Test de primalidad, aplicación a la criptografía

Prof. Marcela Wilder*

Se puede decir que la criptografía es tan antigua como la civilización, cuestiones militares, religiosas o comerciales impulsaron desde tiempos remotos el uso de escrituras secretas.

El primer caso claro de uso de métodos criptográficos se dio durante la guerra entre Atenas y Esparta, el cifrado se basaba en la alteración del mensaje original mediante la inclusión de símbolos innecesarios que desaparecían al enrollar la lista en un rodillo llamado escitala, el mensaje quedaba claro cuando se enrollaba la tira de papel alrededor de un rodillo de longitud y grosor adecuados.

Cuando Julio César enviaba mensajes a sus generales, él no confiaba en sus mensajeros, de manera que reemplazaba en sus mensajes cada A por una D, cada B por una E y así con todo el alfabeto. Solamente las personas que conocían la regla de "corrimiento por 3" podían descifrar sus mensajes.

Definiciones

Los datos que pueden ser leídos y entendidos sin ninguna medida especial son llamados *texto plano o texto en claro.*

El método de ocultar texto plano es llamado *encripción.*

El resultado de encriptar un texto plano es otro llamado texto cifrado que es incoherente e ilegible.

Se utiliza el encriptado de datos para garantizar que la información es oculta para cualquiera que no sea el destinatario.

El proceso de revertir el texto cifrado a su texto plano original se llama *desencripción.*

La *criptografía* es el conjunto de técnicas o procedimientos que alteran los símbolos de información sin alterar el contenido, convirtiendo a la información modificada en un conjunto de símbolos sin contenido para las partes que no disponen de las técnicas.

El conjunto de procedimiento que garantizan la seguridad de la información y utilizan técnicas criptográficas se denominan *criptosistemas.* El elemento fundamental de un criptosistema es la *"clave"*.

El criptoanálisis es el conjunto de metodologías y técnicas que permiten recuperar la información que ha sido previamente tratada por un procedimiento criptográfico, sin conocer "a priori" la técnica utilizada para la criptografía.

Objetivos de la criptografía

1 - Mantener la confiabilidad del mensaje, es decir, que la información contenida en el mensaje permanezca secreta.

2 - Garantizar la autenticidad del mensaje, el emisor y el receptor; esto es:
   • El mensaje recibido ha de ser realmente enviado.

---

* Profesora de Matemática (UBA). Docente de la Facultad de Ciencia y Tecnología de la Universidad de Palermo.

• El remitente y el destinatario han de ser realmente quienes dicen que son y no remitentes y/o destinatarios fraudulentos.

Métodos criptográficos

Existen dos tipos de métodos criptográficos: los *métodos simétricos* y los *métodos antisimétricos.*

En los métodos simétricos, la llave de cifrado coincide con la llave de descifrado, la llave debe permanecer secreta; el emisor y el receptor se ponen de acuerdo previamente o existe un centro de distribución de llaves. Estos métodos son propios de la criptografía clásica o criptografía de llave secreta.

En los métodos antisimétricos la llave de cifrado es diferente, la llave de cifrado es conocida por el público, mientras que la llave de descifrado solo por el usuario. Estos métodos corresponden a la criptografía de la llave pública introducida por Diffie y Hellman en 1976. Más tarde, comenzaron a aparecer implantaciones concretas de criptosistemas de clave pública.

Un sistema realmente elegante fue propuesto por los investigadores del M.I.T, Rivest, Shamir y Adleman en 1978 conocido como criptosistema RSA que ha resistido hasta hoy los ataques criptográficos.

Conceptos matemáticos previos al método RSA

Definición: sea $n \in N$, n es *primo* si y solo si tiene exactamente cuatro divisores. Por ejemplo, los números 2, 3, 5, 7, 11, etc son primos pues sus divisores son, el 1, el -1, el mismo número y su inverso, es decir, los divisores del número 3 son: 1, -1, 3 y -3.

Los números que no son primos (excepto el 1) se llaman compuestos, por ejemplo: los números 4, 6, 8, 9, 10, etc son compuestos.

Teorema fundamental de la aritmética: Todo número entero positivo se puede escribir en forma única (salvo en el orden de los factores) como producto de sus factores primos. Por ejemplo 12 = 2.2.3.

Máximo común divisor: Dados dos números enteros a y b, el máximo común divisor entre a y b, denotado como mcd(a, b), es el mayor número entero d que divide tanto a a como a b.

Congruencia: Dos números a y b en Z son congruentes módulo m, denotado $a \equiv b$ (m) si su diferencia es un múltiplo de m, es decir, si m es divisor de (a - b); o lo que es igual, a y b tienen el mismo resto al ser divididos por m.

Función de Euler: $\varnothing(n) = \#\{i \in N / 1 \leq i \leq n \text{ y } m.c.d.(1; n) = 1\}$, es decir, $\varnothing(n)$ es igual al número de números naturales menores o iguales a n, que no tienen ningún factor primo en común con n.

Propiedad: Si n = p.q con p y q primos distintos, entonces $\varnothing(n) = (p - 1). (q - 1)$.

Método RSA

Se eligen 2 números primos suficientemente grandes (aproximadamente 100 cifras) a los que llamaremos p y q.

Tomaremos n = p.q; por lo tanto, $\varnothing(n) = (p - 1). (q - 1)$.

Se elige un número e, primo, menor a $\varnothing(n)$.

Se calcula d, el único entero positivo, menor a $\varnothing(n)$ que cumple que $d.e \equiv 1\ (\varnothing(n))$.

Para codificar, se elige un número para cada letra. Llamaremos m a este número.

Se busca c tal que $m^e \equiv c\ (\varnothing(n))$. Por lo que el número m se codifica por medio del número c.

Para decodificar esto, se puede demostrar que $c^d \equiv m\ (\varnothing(n))$, lo cual nos permite recuperar el dato original.

Ejemplo:

Elegimos dos primos, en este caso pequeños p = 13 y q = 17.

Por lo tanto, n = p.q = 13.17 = 221.

$\varnothing(n) = (p - 1).(q - 1) = (13 - 1).(17 - 1) = 12.16 = 192.$

Elegimos un número primo e < 192, por ejemplo e = 11.

Calculamos d, el único número entero menor a 192 / $d.11 \equiv 1\ (192)$, este número d = 35.

Vamos a codificar la palabra CRIPTOGRAFIA.

Elegimos la siguiente numeración A→02; B→03; C→04; D→05, etc.

Por lo tanto la palabra CRIPTOGRAFIA se codificará: 04201018221708200207100 2, donde cada letra se identifica con un número de 2 cifras.

Entonces:

C: $4^{11} \equiv 166\ (221)$ R: $20^{11} \equiv 41\ (221)$ I: $10^{11} \equiv 173\ (221)$ P: $18^{11} \equiv 86\ (221)$

T: $22^{11} \equiv 198\ (221)$ O: $17^{11} \equiv 153\ (221)$ G: $8^{11} \equiv 70\ (221)$ R: $20^{11} \equiv 41\ (221)$

A: $2^{11} \equiv 59\ (221)$ F: $7^{11} \equiv 184\ (221)$ I: $10^{11} \equiv 173\ (221)$ A: $2^{11} \equiv 59\ (221)$

Entonces el texto cifrado será: 166041173086198153070041059184173059

Eligiendo bloques de tres dígitos, podremos decodificar así:

$166^{35} \equiv 4\ (221) \rightarrow C$ $41^{35} \equiv 20\ (221) \rightarrow R$ $173^{35} \equiv 10\ (221) \rightarrow I$ $86^{35} \equiv 18\ (221) \rightarrow P$

$198^{35} \equiv 22\ (221) \rightarrow T$ $153^{35} \equiv 17\ (221) \rightarrow O$ $70^{35} \equiv 8\ (221) \rightarrow G$ $41^{35} \equiv 20\ (221) \rightarrow R$

$59^{35} \equiv 2\ (221) \rightarrow A$ $184^{35} \equiv 7\ (221) \rightarrow F$ $173^{35} \equiv 10\ (221) \rightarrow I$ $59^{35} \equiv 2\ (221) \rightarrow A$

En realidad, en este ejemplo, hemos acabado construyendo un cifrador por sustitución mono-alfabético, con los problemas de seguridad que esto representa. Pero podemos, a partir de la secuencia numérica inicial, codificar bloques mayores de dos cifras, con lo que la dificultad para el análisis mediante frecuencias de aparición aumenta.

La seguridad del criptosistema RSA depende de la suposición de calcular d a partir de e, es equivalente a factorizar n. Si un criptosistema logra factorizar n (encontrar p y q), puede calcular $\varnothing(n)$ y encontrar el exponen- te secreto d a partir del exponente público e.

Hace pocos años se consiguió averiguar la clave que estableció uno de los creadores del RSA en 1977, tras codificar un mensaje con un número secreto de 129 cifras. Ellos aseguraron que nadie conseguiría descifrarlo, y para desafiar a los escépticos, acordaron pagar $ 100 a quien lograra descifrar la clave. En aquellos tiempos, se calculó que utilizando el mejor computador se tardaría 40 cuatrillones de años, por lo que teniendo en cuenta el período de vida estimado para el ser humano, era inimaginable intentar resolverlo.

En 1992, un profesor de matemática, Arjen Lenstra, de la universidad de Bellcore, tras oír una conversación entre un grupo de amigos que planeaban utilizar la potencia de internet para resolver ciertas claves codificadas, les explicó el reto de Rivest. Desde entonces, idearon un método que consistía en averiguar pares de números llamados vectores, para relacionarlos como lo estaban los dos números primos originales. Implementaron un programa

que permitía a todo el mundo bajarse a través de internet, con lo cual podían contar con la ayuda de miles de personas. Estos Crackers necesitaron medio millón de vectores para resolverlo. En total cooperaron más de 1600 máquinas de las más diversas características.

El final feliz se obtuvo el 26 de abril de 1994.

Rivest declaró más tarde, que su estimación del tiempo a emplear para descifrar su clave basada en la potencia de los ordenadores de los años 70. Atkins, quien logró descifrarlo, calcula que si se utiliza la clave recomendada actualmente para RSA de 1024 bits, con la potencia de las máquinas que existen hoy, él y su equipo, tardarían algunos pocos millones de años en resolverlo.

TEST DE PRIMALIDAD

El problema de la primalidad consiste en determinar, de forma aleatoria, números primos grandes. El problema surge, entre otras razones porque el criptosistema RSA necesita, para ser implementado, dos números primos p y q, de longitud grande (de alrededor de 100 dígitos). En general, para resolver este problema se recurre a los test de primalidad y de pseudoprimalidad.

Un test de primalidad es un criterio para decidir si un número dado es o no primo.

Supongamos que n es un entero impar grande. El test de primalidad más sencillo es el test de las divisiones sucesivas. El método consiste en tomar un número entero impar m y ver si divide o no a n. Si m no es ni n ni 1, entonces n es compuesto; en caso contrario, n pasa el intento de división por m. Los valores de m para asegurar todos los posibles casos debe ir desde 3 hasta el entero más cercano a raíz de n, es decir, m es impar con $2 < m \leq \sqrt{n}$.

El tiempo de computación para llevar a cabo el test anterior es demasiado elevado para llevarlo a la práctica, por lo que generalmente se recurre a otro tipo de pruebas.

Un test de pseudoprimalidad es un criterio para decidir, con un alto grado de probabilidad, si un número dado es o no primo.

Si el número n pasa por el test de pseudoprimalidad, entonces puede que sea primo; en caso contrario, es seguro que el número no es primo. Con este tipo de test se asegura si un número no es primo, mientras que la propiedad de ser primo es probabilística.

En general hay dos procedimientos para obtener números primos de forma más o menos rápida. El procedimiento depende, fundamentalmente, del tamaño del número que se desea. Así habitualmente el procedimiento seguido para obtener números primos de tamaño grande consiste en generar aleatoriamente, enteros impares y aplicarles un test de pseudoprimalidad.

Test de pseudoprimalidad:

1- Teorema de Tchebycheff:

La cantidad de números primos menores o iguales que x, $\pi(x)$, es asintóticamente equivalente a $x/\ln x$; es decir,

$$\lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1$$

Ejemplo: Según este resultado, la proporción de números primos entre el total de números impares de 100 cifras sería: Y por lo tanto, es de esperar que el número de test de primalidad necesarios para encontrar un número primo de 100 dígitos sea de 115.

2- Teorema (pequeño) de Fermat:

Si n es primo, entonces para cualquier b con mcd(b, n) = 1, se tiene que: $b^{n-1} \equiv 1\ (n)$.

Sin embargo puede que n no sea primo y que siga verificando la congruencia anterior.

Si n es un número compuesto impar y b es un número entero con mcd(b, n) de modo que se verifica la congruencia, entonces n se llama un pseudoprimo base b.

Sin embargo, puede suceder que un número n sea compuesto y verifique la congruencia. Los números que verifican la propiedad anterior para cualquier b se llaman números de Carmichael. Los primeros números de Carmichael son 565, 1105 y 1729. Estos números son bastante raros de encontrar; basta decir que hay 255 números de Carmichael menores a 100 millones.

Ejemplo. Consideremos el número n=91. Este número es un pseudoprimo de base b=3, puesto que $3^{90} \equiv 1(91)$; sin embargo, 91 no es un pseudoprimo de base b=2, puesto que $2^{90} \equiv 64\ (91)$.

Un primer test de pseudoprimalidad para determinar si un número n es primo consiste en hacer pasar a n por el test anterior para t valores de b elegidos independientemente. La probabilidad de que el número n no primo pase por los t test es de 2^{-t}.

$\frac{1}{115}$

3- Test de Solovay-Strassen:

Supongamos que n es un entero positivo impar y queremos saber si n es un número primo o compuesto. Para ello se eligen k enteros 0<b<n aleatoriamente. Para cada uno de estos números b se calculan los valores de $b^{(n-1)/2}$ y de $\left(\frac{b}{n}\right)$.

Si estos dos valores no son congruentes módulo n, entonces n es un número compuesto y el test se detiene. En otro caso, se prueba el siguiente valor de b. Si los valores aleatoriamente calculados son congruentes para k valores de b, independientemente elegidos, entonces hay una probabilidad menor a 2^{-k} de que n no sea primo.

Si n es un número entero impar y b es un entero con mcd(n, b)=1 y se verifica la congruencia anterior, entonces n se llama un pseudoprimo de Euler base b.

Donde $\left(\frac{b}{n}\right) = \begin{cases} 0 & si\ \ n\ \ divide\ \ a\ \ \ b \\ 1 & si\ \ b\ \ es\ \ un\ \ residuo\ \ cuadrático\ \ módulo\ \ n \\ -1 & si\ \ b\ \ no\ \ es\ \ un\ \ residuo\ \ cuadrático\ \ módulo\ \ n \end{cases}$

Sea $a \in Z$ y p primo, p> 2, diremos que a es un residuo cuadrático módulo p, si y solo si existe $x \in Z$ tal que $x \equiv a\ (p)$.

TEST DE PRIMALIDAD:

1- Criba de Eratóstenes:

Eratóstenes que vivió en el siglo III a.c., planteó la forma de listar los números primos hasta un cierto n, de la siguiente manera: se escriben los número naturales desde 2 hasta n, se suprimen todos los múltiplos de 2. Se busca el siguiente entero que no está tachado que, en este caso será el 3, este número es primo. Se suprimen de la lista todos los múltiplos de 3, el siguiente entero que está todavía en la lista es primo, en este caso será el 5, se suprimen todos sus múltiplos. Se continúa de esta manera hasta llegar a n. Claramente, este procedimiento, aunque muy ingenioso, sobre todo teniendo en cuenta la época en que fue creado, es muy trabajoso para encontrar números primos de 100 cifras.

2- Teorema de Wilson:

n es primo si y solo si, $(n-1)! \equiv -1\ (n)$. Este teorema, aunque muy eficaz, tiene la dificultad de la dimensión del número (n-1)! Cuando n es muy grande.

3- Teorema de Merssene:

Si $2^n - 1$ es primo, entonces n es primo. Si n es primo y $2^n - 1$ es primo, entonces $2^n - 1 = Mn$ se llama primo de Mersenne.

No se sabe si existen infinitos primos de Mersenne, el último conocido es $2^{216091} - 1$.

4- Test de Lucas:

M_p es primo si y solo si, $\dfrac{U_{2^p}}{U_{2^{p-1}}} \equiv 0\ (M_p)$

Con: u_n = término n-ésimo de la sucesión de Fibonacci

Sucesión de Fibinacci: {1,1, 2, 3, 5, 8, 13, 21, ...}

La criptografía y en especial los métodos que utiliza, está sujeta a grandes avances. En 1917 el algoritmo de Vigenere fue descrito como "irrompible" por la prestigiosa revista "Scientific American". Hoy en día, un mensaje con él codificado, no resistiría más de dos minutos de tiempo de computación. El mundo avanza rápido, y con él la matemática y los ordenadores. Lo que ayer parecía imposible, hoy es de simplicidad casi trivial. Los computadores cuánticos amenazan con ser capaces de romper cualquier clave en un tiempo muy pequeño.

Los matemáticos no han dicho la última palabra en lo que a algoritmos de factorización se refiere.

A pesar de que se han estudiado nuevas técnicas como son las basadas en el estudio de las curvas elípticas y las ya conocidas de los logaritmos discretos para hacer más difícil la labor del criptoanalista, el triunfo puede ser efímero.