MẬT MÃ ỨNG DỤNG TRONG AN TOÀN THÔNG TIN

(Applied Cryptography In Secure Information System)

ThS. Bùi Hữu Đông
buihuudong19@gmail.com
0903.82.36.46

Học viện Kỹ thuật Mật mã, khoa An toàn thông tin

Ngày 19 tháng 11 năm 2020

Buổi 05

MÃ HÓA KHÓA CÔNG KHAI (Public-Key Cryptography)

Giới thiệu

- Nhược điểm cố hữu của mã hóa đối xứng:
  - Vấn đề làm thế nào để trao đổi khóa giữa người gửi và người nhận. Cần một kênh an toàn riêng => chi phí lớn và chậm trễ về mặt thời gian
  - Tính bí mật của khóa: không có cơ sở quy trách nhiệm nếu khóa bị tiết lộ

Giới thiệu

- Sự ra đời của mã hóa bất đối xứng (asymetric cryptography):
  - Vào năm 1976 Whitfield Diffie và Martin Hellman đã tìm ra một phương pháp mã hóa khác mà có thể giải quyết được hai vấn đề trên
  - Việc thực hiện hệ mật khoá công khai thì lại được Rivest. Shamin và Adieman đưa ra đầu tiên vào năm 1977
  - Để khắc phục điểm yếu của mã hóa đối xứng người ta tập trung vào nghiên cứu theo hướng: có phương pháp nào để việc mã hóa và giải mã dùng hai khóa khác nhau?
  - Có nghĩa là: $C = E(P, K_1)$ và $P = D(C, K_2)$

Giới thiệu

- Có hai phương án áp dụng phương pháp trên:
  - *Phương án 1*: Người nhận giữ bí mật khóa K_2 còn K_1 thì công khai và người gửi sẽ dùng K_1 để mã hóa => Đảm bảo *tính bí mật* của quá trình truyền dữa liệu và cũng không cần truyền khóa K_1 này trên kênh an toàn.
  - *Phương án 2*: Người gửi giữ khóa bí mật K_1 và K_2 hoàn toàn công khai. Người gửi dùng K_1 để mã hóa và người nhận dùng K_2 để giải mã, vì K_2 là public nên kẻ phá mã có thể dùng K_2 để giải mã => *không đảm bảo tính bí mật*, nhưng đảm bảo *tính chứng thực* và *tính chống chối bỏ*
- Áp dụng kết hợp 2 phương án trên => giải quyết được nhược điểm của mã hóa đối xứng.

Một số quy ước

- Ta quy ước một số ký hiệu sau:
  - Khóa bí mật - khóa riêng (private key) - tránh nhầm với khóa bí mật trong hệ mã đối xứng. Ký hiệu: K_R
  - Khóa công khai (public key). Ký hiệu là K_U
  - Bản rõ: P, bản mã: C
  - Phương án 1 viết thành:

$$C = E(P, K_U)$$
$$P = D(C, K_R)$$

  - Phương án 2 viết thành:

$$C = E(P, K_R)$$
$$P = D(C, K_U)$$

Hàm một chiều

- Liệu có một mô hình quá trình mã và giải mã dùng hai khóa khác nhau không?
  - Chắc chắn K_U và K_R không thể hoàn toàn độc lập. và chúng phải có một mối quan hệ nào đó với nhau. Có nghĩa:

$$K_R = f(K_U)$$

  - Việc tính $K_R = f(K_U)$ phải là bất khả thi về mặt thời gian
  - Để có được K_U và K_R người ta dùng các *hàm một chiều - oneway function*

Hàm một chiều

- Định nghĩa:

Cho hàm $f : X \to Y$ được gọi là hàm một chiều nếu tính $y = f(x)$ với $\forall x \in X$ là dễ nhưng việc tìm x khi biết y lại là vấn đề khó.

- Tại sao lại dùng hàm một chiều?

  Vì, các hàm một chiều có tính chất là hàm nghịch đảo của chúng rất khó thực hiện.

Hàm một chiều

- Thí dụ:
  - Việc sinh ra hai số nguyên tố lớn p, q và tính tích $N = pq$ thì thực hiện dễ dàng
  - Tuy nhiên nếu chỉ cho trước N và thực hiện phân tích N để tìm lại hai số nguyên tố p, q là việc hoàn toàn bất khả thi về mặt thời gian.

Một số phương pháp mã hóa thuộc khóa công khai

- Hệ mật mã RSA
- Phương pháp Knapsack
- Phương pháp đường cong elliptic
- Hệ mật xếp balô Merkle – Hellman
- Hệ mật McEliece
- Hệ mật ElGamal
- Hệ mật Chor – Rivest

Kiến thức cơ bản lý thuyết số

- Một số khái niệm:
  - *Phép chia modulo*: Phép chia modulo là phép chia lấy phần dư. Tổng quát:

    $a \ mod \ n = r$ với $a >= 0; n > 0; 0 <= r <= n - 1$

    Thí dụ: $27 \ mod \ 8 = 3$, $35 \ mod \ 9 = 8$
  - *Đồng dư*: Nếu hai số a, b có cùng số dư trong phép chia cho n thì ta nói rằng a và b là đồng dư trong phép chia modulo cho n.

    Kí hiệu: $\equiv$

    $a \equiv b \ (mod) \ n$ hay $a \equiv b \ mod \ n$

Kiến thức cơ bản lý thuyết số

- Một số khái niệm:
  - *Nhận xét*: phép toán modulo phân hoạch tập số tự nhiên N thành n lớp tương đương đồng dư ứng với các giá trị của r trong tập $\{0, 1, 2, 3, ..., n-1\}$
  - *Thí dụ*: với $n = 4$ ta có 4 lớp tương đương sau:

$$\{0, 4, 8, 12, 16, ...\}$$

$$\{1, 5, 9, 13, 17, ...\}$$

$$\{2, 6, 10, 14, 18, ...\}$$

$$\{3, 7, 11, 15, 19, ...\}$$

Kiến thức cơ bản lý thuyết số

- *Tính chất của phép modulo*: Cho a, b và n là các số nguyên, phép modulo có các tính chất:
  - $(a + b)\ mod\ n = [(a\ mod\ n) + (b\ mod\ n)\ ]\ mod\ n$
  - $(a - b)\ mod\ n = [(a\ mod\ n)$ - $(b\ mod\ n)\ ]\ mod\ n$
  - $(a \times b)\ mod\ n = [(a\ mod\ n) \times (b\ mod\ n)\ ]\ mod\ n$
- *Ước số* : nếu $a\ mod\ n = 0$ hay $a = 0\ mod\ n$, nghĩa là a chia hết cho n hay n là ước số của a

  Sinh viên về tìm hiểu thuật toán Euclid và Euclid mở rộng.
- *Số nguyên tố* : Số p được gọi là số nguyên tố khi nó chỉ chia hết cho 1 và chính nó.

Kiến thức cơ bản lý thuyết số

- *Số nguyên tố cùng nhau*: hai số a và b được gọi là nguyên tố cùng nhau nếu $gcd(a, b) = 1$

  Ký hiệu: $a \perp b$

  Thí dụ: $3 \perp 8$, $7 \perp 9$
- *Phần tử nghịch đảo của phép nhân modulo*: Nếu hai số nguyên a và n nguyên tố cùng nhau, thì tồn tại số nguyên w sao cho:

$$a.w = 1 \, mod \, n$$

  Ta gọi w là phần tử nghịch đảo của a trong modulo n.

  Ký hiệu: a^{-1}

Kiến thức cơ bản lý thuyết số

- *Phần tử nghịch đảo*: Thí dụ:

  ◇ $n = 10, a = 7$ là hai số nguyên tố cùng nhau, do đó tìm được $a^{-1} = 3(21 \equiv 1\ mod\ 10)$

a^{-1}	0	1	2	3	4	5	6	7	8	9
a^{-1} x 7	0	7	4	1	8	5	2	9	6	3

  ◇ $n = 10, a = 2$ không phải là hai số nguyên tố cùng nhau, ta có bảng phép nhân sau:

a^{-1}	0	1	2	3	4	5	6	7	8	9
a^{-1} x 2	0	2	4	6	8	0	2	4	6	8

  Để tính chúng ta dùng thuật toán Euclid mở rộng.

Kiến thức cơ bản lý thuyết số

- Định lý Fermat:
  - *Nội dung:* Nếu p là số nguyên tố và a là số nguyên không chia hết cho p thì $a^{p-1} \equiv 1 \, mod \, p$
  - *Chứng minh*: ta xét $X = \{a \, mod \, p, 2a \, mod \, p, ..., (n-1)a \, mod \, p\}$

    Nhân xét:
    - vì $gcd(a, p) = 1$ nên tập X không có phần tử nào $= 0$
    - vì không tồn tại phần tử thứ i và j $(i \neq j)$ sao cho: $ia \, mod \, p = ja \, mod \, p$. Vì $gcd(a, p) = 1$ nên tồn tại a^{-1} trong phép modulo p. Do đó, nếu $ia \equiv ja \, mod \, p$ thì $iaa^{-1} \equiv jaa^{-1} \, mod \, p$. Tức là $i \equiv j \, mod \, p$ => trái với giả thiết $i \neq j$

      Từ hai nhận xét trên ta suy ra các phần tử của X sẽ là một hoán vị của các giá trị $\{1, 2, 3, ..., p-1\}$. Do đó:

Kiến thức cơ bản lý thuyết số

Từ hai nhận xét trên ta suy ra các phần tử của X sẽ là một hoán vị của các giá trị $\{1, 2, 3, ..., p-1\}$. Do đó:

$$a\,x\,2a\,x\,...,(p-1)\,a \equiv [1\,x\,2\,x, ..., x\,(p-1)]\,mod\,n$$

$$=> a\,x\,a\,x\,, ..., \,x\,a \equiv a^{p-1} \equiv 1\,mod\,n$$

Thí dụ:

- $p = 5, a = 7 => 7^4 = 49.49 = 2401, 2401 \equiv 1\,mod\,5$
- $p = 7, a = 4 => 4^6 = 64.64 = 4096, 4096 \equiv 1\,mod\,7$

Kiến thức cơ bản lý thuyết số

- Phép logarit rời rạc:
  - Ta định nghĩa phép lũy thừa modulo như sau, để tính y từ a, x và n là các số nguyên:

$$y = a^x \, mod \, n = (a.a...a) \, mod \, n$$

  - Ta chỉ xét trường hợp n là số nguyên tố. Bảng sau minh họa các giá trị của phép lũy thừa modulo với $n = 19$, a và x từ 1 đến 18

Kiến thức cơ bản lý thuyết số

- Bảng sau minh họa các giá trị của phép lũy thừa modulo với $n = 19$, a và x từ 1 đến 18

a	a^2	a^3	a^4	a^5	a^6	a^7	a^8	a^9	a^{10}	a^{11}	a^{12}	a^{13}	a^{14}	a^{15}	a^{16}	a^{17}	a^{18}
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	4	8	16	13	7	14	9	18	17	15	11	3	6	12	5	10	1
3	9	8	5	15	7	2	6	18	16	10	11	14	4	12	17	13	1
4	16	7	9	17	11	6	5	1	4	16	7	9	17	11	6	5	1
5	6	11	17	9	7	16	4	1	5	6	11	17	9	7	16	4	1
6	17	7	4	5	11	9	16	1	6	17	7	4	5	11	9	16	1
7	11	1	7	11	1	7	11	1	7	11	1	7	11	1	7	11	1
8	7	18	11	12	1	8	7	18	11	12	1	8	7	18	11	12	1
9	5	7	6	16	11	4	17	1	9	5	7	6	16	11	4	17	1
10	5	12	6	3	11	15	17	18	9	14	7	13	16	8	4	2	1
11	7	1	11	7	1	11	7	1	11	7	1	11	7	1	11	7	1
12	11	18	7	8	1	12	11	18	7	8	1	12	11	18	7	8	1
13	17	12	4	14	11	10	16	18	6	2	7	15	5	8	9	3	1
14	6	8	17	10	7	3	4	18	5	13	11	2	9	12	16	15	1
15	16	12	9	2	11	13	5	18	4	3	7	10	17	8	6	14	1
16	9	11	5	4	7	17	6	1	16	9	11	5	4	7	17	6	1
17	4	11	16	6	7	5	9	1	17	4	11	16	6	7	5	9	1
18	1	18	1	18	1	18	1	18	1	18	1	18	1	18	1	18	1

Kiến thức cơ bản lý thuyết số

- Nhận xét:
  - Nhìn vào bảng trên, ta thấy rằng không phải hàng nào cũng có đầy đủ các giá trị từ 1 đến 18. Xét hàng $a = 11$, ta có:
    - $11^1 \equiv 11 \; mod \; 19$
    - $11^2 = 121 \equiv 7 \; mod \; 19$
    - $11^3 = 1331 \equiv 1 \; mod \; 19$
    - $11^4 = 11^3.11 \equiv 11 \; mod \; 19$ (giống hàng đầu)
    - $11^5 \equiv 11^2 \; mod \; 19$
    - .....

      Do đó hàng $a = 11$ (tương ứng với dãy $11^1, 11^2, \ldots, 11^{18}$) chỉ có ba giá trị 11, 7, 1 được lặp lại theo chu kỳ
  - Trong bảng trên chỉ có các giá trị $a = 2, 3, 10, 13, 14, 15$ là làm cho dãy $a^1, a^2, \ldots, a^{18}$ có đầu đủ các giá trị từ 1 đến 18 với phép modulo 19. Như vậy chỉ có $a = 2, 3, 10, 13, 14, 15$ thì phép lũy thừa modulo trên mới *khả nghịch*.

Kiến thức cơ bản lý thuyết số

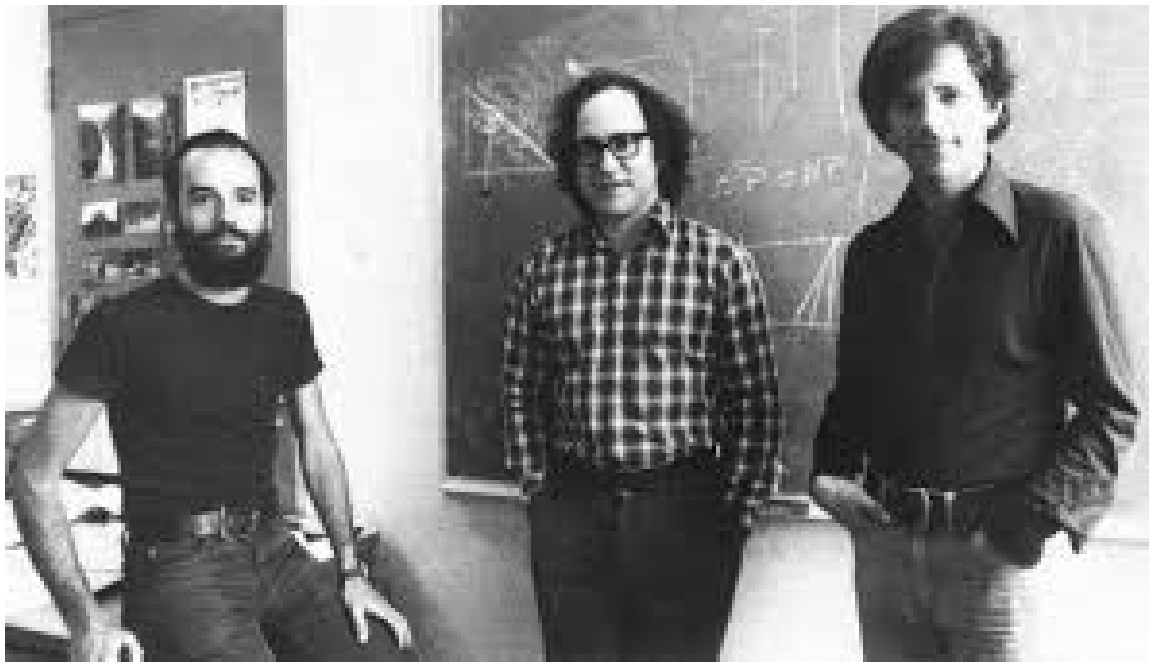
- Nhận xét <tiếp>:
  - Trong trường hợp tổng quát với mỗi n chỉ có một số trường hợp của a thì phép lũy thừa là khả nghịch. Lúc này a được gọi là *primitive root* của n
  - Và cũng tương tự như số thực, nếu biết y, a và n, muốn tìm lại x thì ta cũng dùng hàm logarith, được gọi là logarith rời rạc

$$x = dlog_{a,n}y$$

  - Tuy nhiên không giống như số thực, việc tính logarith rời rạc đã được chứng minh là rất tốn kém về mặt thời gian. Và được xem như là bất khả thi nếu a và n là các số lớn => phép lũy thừa modulo cũng được xem là *hàm một chiều* và được ứng dụng trong phương pháp trao đổi khóa Diffie – Hellman

Hệ mật mã công khai RSA

- Giới thiệu:
  - Là phương pháp mã hóa với khóa công khai
  - RSA được xây dựng bởi các tác giả Ron Rivest, Adi Shamir và Len Adleman tại học viện MIT vào năm 1977



Hệ mật mã công khai RSA - Giới thiệu

- 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 bít của khối.
- Kích thước thường dùng của i là 1024 bít
- RSA sử dụng hàm một chiều là vấn đề phân tích một số thành thừa số nguyên tố.

Nguyên tắc hoạt động của RSA

Để thực hiện mã hóa và giải mã, RSA dùng phép lũy thừa modulo của lý thuyết số. Với các bước:

1) Chọn hai số nguyên tố lớn p và q và tính $N = pq$. Với p và q sao cho:

   $P < 2^{i-1} < N < 2^i$. Với $i = 1024$ thì N là số nguyên dài khoảng 309 ký số
2) Tính $n = (p-1)(q-1)$
3) Tìm e sao cho $gcd(e, n) = 1$
4) Tìm số d sao cho: $e.d \equiv 1 \, mod \; n$. Với d là phần tử nghịch đảo của e trong phép modulo
5)

Nguyên tắc hoạt động của RSA

5) Hủy bỏ n, p và q. Chọn khóa công khai K_U là cặp (e, N), khóa riêng K_R là cặp (d, N)
6) Quá trình mã hóa:
   - Theo phương án 1, mã hóa bảo mật: $C = E(P, K_U) = P^e \ mod \ N$
   - Theo phương án 2, mã hóa chứng thực: $C = E(P, K_R) = P^d \ mod \ N$
7) Quá trình giải mã:
   - Theo phương án 1, mã hóa bảo mật: $P' = D(C, K_R) = C^d \ mod \ N$
   - Theo phương án 2, mã hóa chứng thực: $P' = D(C, K_U) = C^e \ mod \ N$

Bản rõ P có kích thước $i - 1$ bít, bản mã C có kích thước i bít

Nguyên tắc hoạt động của RSA

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: $P = P'$ (Nhóm tiểu luận => chứng minh)
- Không thể suy ra 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 vậy, phải tính được n vì $n = (p-1)(q-1)$ nên cần tìm p và q. Và vi $N = pq$ nên chỉ tính được p, q từ N=> bất khả thi vì $N = pq$ là hàm 1 chiều => Do đó, không thể tính được K_R từ K_U

Thí dụ về RSA

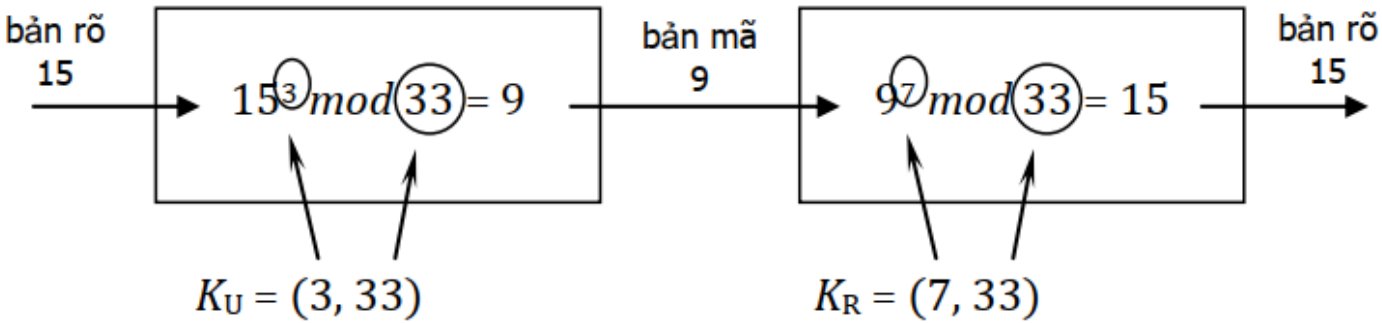
Để minh họa ta sẽ thực hiện một ví dụ về mã hóa RSA với kích thước khóa là 6 bít.

1) Chọn $p = 11$ và $q = 3$, do đó $N = pq = 33$ $(2^5 = 32 < 33 < 64 = 2^6)$
2) $n = (p-1)(q-1) = 20$
3) Chọn $e = 3$ nguyên tố cùng nhau với n
4) Tính nghịch đảo của e trong phép modulo n được $d = 7$ $(3x7 = 21)$
5) Khóa công khai $K_U = (e, N) = (3, 33)$. Khóa bí mật $K_R = (d, N) = (7, 33)$

Thí dụ về RSA

6) Theo phương án 1 (mã hóa bảo mật): Mã hóa bản rõ $P = 15$.
Ta có:
$C = P^e \ mod \ N = 15^3 \ mod \ 33 = 9 \ (15^3 = 3375 = 102 \, x \, 33 + 9)$

7) Giải mã bản mã $C = 9$, ta có:
$P' = C^d \ mod \ N = 9^7 \ mod \ 33 = 15 = P$ vì
$9^7 = 4.782.696 = 144.938 \, x \, 33 + 15$



Thí dụ về RSA

6) Theo phương án 2 (mã hóa chứng thực): Mã hóa bản rõ $P = 15$. Ta có:

   $C = P^d \ mod \ N = 15^8 \ mod \ 33 = 27$ vì
   $15^7 = 170.859.375 = 5.177.556 \, x \, 33 + 27$

7) Giải mã $C = 27$, ta có:

   $P' = C^e \ mod \ N = 27^3 \ mod \ 33 = 15 = P$ vì
   $27^3 = 19.683 = 596 \, x \, 33 + 15$

Độ phức tạp tính toán trong RSA

Có hai vấn đề về độ phức tạp tính toán trong phương pháp RSA. Đó là các phép tính sinh khóa và các phép tính mã hóa/giải mã

- Phép tính mã hóa/giải mã
- Phép tính sinh khóa

Độ an toàn của RSA

Một số các tấn công phương pháp RSA:

- Vét cạn khóa: Bằng cách thử tất cả các khóa d. Tuy nhiên, Với N lớn, việc tấn công là bất khả thi
- Phân tích N thành thừa số nguyên tố N = pq: do $N = pq$ là hàm 1 chiều => Phân tích bất khả thi với N lớn.

Độ an toàn của RSA

- Bảng sau liệt kê kích thước N của các RSA đã phá mã được cho đến hiện nay:

Số chữ số của N	*Số bít*	*Năm phá mã*	*Thuật toán*
100	322	1991	Quadratic sieve
110	365	1992	Quadratic sieve
120	398	1993	Quadratic sieve
129	428	1994	Quadratic sieve
130	431	1996	GNFS
140	465	1999	GNFS
155	512	1999	GNFS
160	530	2003	Lattice sieve
174	576	2003	Lattice sieve
200	633	2005	Lattice sieve

Bảo mật, chứng thực và không từ chối với mã hóa khóa công khai

- Giả sử An muốn gửi dữ liệu cho Bình dùng mã hóa khóa công khai, trước tiên An và Bình sẽ chọn cặp khóa riêng-khóa công khai. Ký hiệu khóa riêng-khóa công khai của An là K_{RA} và K_{UA}, của Bình là K_{RB} và K_{UB}
- Để gửi dữ liệu bảo mật cho Bình, An sẽ dùng phương án 1: mã hóa dữ liệu bằng khóa công khai K_{UB} của Bình, và Bình dùng khóa riêng K_{RB} để giải mã:

$$C = E(P, K_{UB})$$

$$P = D(C, K_{RB})$$

Bảo mật, chứng thực và không từ chối với mã hóa khóa công khai

- Để đảm bảo tính chứng thực và An không từ chối trách nhiệm gửi dữ liệu, An sẽ dùng phương án 2: An mã hóa dữ liệu bằng khóa riêng K_{RA}, và Bình dùng khóa công khai K_{RA} của An để giải mã:

$$C = E(P, K_{RA})$$

$$P = D(C, K_{UA})$$

- Tuy nhiên mô hình như trên lại không đảm bảo tính bảo mật. Vì không chỉ riêng Bình, kẻ trộm (T) cũng biết được khóa công khai K_{UA} của An. Do đó T có thể giải mã bản mã C và biết được nội dung bản rõ M

Bảo mật, chứng thực và không từ chối với mã hóa khóa công khai

- Để giải quyết vấn đề trên, người ta kết hợp tính bảo mật, chứng thực và không từ chối qua mô hình sau:

$$C = E(E(P, K_{RA}), K_{UB})$$

$$M = D(D(C, K_{RB}), K_{UA})$$

Bài tập

1. Thực hiện mã hóa và giải mã bằng phương pháp RSA với $p = 3, q = 11, e = 7, P = 5$ theo hai trường hợp mã hóa bảo mật và mã hóa chứng thực
2. Tìm hiểu thuật toán Miller-Rabin để kiểm tra tính nguyên tố của một số nguyên cho trước.
3. Viết chương trình thể hiện thuật toán Euclid mở rộng áp dụng cho các số nguyên nhỏ 32 bít
4. Viết chương trình sinh một số nguyên tố nhỏ (32 bít) dùng thuật toán Miller-Rabin
5. Viết chương trình mã hóa file bằng thuật toán RSA trên số nguyên nhỏ
6. Viết chương trình thực hiện các phép toán +, -, *, mod trên các số nguyên lớn (kích thước tối đa một số nguyên là 1024 bít). Gợi ý: mỗi số nguyên được biểu diễn bằng một mảng các phần tử 32 bít.