Cátedra No. 13 : 8 de Octubre

Transcriptor: Gregorio González

1 Ataques a RSA

1.1 Calcular el Exponente de Decripción (Continuación)

En la clase anterior se discutió que un algoritmo que pudiese calcular el exponente de decripción de RSA podría utilizarse como subrutina para factorizar el módulo de RSA. A continuación mostraremos cuál es este algoritmo de factorización.

La idea de este algoritmo es utilizar la subrutina para encontrar una raíz cuadrada no trivial de 1 módulo n, $n = pq$, lo cual ya vimos que es equivalente a factorizar n.

Sea A el algoritmo que toma la clave pública de RSA, (e, n), donde e es el exponente de decripción y $n = pq$, p y q primos, y entrega el exponente de decripción d, i.e.

$$x^{ed} \equiv x \pmod{n}, \quad \forall x \in \mathbb{Z}_n.$$

A continuación describiremos un algoritmo probabilista que utiliza el algoritmo A como oráculo para factorizar n con probabilidad mayor o igual que $1/2$. Luego, si este procedimiento se repite m veces, factorizaremos n con probabilidad mayor o igual que $1 - 1/2^m$.

El algoritmo es el siguiente:

1. Se elige $w \in \{1, \ldots, n-1\}$ al azar.
2. $x = \text{mcd}(w, n)$
3. Si $1 < x < n$ parar, pues $x = p$ o $x = q$.
4. $d = A((e, n))$.
5. Calcular s tal que $ed - 1 = 2^s r$, donde r es impar.
6. Calcular $v = w^r \pmod{n}$.

7. Si $v \equiv 1 \pmod n$ parar, el algoritmo fracasa.
8. Mientras $v \not\equiv 1 \pmod n$ $\{v_0 = v, v = v^2 \pmod n\}$.
9. SI $v_0 \equiv -1 \pmod n$ parar, fracasó el algoritmo. En caso contrario, $\mathrm{mcd}((v_0+1), n) = p$ o q.

Analicemos brevemente el algoritmo. Primero que nada, si tenemos la suerte de elegir w múltiplo de p o q, entonces podemos factorizar n inmediatamente (paso 2). Si w es primo relativo con n, entonces calculamos w^r, w^{2r}, w^{4r},..., elevando al cuadrado sucesivamente, hasta que:

$$w^{2^t r} \equiv 1 \pmod n$$

para algún t. Como:

$$ed - 1 \equiv 2^s r \equiv 0 \pmod \phi)(n),$$

sabemos (por el lema de Euler) que $w^{2^s r} \equiv 1 \pmod n$ (paso 5). Luego, el ciclo "mientras"se acaba en a lo más s iteraciones. Al final de dicho ciclo hemos encontrado un valor de v_0 tal que $v_0^2 \equiv 1 \pmod n$ pero $v_0^2 \not\equiv 1 (mod n)$. Si $v_0^2 \equiv -1 \pmod n$ el algoritmo falló. En caso contrario v_0 es una raíz no trivial de 1 módulo n y podemos factorizar n.

No probaremos que el algoritmo tiene éxito con probabilidad al menos 1/2. Para ver dicha demostración, referirse al libro de Douglas R. Stinson[1].

1.2 Información Parcial

El otro ataque que discutiremos se basa en información parcial sobre el texto plano que podría "filtrarseª través de una encripción RSA.

Si un criptosistema permite encontrar información parcial sobre el texto plano encriptado, esto representa una debilidad en nuestro sistema, ya que si se quisiera utilizar el criptosistema con alguna modificación se corre el riesgo de perder toda seguridad. Por ejemplo, si un sistema permite hallar el tercer bit del texto plano encriptado, no podríamos usar el sistema para enviar sólo dos mensajes, sí y no, como un uno o un cero en el tercer bit y el resto con bits aleatorios "para despistar". En este caso un criptosistema que era muy seguro pasa a ser in'util. No siempre es tan significativo lo que se puede ganar con un sólo bit, pero definitivamente es un factor limitante al utilizar el criptosistema.

Los dos ejemplos de información parcial de RSA que consideraremos son:

[1] *"Cryptography, Theory and Practice"*, CRC Press, 1995, p'agina 149.

1. Dado $y = e_k(x)$, calcular $paridad(y)$, donde $paridad(y)$ entrega el bit menos significativo de x.
2. Dado $y = e_k(x)$, calcular $mitad(y)$, donde $mitad(y) = 0$ si $0 \leq x < n/2$ y $mitad(y) = 1$ si $n/2 < x \leq n$.

Probaremos que, dada $e_k(\cdot)$ la función de encripción de RSA, cualquier algoritmo que calcule $paridad(y)$ o $mitad(y)$ puede ser usado como oráculo para construír un algoritmo que obtenga el texto plano x. Lo que esto significa es que, dado un texto cifrado, calcular el bit menos significativo del texto plano es polinomialmente equivalente a calcular el texto plano completo. Con ello probamos que no se puede obtener dicha información parcial (bajo el supuesto que RSA es seguro).

Primero que nada, probaremos que calcular $paridad(y)$ es polinomialmente equivalente a calcular $mitad(y)$. En efecto, se tiene que:

1. $paridad(y) = mitad(y \cdot e_k(2^{-1}) \pmod n)$
2. $mitad(y) = paridad(y \cdot e_k(2) \pmod n)$

Probemos la segunda identidad.

$$\begin{aligned} paridad(y \cdot e_k(2)(modn)) &= paridad(e_k(x) \cdot e_k(2)) \\ &= paridad(e_k(2x \pmod n)) \end{aligned}$$

Esto es porque RSA es multiplicativa, i.e. $e_k(x_1) \cdot e_k(x_2) = e_k(x_1 \cdot x_2 \pmod n)$. Ahora bien, si $0 \leq x < n/2$ entonces $0 \leq 2x < n$, luego $2x \pmod n$ es par y $paridad(e_k(2x \pmod n)) = 0$. Si $n/2 < x \leq n$ entonces $n < 2x \leq 2n$, luego $2x \pmod n = 2x - n$, y como n es impar y $2x$ par, $2x \pmod n$ es impar. Entonces $paridad(e_k(2x \pmod n)) = 1$. Pero esto coincide con $mitad(y)$, luego se tiene la igualdad.

La otra identidad se prueba de manera análoga. Es posible calcular $x = d_k(y)$, dado un algoritmo que calcule $mitad(y)$. Observemos que:

$$\begin{aligned} mitad(e_k(x)) = 0 &\leftrightarrow x \in [0, n/2) \\ mitad(e_k(2x)) = 0 &\leftrightarrow x \in [0, n/4) \cup [n/2, 3n/4) \\ mitad(e_k(4x)) = 0 &\leftrightarrow x \in [0, n/8) \cup [n/4, 3n/8) \cup [n/2, 5n/8) \cup [3n/4, 7n/8) \end{aligned}$$

y así sucesivamente. Luego es posible encontrar x mediante una búsqueda binaria si podemos calcular $paridad(y)$ o $mitad(y)$.

2 Otros Criptosistemas

2.1 Criptosistema de Rabin

Ahora presentaremos el Criptosistema de Rabin, el cual es computacionalmente seguro contra todo ataque de texto plano elegido bajo el supuesto que es computacionalmente infactible factorizar n'umeros de la forma $n = p \cdot q$, donde p y q son primos.

Sea n el producto de dos primos distintos p y q, con $p, q \equiv 3 (mod 4)$. Sea $\mathcal{P} = \mathcal{C} = \mathbb{Z}_n$, y definamos

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

Para $K = (n, p, q, B)$ definimos

$$e_K(x) = x(x + B) \pmod{n}$$

y

$$d_K(y) = \sqrt{\frac{B^2}{4} + y} - \frac{B}{2}.$$

La clave pública es (n, B), y p y q se mantienen secretos.

La función de encripción no es inyectiva, así que la decripción es ambigua. En efecto, si w es una raíz cuadrada no trivial de 1, entonces para todo texto plano x hay tres valores distintos que producen el mismo texto cifrado, y son:

$$x,\ -x - B,\ w\left(x + \frac{B}{2}\right) - \frac{B}{2},\ -w\left(x + \frac{B}{2}\right) - \frac{B}{2}.$$

Por ejemplo:

$$\begin{aligned}
e_K\left(w\left(x + \frac{B}{2}\right) - \frac{B}{2}\right) &= \left(w\left(x + \frac{B}{2}\right) - \frac{B}{2}\right) \cdot \left(w\left(x + \frac{B}{2}\right) + \frac{B}{2}\right) \\
&= w^2\left(x + \frac{B}{2}\right)^2 - \frac{B^2}{4} \\
&= x(x - B) \pmod{n} \\
&= e_K(x)
\end{aligned}$$

En general, el receptor del mensaje no tiene forma de distinguir entre estos cuatro textos planos posibles, a menos que el texto plano contenga suficiente redundancia como para eliminar tres de las cuatro posibilidades.

Para decriptar debemos resolver:

$$x^2 + Bx \equiv y \pmod{n}.$$

o equivalentemente,

$$x_1^2 \equiv y + \frac{B^2}{4} \pmod{n}$$

donde $x_1 = x + \frac{B}{2}$.

Sea entonces $C = y + \frac{B^2}{4}$. Observemos que C es un residuo cuadrático si la encripción se hizo correctamente. Las soluciones de la ecuación $x_1^2 \equiv C \pmod{n}$ se pueden obtener mediante el Teorema Chino del Resto a partir de las soluciones de las ecuaciones:

$$\begin{aligned} x_1^2 &\equiv C \pmod{p} \\ x_1^2 &\equiv C \pmod{q} \end{aligned}$$

Cada una de estas ecuaciones tiene dos raíces. Estas se pueden combinar para obtener las cuatro soluciones de la ecuación original (por el TCR). Por el criterio de Euler sabemos que si se encriptó correctamente C es un residuo cuadrático módulo p y módulo q. Pero esto no basta para poder resolver las ecuaciones. En realidad el elemento clave para esto es el hecho que $p \equiv 3 \pmod{4}$ y $q \equiv 3 \pmod{4}$, pues en este caso tenemos un algoritmo determinista polinomial que nos permite calcular las raíces cuadradas de residuos cuadráticos módulo p o q. Las raíces de la primera ecuación vienen dadas por: $\pm C^{\frac{p+1}{4}} (mod p)$. En efecto:

$$\begin{aligned} \left(\pm C^{\frac{p+1}{4}}\right)^2 &\equiv C^{\frac{p+1}{2}} \pmod{p} \\ &\equiv C^{\frac{p-1}{2}} \cdot C \pmod{p} \\ &\equiv C \pmod{p}. \end{aligned}$$

esto debido a que C es residuo cuadrático, luego por el criterio de Euler se tiene que: $C^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. Análogamente, las raíces de la segunda ecuación son: $\pm C^{\frac{q+1}{4}} \pmod{q}$. Teniendo las soluciones de ambas ecuaciones se obtienen directamente las cuatro raíces de C módulo n usando el Teorema Chino del Resto.

Cabe hacer dos observaciones:

1. El adversario no sabe cuál de las cuatro es la decripción correcta, pero el receptor tampoco. Esto se resuelve acordando ciertas reglas de redundancia en el mensaje (i.e. el mensaje se transmite como dos copias concatenadas, el tercer bit se repite tres veces, etc.) para poder distinguir cuál de las cuatro raíces corresponde al mensaje original.

2. Es interesante el hecho que para el caso en que un primo p es congruente a 1 módulo 4 no se conoce ningún algoritmo polinomial determinista para determinar las raíces cuadradas de residuos cuadráticos módulo p, i.e. el hecho que $p \equiv 3 \pmod 4$ y $q \equiv 3 \pmod 4$ es vital para el algoritmo de decripción (Sin embargo, se conoce un algoritmo probabilista de tiempo polinomial que cumple este propósito).