Criptosistemul RSA

- Este primul criptosistem cu cheie publică.
- A fost inventat de Ron **R**ivest (cu idei de criptare), Adi **S**hamir (cu idei de criptare dar şi de atac) şi Leonard **A**dleman (doar cu idei de atac).
- În 1978 ei publică descoperirea lor în articolul "A method for obtaining digital signatures and public key cryptosystems".
- Securitatea sa se bazează pe dificultatea descompunerii numerelor foarte mari în factori primi (nu există algoritmi eficienţi de descompunere a numerelor foarte mari în factori primi).
- Ellis, Cocks şi Williamson, lucrând pentru BSS (British Security Service), au dezvoltat un sistem echivalent încă din 1973, dar descoperirea lor a fost declasificată abia în 1997.

Descriem mai întâi funcţionarea acestui criptosistem. Presupunem că Alice trimite mesajul iar Bob decriptează.

1. **Generarea cheilor**

   - Bob alege (sau generează) două numere prime distincte foarte mari (la întâmplare şi independente) p şi q ($p \neq q$), de cel puţin 512 de biţi fiecare şi calculează produsul lor $n = pq$.
   - Bob alege un număr întreg e astfel încât
     $$1 < e < \varphi(n) = (p-1)(q-1) \text{ şi } \text{ c.m.m.d.c. } (e, \varphi(n)) = 1,$$
     unde φ este indicatorul Euler.
   - Bob determină unicul număr d satisfăcând
     $$1 < d < (p-1)(q-1) \text{ şi } d \cdot e \equiv 1 \bmod \varphi(n).$$
     Deoarece cmmdc$(e, \varphi(n)) = 1$, d există (este chiar inversul lui e modulo $\varphi(n)$) şi se determină folosind algoritmul lui Euclid extins.

   Cheia publică a lui Bob este perechea (n, e). Numărul n se numeşte *modulul RSA* iar e se numeşte *exponent de criptare*. Cheia sa secretă este d. Numărul d se numeşte *exponent de decriptare*.

   Să remarcăm că dacă p şi q ar fi cunoscuţi, putem calcula cheia secretă d din exponentul e. Deci dacă Oscar, atacatorul, ar putea găsi descompunerea lui n în factori primi, atunci ar putea găsi uşor cheia secretă a lui Bob.

2. **Criptarea RSA**

- Presupunem că Alice vrea să cripteze textul simplu $m \in \{m \in \mathbb{N} \,|\, 1 < m < n\}$. Alice cunoaşte cheia publică a lui Bob (n, e) şi calculează textul criptat c (folosind exponenţierea rapidă) prin

$$c \equiv m^e \mod n.$$

3. **Decriptarea RSA**

- Bob va decripta mesajul c folosind cheia sa privată d prin

$$m = c^d \mod n.$$

Deecriptarea funcţionează datorită următorului rezultat:

Propoziţia 1. *Fie (n, e) cheia publică RSA şi d cheia secretă RSA corespunzătoare. Atunci*

$$(m^e)^d \mod n = m$$

pentru orice întreg m cu $0 \leq m < n$.

Demonstraţie: Deoarece $ed \equiv 1 \mod \varphi(n) = (p-1)(q-1)$, atunci există un întreg t astfel încât $ed = 1 + t(p-1)(q-1)$. Rezultă

$$(m^e)^d = m^{ed} = m^{1+t(p-1)(q-1)} = m\left(m^{(p-1)(q-1)}\right)^t.$$

Obţinem uşor acum că

$$(m^e)^d = m\left(m^{(p-1)}\right)^{(q-1)t} \equiv m \mod p.$$

Într-adevăr, dacă p este divizor al lui m, identitatea este trivială căci ambii membri ai congruenţei sunt 0 mod p. Dacă p nu este divizor al lui m (adică cmmdc $(p, m) = 1$), folosind Mica Teoremă a lui Fermat, avem $m^{p-1} \equiv 1 \mod p$. Analog se arată că

$$(m^e)^d \equiv m \mod q.$$

Deoarece p şi q sunt prime şi distincte, folosind Lema Chineză a resturilor, obţinem

$$(m^e)^d \equiv m \mod n,$$

şi pentru că $0 \leq m < n$, rezultă aserţiunea din propoziţie. ■

Exemplul 1. *Bob alege factorii primi $p = 11$ şi $q = 23$. Apoi calculează produsul lor (modulul RSA) $n = 253$ şi $\varphi(253) = (p-1)(q-1) = 10 \times 22 = 4 \times 5 \times 11 = 220$.*

Bob alege exponentul de criptare $e = 3$. Se verifică c.m.m.d.c $(3, 220) = 1$.

Cheia publică este perechea $(253, 3)$.

Aplicând algoritmul lui Euclid extins Bob obţine exponentul de decriptare (cheia secretă) $d = 147$.

Folosind cheia publică, Alice criptează mesajul $m = 26$ prin $c = 26^3$ mod $253 = 119$ şi trimite lui Bob.

Folosind cheia sa secretă, Bob decriptează mesajul primit prin 119^{147} mod 253 şi se obţine mesajul iniţial necifrat, adică m=26.

Securitatea cheii secrete

Aşa cum am mai amintit, în cazul unui sistem criptografic cu cheie publică şi în particular în cazul criptosutemului RSA, este imposibil (sau deosebit de dificil) să calculăm cheia secretă din cheia publică. Se poate arăta că dificultatea găsirii lui d, ştiind pe (n, e), este aceeaşi cu dificultatea găsirii factorilor primi p şi q ai lui n. Aceasta nu demonstrează direct dificultatea calculării cheii secrete, dar reduce demonstarea dificultăţii la o problemă faimoasă, aceea a descompunerii în factori primi a unui număr. Nu se ştiu demonstraţii care să dovedească dificultatea factorizării modulului RSA. Totuşi dacă factorii p şi q sunt suficient de mari, atunci nimeni nu poate şti cum se descompune n în factori.

Faptul că securitatea criptosistemului RSA depinde de o problemă importantă de matematică, "aflarea unui algoritm eficient de descompunere în factori primi a numerelor intregi", are avantajul că rezolvarea celei din urmă nu poate fi ţinută secret, cum s-ar putea întâmpla cu rezolvările problemelor de criptografie. Evident nimeni nu poate garanta că în lume nu există astăzi un astfel de algoritm, şi prin urmare este periculos să implementăm un criptosistem bazat numai pe RSA.

Presupunem că Oscar cunoaşte descompunerea $n = pq$, atunci el ştie că $\varphi(n) = (p-1)(q-1)$ şi află uşor d astfel ca $de \equiv 1 \bmod \varphi(n)$.

Putem arăta acum că este adevărat şi reciproc, adică este posibil să calculăm factorii primi p şi q ai lui n ştiind pe n, e şi d. Fie

$$s = \max\{t \in \mathbb{N} \mid 2^t \text{ divide } ed - 1\}$$

şi

$$k = \frac{ed-1}{2^s}.$$

Pentru calculul factorizării lui n, avem nevoie de următoarea lemă.

Lema 1. *Pentru orice întreg a prim cu n, ordinul lui a^k, în grupul $(\mathbb{Z}_n*, \cdot)$ este un element al mulţimii $\{2^i \mid 0 \leq i \leq s\}$.*

Demonstraţie: Fie a un întreg prim cu n, atunci $a^{ed-1} \equiv 1 \bmod n$. De ce? Deoarece $ed \equiv 1 \,(\mathrm{mod}\, n)$, rezultă că $ed = 1 + t\varphi(n)$ iar congruenţa rezultă din Teorema lui Euler.

Pe de altă parte, deoarece $ed - 1 = k2^s$, aceasta implică $(a^k)^{2^s} \equiv 1 \bmod n$. Deci ordinul lui a^k este un divizor al lui 2^s.■

Algoritmul de factorizare a lui n ştiind pe e şi d se bazează pe următoarea teoremă.

Teorema 1. *Fie a un întreg prim cu n. Dacă ordinul lui a^k mod p este diferit de ordinul lui a^k mod q, atunci*

$$1 < \ cmmdc\ (a^{2^t k} - 1, n) < n \text{ pentru un } t \in \{0, 1, 2, \ldots, s-1\}.$$

Demonstraţie: Din lema 1, ordinele lui a^k mod p şi a^k mod q se află în mulţimea $\{2^i, \mid 0 \leq i \leq s\}$. Fără a restrânge generalitatea presupunem că ordinul lui a^k mod p este mai mare decât ordinul lui a^k mod q. Presupunem că ordinul lui a^k mod q este 2^t şi deci $t < s$.

Avem $a^{2^t k} \equiv 1 \bmod q$ dar $a^{2^t k} \not\equiv 1 \bmod p$. Atunci cmmdc $(a^{2^t k} - 1, n) = q$.■

Pentru a factoriza pe n vom proceda astfel:

1. Alegem la întâmplare un întreg a în mulţimea $\{1, \ldots, n-1\}$.

2. Calculăm $g = \text{cmmdc}\ (a, n)$.

3. Dacă $g = 1$, atunci calculăm $g = \text{cmmdc}\ (a^{2^t k} - 1 \bmod n, n)$ pentru $t = s-1, s-2, \ldots$ până când $g > 1$ sau $t = 0$.

4. Dacă $g > 1$, atunci $g = p$ sau $g = q$. Deci, factorizarea lui n a fost făcută şi algoritmul se încheie. Altfel algoritmul nu a reuşit cu baza a aleasă şi alegem alt a.

Dacă algoritmul nu a reuşit cu numărul a ales, atunci se rulează din nou cu un alt a. Se poate arăta că probabilitatea ca algoritmul să reuşească este de cel puţin $1/2$. Astfel, probabilitatea de succes după r iteraţii este de cel puţin $1 - 1/2^r$. Mai exact se poate arăta că numărul întregilor a din mulţimea $\{1, 2, \ldots, n-1\}$ primi cu n astfel încât ordinul lui $a^k \bmod p$ este diferit de ordinul lui $a^k \bmod q$ este cel puţin $(p-1)(q-1)/2$.

Exemplul 2. *Ne întoarcem la exemplul de mai sus.*
Avem $n = 253$*,* $e = 3$*, şi* $d = 147$*. Deci* $ed - 1 = 440 = 2^3 \cdot 55$*, adică* $s = 3$ *şi* $k = 55$*.*
Alegem $a = 2 \in \{1, 2, \ldots, n-1\}$ *şi se calculează* $gcd(2, 253) = 1$*.*
Pentru $t = 2$ *calculăm cmmdc* $(2^{220} - 1, 253) = 253$
Pentru $t = 1$ *calculăm cmmdc* $(2^{110} - 1, 253) = 253$*.*
Pentru $t = 0$ *calculăm cmmdc* $(2^{55} - 1, 253)$ *= cmmdc* $(207, 253) = 23$*.*
Obţinem în final descompunerea lui $n = 23 \cdot 11$*.*

Atacul modulului comun (vezi cursul scris pe foi).