Bài 1. Hệ mật khóa công khai RSA.

Giới thiệu

Cùng với sự phát triển của khoa học và công nghệ, xuất hiện nhu cầu về các Hệ mật hiện đại. Các hệ mật này cho phép 2 đối tác chưa hề có liên hệ mà vẫn tiến hành thỏa thuận và ký kết các Hợp đồng online.

Nội dung

1. Mô tả hệ mật mã RSA.

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

$S = (P, C, K, E, D)$ (1)

trong đó P là tập ký tự bản rõ, C là tập ký tự bản mã, 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 $x \in P$, thuật toán lập mã E cho ta ký tự mã tương ứng $y = E(K', x) \in C$, và với ký tự mã

y thuật toán giải mã D sẽ cho ta lại ký tự bản rõ x : $D\,(K'', y) = D\,(K'', E\,(K', x)) = x$.

Các bước Xây dựng hệ mật

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, sơ đồ chung của hệ mật mã RSA được định nghĩa bởi danh sách (1), trong đó:

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

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

E và D được xác định bởi:

$E(K', x) = x^e \bmod n = y$, với mọi $x \in P$,

$D(K'', y) = y^d \bmod n$, với mọi $y \in C$.

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

$$D(K'', E(K', x)) = x .$$

Thực vậy, do $e.d \equiv 1 (\bmod \phi(n))$ 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

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

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

$$x^{t\varphi(n)+1} \equiv x(mod\, p),$$

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

từ đó suy ra $x^{t\varphi(n)+1} \equiv x(mod\, n)$, tức $D\ (K'',\ E\ (K',\ x)) = x$.

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 $x = 3650502^{422191} \bmod 6012707 = 5234673$.

VD2. x = DUNG = 66592

Mã hóa

$y = x^e \bmod n = 66592^{3674911} \bmod n = 1132870$

Giải mã : $y^d \bmod n = 1132870^{422191} \bmod n = 66592$.

Bài tập : x là tên của SV

Bài tập 1. Thuật toán Euclid mở rộng

a = 127, b = 75 tính $b^{-1} \bmod a = ?$

Bài tập 2

a = 2357, b = tháng sinh||ngày sinh

Thuật toán lũy thừa theo modulo :
Cần tính $b^n mod m$

ALGORITHM 5 Modular Exponentiation.

procedure *modular exponentiation*(b: integer, $n = (a_{k-1}a_{k-2}\ldots a_1a_0)_2$, m: positive integers)
$x := 1$
$power := b$ **mod** m
for $i := 0$ **to** $k - 1$
    if $a_i = 1$ **then** $x := (x \cdot power)$ **mod** m
    $power := (power \cdot power)$ **mod** m
return x {x equals b^n **mod** m}

Bài tập

x = b= 10, n = 50, m = 2023

Tính 10^{50}mod2023 = 1885

x = 23, Tính x^bmod2023

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 E (với tham biến đầu vào là n ,e và x) và hàm 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 E , 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.

Bài tập
Với p và q đã cho

Anh Đỗ Đức Huy

p= 2099	q= 2399
e = 7 chữ số	d = ?

Bước 1. Tính n = p*q,

$\phi(n)$ =(p – 1) (q -1)

Lấy 1 số lẻ có 7 chữ số và dùng thuật toán Euclid để kiểm tra xem có thỏa mãn nguyên tố cùng nhau với $\phi(n)$, nếu thỏa mãn ta lấy làm e. Nếu không ta lại chọn 1 số lẻ gồm 7 chữ số.

Bước 2. Dùng thuật toán Euclid mở rộng để tính $d = e^{-1} \bmod \phi(n)$

Anh Đỗ Đức Huy : e = 5031003

d = ?

Bài tập :

Với p và q đã cho, e là số có 7 chữ số nguyên tố cùng nhau với Phi (n).

1. Tính d
2. Lấy x = Tên đầy đủ các sv
   Và cắt các nhóm phù hợp rồi số hóa
3. Mã hóa
4. Giải mã.
   BT. Lấy p và q gồm 10, 20, 50 chữ số
   Chọn e ít hơn n 1 chữ số
   X = câu thơ yêu thích.

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 $y = x^e \bmod n$, 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 $c_i = x^3 \mathrm{mod} n_i$ 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 modn_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 ô có bản mã được chọn ằ (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 $k \geq 1$ 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 $k \geq 1$ 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 mod$n$ (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.

Ví dụ: p=11, q=19, e=61, d=?

$\Phi(n)=(p-1)*(q-1)$

$=(11-1)*(19-1)=180$

$d=e^{-1}\text{mod } \Phi(n)=61^{-1}\text{mod}180$

$= 121.$

$x=v,\ e_k(x) = 21^{61}\text{mod}209= 109$

$x_1=10^{61}\text{mod}209 = 186$

Ví dụ 2. n=18923, e = 1261,

12423	11524	7243	7459	14303	6127	10964	16399
9792	13629	14407	18817	18830	13556	3159	16647
5300	13951	81	8986	8007	13167	10022	17213
2264	961	17459	4101	2999	14569	17183	15827
12693	9553	18194	3830	2664	13998	12501	18873
12161	13071	16900	7233	8270	17086	9792	14266
13236	5300	13951	8850	12129	6091	18110	3332
15061	12347	7817	7946	11675	13924	13892	18031
2620	6276	8500	201	8850	11178	16477	10161
3533	13842	7537	12259	18110	44	2364	15570
3460	9886	8687	4481	11231	7547	11383	17910
12867	13203	5102	4742	5053	15407	2976	9330
12192	56	2471	15334	841	13995	17592	13297
2430	9741	11675	424	6686	738	13874	8168
7913	6246	14301	1144	9056	15967	7328	13203
796	195	9872	16979	15404	14130	9105	2001
9792	14251	1498	11296	1105	4502	16979	1105
56	4118	11302	5988	3363	15827	6928	4191
4277	10617	874	13211	11821	3090	18110	44
2364	15570	3460	9886	9988	3798	1158	9872
16979	15404	6127	9872	3652	14838	7437	2540
1367	2512	14407	5053	1521	297	10935	17137
2186	9433	13293	7555	13618	13000	6490	5310
18676	4782	11374	446	4165	11634	3846	14611
2364	6789	11634	4493	4063	4576	17955	7965