) = (2^{853}, 1299.\ 949^{853}) \bmod 2579$$
$$= (453, 2396).$$

Và giải mã ta được lại

$$d_{K''}(453, 2396) = 2396.\ (453^{765})^{-1} \bmod 2579 = 1299.$$

## 4.4.2. Tính an toàn của hệ mật mã ElGamal.

Như đã trình bày ở trên, nếu ta xem tính an toàn của hệ mật mã ElGamal là ở việc giữ tuyệt mật khoá bí mật $K''$, thì ta có thể yên tâm vì bài toán phát hiện khoá bí mật có độ khó tương đương với bài toán tính lôgarit rời rạc, mà bài toán này thì như ở các mục 4.1.2 và 2.4.3 đã chứng tỏ, cho đến nay chưa có một thuật toán nào làm việc trong thời gian đa thức giải được nó. Có một điều cảnh báo là nên chú ý chọn môđuyn $p$ là số nguyên tố sao cho $p$ -1 có ít nhất một ước số nguyên tố lớn (xem 2.4.3). Điều đó là thực hiện được

nếu số nguuyên tố $p$ được chọn là số nguyên tố Sophie Germain (tức có dạng $2q+1$, với $q$ cũng là số nguyên tố lớn).

Ngoài ra, còn có khả năng khoá bí mật $K'' = a$ bị lộ do cẩu thả trong việc sử dụng số ngẫu nhiên $k$, đặc biệt là khi *để lộ số k được dùng*. Thực vậy, nếu để lộ số $k$, thì khoá bí mật $a$ được tính ra ngay theo công thức sau đây:

$$a = (x - ky_2)y_1^{-1} \bmod (p-1).$$

Như vậy,một người thám mã có khả năng tấn công theo kiểu "biết cả bản rõ" (xem 1.5.1) có thể phát hiện ra khoá $a$ nếu biết $k$.

Một trường hợp khác làm mất tính an toàn của hệ mật mã ElGamal là việc *dùng cùng một số k cho nhiều lần lập mật mã*. Thực vậy, giả sử dùng cùng một số ngẫu nhiên $k$ cho hai lần lập mã, một lần cho $x_1$, một lần cho $x_2$, và được các bản mã tương ứng $(y_1,y_2)$ và $(z_1,z_2)$. Vì cùng dùng một số $k$ nên $y_1=z_1$. Và do đó theo công thức lập mã ta có $z_2/y_2 = x_2/x_1$, tức là $x_2 = x_1.z_2/y_2$. Như vậy, một người thám mã, một lần "biết cả bản rõ" dễ dàng phát hiện được bản rõ trong các lần sau.

### 4.4.3. Các hệ mật mã tương tự ElGamal.

Hệ mật mã ElGamal được xây dựng dựa trên các yếu tố: một nhóm hữu hạn cyclic $(Z_p^*)$, một phần tử nguyên thuỷ $(\alpha \in Z_p^*)$ sao cho bài toán tính lôgarit rời rạc (tính $a = \log_\alpha \beta$, tức cho $\beta$ tìm $a$ sao cho $\beta = \alpha^a \bmod p$) là rất khó thực hiện. Vì vậy, nếu có đủ các yếu tố đó thì ta có thể xây dựng các hệ mật mã tương tự ElGamal. Như vậy, sơ đồ của một hệ mật mã tương tự ElGamal được cho bởi

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

trong đó: $\mathcal{P}=G$, $\mathcal{C}=G\times G$, với G là một nhóm cyclic hữu hạn;

$$\mathcal{K} = \{K = (K', K'') : K' = (G, \alpha, \beta),\ K'' = a,\ \beta = \alpha^a\},$$

ở đây $\alpha$ là một phần tử nguyên thuỷ của nhóm $G$.

Các thuật toán lập mã $e_{K'} = \mathcal{E}(K',.)$ và giải mã $d_{K''} = \mathcal{D}(K'',.)$ được xác định như sau: Với mỗi $x \in \mathcal{P}=G$, để lập mật mã cho $x$ trước hết ta chọn thêm một số ngẫu nhiên $k$ $(0 \le k \le |G|)$ rồi tính:

$$e_{K'}(x,k) = (y_1, y_2), \quad \text{với} \quad \begin{cases} y_1 = \alpha^k \\ y_2 = x.\beta^k \end{cases}$$

Với mọi số ngẫu nhiên $k$ bất kỳ, ta đều xem $e_{K'}(x,k)$ là mật mã của $x$. Và thuật toán giải mã được xác định bởi

$$d_{K''}(y_1, y_2) = y_2.(y_1^a)^{-1} \bmod p.$$

Phép nhân trong các biểu thức nói trên đều là phép nhân của $G$.

Có hai lớp nhóm thường được sử dụng để xây dựng các hệ mật mã tương tự ElGamal là nhóm nhân của trường Galois GF($p^n$) và nhóm cộng của một đường cong elliptic xác định trên một trường hữu hạn.

1. *Nhóm nhân của trường Galois GF($p^n$)*: Trường Galois GF($p^n$) là trường của các đa thức với hệ số trong $Z_p$ lấy theo môđuyn là một đa thức bậc $n$ bất khả qui; với phép cộng và phép nhân là phép cộng và phép nhân đa thức theo môđuyn đó. Trường có $p^n$ phần tử, có thể xem mỗi phần tử là một đa thức bậc $n$ -1 với hệ số thuộc $Z_p$ = {0,1,2,...,$p$ -1}, thậm chí là một vectơ $n$ chiều mà các thành phần là các hệ số của đa thức đó. Tập tất cả các đa thức khác 0 lập thành nhóm nhân của trường $GF(p^n)$,và người ta chứng minh được rằng nhóm nhân đó là cyclic.

Như vậy, nhóm $G = GF(p^n)\setminus\{0\}$ là nhóm cyclic cấp $p^n$-1. ta có thể chọn một phần tử nguyên thuỷ của nhóm đó, và thiết lập bài toán lôgarit rời rạc tương ứng, từ đó xây dựng được hệ mật mã tương tự ElGamal.

2. *Nhóm công của đường cong elliptic* : Giả sử $p$ là một số nguyên tố > 3. *Đường cong elliptic* $y^2=x^3+a.x+b$ *trên* $Z_p$ **,** trong đó $a,b \in Z_p$ là các hằng số thoả mãn $4a^3+27b^2 \not\equiv 0 \pmod p$, được định nghĩa là tập hợp tất cả các điểm $(x,y) \in Z_p \times Z_p$ thoả mãn phương trình

$$y^2 \equiv x^3+a.x+b \pmod p,$$

cùng với một phần tử đặc biệt mà ta ký hiệu là $\mathcal{O}$ . Tập hợp đó được ký hiệu là $E$. Trên tập $E$ ta xác định một phép cộng như sau : Giả sử $P=(x_1, y_1)$ và $Q = (x_2, y_2)$ là hai điểm của $E$. Nếu $x_1=x_2$ và $y_1= -y_2$ thì ta định nghĩa $P+Q=\mathcal{O}$ ; nếu không thì $P+Q = (x_3, y_3)$, trong đó

$$x_3 = \lambda^2-x_1-x_2 , \quad y_3 = \lambda (x_1-x_3) - y_1 ,$$

với

$$\lambda = \begin{cases} (y_2 - y_1)/(x_2 - x_1), khi\ P \neq Q; \\ (3x_1^2 + a)/2y_1, \qquad khi\ P = Q. \end{cases}$$

Ngoài ra, ta định nghĩa thêm : $P+\mathcal{O} = \mathcal{O}+P = P$.

Tập $E$ với phép toán cộng đó lập thành một nhóm. Nếu $|E| = q$ là số nguyên tố thì nhóm cộng đó là nhóm cyclic, và mọi phần tử khác không ($\neq\mathcal{O}$ ) đều là phần tử nguyên thuỷ. Ta nhớ rằng trong trường hợp này, phần tử nghịch đảo là phần tử đối, phép nâng lên luỹ thừa $n$ là phép nhân với số $n$ , phép lôgarit tương ứng với một kiểu phép chia. Ta có thể xuất phát từ nhóm $E$ này để xây dựng hệ mật mã tương tự ElGamal.

## 4.5. Các hệ mật mã dựa trên các bài toán $\mathcal{NP}$-đầy đủ.

### 4.5.1. Nguyên tắc chung.

Như đã giới thiệu trong chương II, các bài toán $\mathcal{NP}$-đầy đủ là các bài toán mà cho đến nay chưa tìm được một thuật toán với độ phức tạp tính toán đa thức nào để giải chúng. Và tính « khó» của các bài toán đó lại được bảo đảm bằng sự kiện là chỉ cần có một thuật toán với độ phức tạp đa thức giải một bài toán $\mathcal{NP}$-đầy đủ nào đó thì lập tức mọi bài toán $\mathcal{NP}$-đầy đủ đều giải được trong thời gian đa thức.

Đối với một số bài toán $\mathcal{NP}$-đầy đủ, tuy không có thuật toán với độ phức tạp đa thức để giải đối với mọi dữ liệu của bài toán, nhưng có thể có một lớp các dữ liệu mà đối với chúng có thuật toán để giải với thời gian chấp nhận được. Với những bài toán như vậy ta có thể sử dụng để xây dựng các hệ mật mã khoá công khai với nguyên tắc chung như sau : Hệ mật mã sẽ có phép giải mã tương đương với việc tìm lời giải cho bài toán $\mathcal{NP}$-đầy đủ đó; tuy nhiên có một thủ tục để biến một dữ liệu nói chung của bài toán $\mathcal{NP}$-đầy đủ đó thành một dữ liệu thuộc lớp đặc biệt mà đối với nó có thể giải được bởi một thuật toán với độ phức tạp thời gian chấp nhận được. Như vậy, ta đã biến được phép lập mã thành một hàm « cửa sập một phía », và đó là cơ sở để xây dựng hệ mật mã khoá công khai tương ứng.

Ta sẽ xét sau đây hai trường hợp xây dựng được các hệ mật mã khoá công khai theo cách như vậy : một là hệ mật mã Merkle-Hellman dựa trên bài toán sắp ba lô (hay bài toán tổng tập con), và hai là hệ mật mã Mc-Eliece dựa trên bài toán giải mã tuyến tính tự sửa sai.

### 4.5.2. Hệ mật mã Merkle-Hellman.

Bài toán sắp ba lô (tức bài toán KNAPSACK, cũng được gọi là bài toán tổng tập con) được đặt ra như sau: Cho một tập các số nguyên dương $\{a_1, a_2, ..., a_n\}$ và một số nguyên dương $s$. Hãy xác định xem có hay không một tập con các $a_j$ mà tổng của chúng bằng $s$. Một cách tương đương, hãy xác định xem có hay không các $x_i \in \{0,1\}$ $(1 \le i \le n)$ sao cho $\sum_{i=1}^{n} a_i x_i = s$.

Bài toán này là $\mathcal{NP}$-đầy đủ, tuy nhiên nếu ta hạn chế bài toán trên các dữ liệu $I = (\{a_1, a_2, ..., a_n\}, T)$, trong đó $\{a_1, a_2, ..., a_n\}$ là dãy *siêu tăng*, tức là dãy thoả mãn điều kiện

$$\forall j = 2, 3, ..., n: \; a_j > \sum_{i=1}^{j-1} a_i,$$

thì việc tìm trả lời là khá dễ dàng, chẳng hạn có thể bằng thuật toán đơn giản dưới đây:

1. **for** $i = n$ **downto** 1 **do**
   **if** $T > a_i$ **then** $T = T - a_i$, $x_i = 1$, **else** $x_i = 0$
2. **if** $\sum_{i=1}^{n} x_i.a_i = T$ **then** $X = (x_1, ..., x_n)$ is the solution of problem,
   **else** there is no solution.

Bây giờ, để chuẩn bị xây dựng một sơ đồ mật mã Merkle-Hellman, ta chọn trước một số nguyên dương $n$ và một số nguyên tố $p$ đủ lớn. Với mỗi người tham gia sẽ được chọn một bộ khoá $K = (K', K'')$, trong đó khoá bí mật $K'' = (A, p, a)$ gồm một dãy siêu tăng $A = \{a_1, a_2, ..., a_n\}$ thoả mãn $\sum_{i=1}^{n} a_i < p$, và một số $a$, $1 < a < p$ ; khoá công khai $K' = \{b_1, ..., b_n\}$ với $b_i = a.a_i \bmod p$.

*Sơ đồ hệ mật mã Merkle-Hellman* được định nghĩa bởi

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

trong đó $\mathcal{P} = \{0,1\}^n$, $\mathcal{C} = \{0, 1, ..., n(p-1)\}$, $\mathcal{K}$ là tập các bộ khoá $K = (K', K'')$ như được xây dựng ở trên. Các thuật toán lập mật mã và giải mã được xác định bởi:

Với mọi $x = (x_1, ..., x_n) \in \mathcal{P}$ thuật toán lập mã cho ta

$$\mathcal{E}(K', x) = \sum_{i=1}^{n} x_i.b_i \; ;$$

và với mọi $y \in \mathcal{C}$, ta tính $z = a^{-1}.y \bmod p$, rồi sau đó giải bài toán sắp balô đối với dữ liệu $I = (\{a_1, a_2, ..., a_n\}, z)$ ta sẽ được lời giải $(x_1, ..., x_n)$, lời giải đó là giá trị của $\mathcal{D}(K'', y)$.

Thí dụ: Chọn $n = 6$, khoá bí mật có $p = 737$, $A = \{12, 17, 33, 74, 157, 316\}$, $a = 635$. Tính được khoá công khai là {250, 477, 319, 559, 200, 196}. Với bản rõ $x = 101101$ ta có bản mã tương ứng là $y = 1324$. Để giải mã, trước hết tính $z = a^{-1}.y \bmod p = 635^{-1}.1324 \bmod 737 = 435$, sau đó giải bài toán sắp balô với dãy siêu tăng $A$ và $z$ ta được

$$435 = 12 + 33 + 74 + 316,$$

tức được lời giải $x = (1,0,1,1,0,1)$.

Hệ mật mã Merkle-Hellman được đề xuất khá sớm, từ năm 1978, đến năm 1985 Shamir tìm được một phương pháp thám mã trong thời gian đa thức dựa vào một thuật toán của Lenstra giải bài toán qui hoạch động. Tuy nhiên, sau đó, vào năm 1988, Chor và Rivest có đưa ra một cách khác xây dựng hệ mật mã cũng dựa vào bài toán sắp balô, cho đến nay vẫn giữ được an toàn.

### 4.5.3. Hệ mật mã McEliece.

Hệ mật mã McEliece được xây dựng dựa vào tính $\mathcal{NP}$-đầy đủ của bài toán giải mã tuyến tính tự sửa sai (trong lý thuyết truyền tin). Bài toán được đặt ra như sau: giả sử nguồn tin là tập các từ $k$ bit nhị phân, tức tập hợp $\{0,1\}^k$, được truyền đi trên một kênh có nhiễu, tức là nếu truyền trực tiếp các dãy từ $k$ bit thì thông tin mà ta nhận được có thể bị sai lệch và ta không nhận được đúng thông tin được truyền đi. Để khắc phục những sai lệch đó người ta tìm cách mã hoá nguồn tin gốc bằng cách thêm cho mỗi từ $k$ bit mang thông tin một số bit dùng để tự hiệu chỉnh, tức là thực hiện một phép mã hoá biến mỗi từ $k$ bit ban đầu thành một từ $n$ bit, với $n > k$, được gọi là từ mã. Phép mã hoá tuyến tính là phép mã hoá được thực hiện bằng cách nhân từ $k$ bit ban đầu $x$ với một ma trận $G$ cấp $k \times n$ để được từ mã $n$ bit $y$, $y = x.G$ (các phép toán cộng và nhân được thực hiện theo mod2). Ta định nghĩa khoảng cách Hamming giữa hai từ mã $n$ bit là số các vị trí mà tại đó hai từ mã có giá trị khác nhau; khoảng cách $d$ của hệ mã là khoảng cách Hamming bé nhất giữa hai từ mã bất kỳ. Như vậy, một hệ mã tuyến tính được xác định bởi một ma trận $G$ (gọi là ma trận sinh), và được đặc trưng bởi ba số $[n,k,d]$. Nếu $d = 2t+1$, thì hệ mã có khả năng tự sửa sai đến $t$ sai ngẫu nhiên nhiễm phải do nhiễu của kênh truyền. Tuy nhiên, việc tự sửa sai (tức là khi nhận được từ mã có thể có đến $t$ sai ta tìm lại được đúng từ $k$ bit thông tin ban đầu) của các hệ mã tuyến tính như vậy nói chung khá phức tạp, và bài toán giải mã tuyến tính tự sửa sai đã được chứng minh là một bài toán $\mathcal{NP}$-khó, tức cho đến nay chưa biết có thuật toán nào làm việc trong thời gian đa thức giải được nó. Mặc dầu vậy, người ta đã tìm được một số lớp riêng các hệ mã tuyến tính mà đối với chúng có thể xây dựng được những thuật toán giải mã tự sửa sai làm việc có hiệu quả, các hệ mã Goppa là một lớp như vậy. Hệ mã Goppa là một loại hệ mã tuyến tính có các đặc trưng $n = 2^m$, $d = 2t+1$, $k = n - mt$, có ma trận sinh $G$ cấp $k \times n$ được xây dựng dựa trên một số tính chất đại số của trường GF($2^m$)-mà ở đây ta không đi vào các chi tiết.

Để có một hệ mật mã McEliece, trước hết ta chọn một hệ mã Goppa với ma trận sinh $G$ và các đặc trưng trên, sau đó dùng một

ma trận $S$ khả nghịch cấp $k \times k$ trên $Z_2$ và một ma trận hoán vị $P$ cấp $n \times n$ (cũng có các phần tử trong $Z_2$) để biến hệ mã Goppa với ma trận sinh $G$ thành một hệ mã tuyến tính "phổ biến" với ma trận sinh $G^* = SGP$; vậy là đã biến hệ mã Goppa có thuật toán giải mã hiệu quả thành một hệ mã tuyến tính nói chung mà ta chỉ biết việc giải mã tự sửa sai đối với nó là $\mathcal{NP}$-khó. Hệ mật mã mà ta xây dựng sẽ có thuật toán giải mã là "dễ" đối với người trong cuộc như giải mã Goppa, và là "khó" đối với người ngoài như giải mã tuyến tính nói chung!

Như vậy, một *hệ mật mã khoá công khai McEliece* được xác định bởi

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

trong đó $\mathcal{P} = \{0,1\}^k$, $\mathcal{C} = \{0,1\}^n$, $\mathcal{K}$ là tập hợp các bộ khoá $K = (K', K'')$, với khoá bí mật $K'' = (G,S,P)$ gồm một ma trận sinh $G$ của một hệ mã Goppa, một ma trận khả nghịch $S$ cấp $k \times k$ trên $Z_2$ và một ma trận hoán vị $P$ cấp $n \times n$; khoá công khai $K' = G^*$ là ma trận "đã được biến đổi" nói trên.

Thuật toán lập mật mã $\mathcal{E}(K',.): \mathcal{P} \to \mathcal{C}$ được xác định bởi

$$\mathcal{E}(K', x) = x.G^* + \boldsymbol{e},$$

trong đó $\boldsymbol{e} \in \{0,1\}^n$ là một vectơ ngẫu nhiên có trọng số $t$, tức có $t$ thành phần là 1. Thuật toán giải mã $\mathcal{D}(K'',.)$ được thực hiện theo ba bước như sau với mọi $y \in \mathcal{C} = \{0,1\}^n$:

1. Tính $y_1 = y.P^{-1}$,
2. Giải mã Goppa đối với $y_1$, giả sử được $x_1$.
3. Tính $\mathcal{D}(K'', y) = x_1.S^{-1}$.

Dễ thử lại rằng các thuật toán lập mật mã và giải mã xác định như trên là hợp thức, vì với mọi $x \in \mathcal{P} = \{0,1\}^k$, ta đều có

$$\mathcal{D}(K'', \mathcal{E}(K', x)) = x,$$

Đẳng thức đó đúng với mọi vectơ $\boldsymbol{e}$ bất kỳ có trọng số $\le t$. Hệ mật mã này cũng tương tự như hệ mật mã ElGamal ở chỗ khi lập mật mã ta có thể chọn thêm cho dữ liệu vào một yếu tố ngẫu nhiên; về sau ta sẽ gọi những hệ mật mã như vậy là hệ mật mã xác suất.

Yếu tố chủ yếu bảo đảm tính an toàn của các hệ mật mã McEliece là ở chỗ từ khoá công khai $G^*$ khó phát hiện ra khoá bí mật $(G,S,P)$ và ở tính $\mathcal{NP}$-khó của bài toán giải mã tuyến tính tự sửa sai nói chung. Cũng cần nhớ rằng độ an toàn còn phụ thuộc vào việc chọn các tham số $k,n,t$ đủ lớn; theo gợi ý của các nghiên cứu thực nghiệm thì đủ lớn có nghĩa là $n \approx 1024$, $k \approx 644$, $t \approx 38$. Với những đòi hỏi đó thì kích cỡ của các ma trận $G$, $S$, $P$ và $G^*$ sẽ quá

lớn, khá bất tiện cho việc thực thi trong thực tế, vì vậy mà các hệ mật mã McEliece chưa được sử dụng phổ biến lắm.

## 4.6. Các hệ mật mã xác suất khoá công khai.

### 4.6.1. Đặt vấn đề và định nghĩa.

Mật mã xác suất là một ý tưởng được đề xuất bởi Goldwasser và Micali từ năm 1984, xuất phát từ yêu cầu giải quyết một vấn đề sau đây: Giả thiết ta có một hệ mật mã khoá công khai, và ta muốn lập mật mã cho bản rõ chỉ gồm một bit. Điều đó thường gặp khi ta muốn bí mật truyền đi một thông tin chỉ có nội dung là *có* hoặc *không*, tức là một thông tin đặc biệt quan trọng nhưng chỉ gồm một bit. Nếu ta dùng một hệ mật mã khoá công khai thông thường, thì bản mật mã được truyền đi sẽ là $e_{K'}(0)$ hoặc $e_{K'}(1)$, một người thám mã có thể không biết cách giải mã, nhưng lại hoàn toàn có thể tính trước các giá trị $e_{K'}(0)$ và $e_{K'}(1)$, và khi lấy được bản mã truyền đi trên kênh truyền tin công cộng, chỉ cần so sánh bản mã nhận được đó với hai bản $e_{K'}(0)$ và $e_{K'}(1)$ đã được tính sẵn là đủ biết được thông tin mật được truyền đi là 0 hay là 1. Các hệ mật mã khoá công khai sở dĩ có được tính bảo mật là vì từ thông tin về bản mã khó lòng khai thác được thông tin gì về bản rõ, nhưng rõ ràng điều đó không còn được bảo đảm nếu số các bản rõ là rất ít, chẳng hạn như khi các bản rõ có độ dài cực ngắn, hay như trường hợp trên, số các bản rõ chỉ là hai, cụ thể là 0 và 1.

Mục đích của việc xây dựng mật mã xác suất là để bảo đảm không một thông tin nào về bản rõ có thể khai thác được (trong thời gian đa thức) từ bản mã; điều này, đối với các hệ mật mã khoá công khai, có thể được thực hiện bằng cách tạo cho một bản rõ nhiều bản mã khác nhau thu được một cách ngẫu nhiên với việc sử dụng các số ngẫu nhiên trong tiến trình lập mã. Sau đây là định nghĩa về một hệ mật mã xác suất khoá công khai:

Định nghĩa. Một *hệ mật mã xác suất khoá công khai* được xác định bởi một bộ

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

trong đó $\mathcal{P}$, $\mathcal{C}$, $\mathcal{K}$ được hiểu như đối với các hệ mật mã khoá công khai thông thường, $\mathcal{R}$ là một tập các phần tử ngẫu nhiên, và với mỗi $K = (K', K'') \in \mathcal{K}$, thuật toán lập mật mã $e_{K'} = \mathcal{E}(K', .): \mathcal{P} \times \mathcal{R} \to \mathcal{C}$ và giải mã $d_{K''} = \mathcal{D}(K'', .): \mathcal{C} \to \mathcal{P}$ thoả mãn đẳng thức:

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

Ngoài ra, ta mong muốn một *điều kiện an toàn* như trong định nghĩa sau đây được thoả mãn: ta ký hiệu $p_{K,x}$ là phân bố xác