Miscelánea Matemática **22**, *51–56 (1995)*

Pruebas de Conocimiento Cero

Jorge L. Ramírez Alfonsín
Instituto de Matemáticas,
Universidad Nacional Autonoma de México
México, D.F.

1 Introducción.

En una noche con mucha neblina un espía regresa al castillo después de haber cumplido su misión en el campo enemigo. Cuando estaba cerca de la entrada del castillo una voz tenue le pregunta ¿Cuál es la contraseña? Pero, ¿Es un amigo o un enemigo el que preguntó? ¿Cómo puede el espía mostrar que él sabe la contraseña sin revelársela a un posible impostor?

El anterior problema es conocido como *el dilema del espia.* Un método que ha sido propuesto para el intercambio de claves en este contexto es la *la prueba de conocimiento cero.* En esta nota, expondré dicho método.

2 Método.

Un *protocolo interactivo* consiste en dos algoritmos que llamaremos P (probador) y V (verificador) que leen una entrada de cadena común w de longitud $|w|$ y después de hacer sus cálculos respectivos se comunican en forma alternada para determinar si w tiene una propiedad específica.

El verificador es *acotado polinomialmente*, esto es, debe de terminar finalmente y su tiempo total de cálculo debe de estar acotado por un polinomio fijo en $|w|$. Cuando el protocolo termina, el verificador, V, tiene como datos de salida *aceptado* o *rechazado*, dependiendo si w tiene la propiedad (donde w pertenece a un cierto conjunto E). El verificador es *probabilístico*, esto es, le es permitido hacer selecciones al azar durante sus cálculos (digamos, de acuerdo a los resultados de lanzamientos de una moneda). El probador tiene un poder computacional "ilimitado", esto es, gran capacidad de memoria además de poder ser probabilístico.

Un *lenguaje* $L \subseteq E$ es un conjunto de cadenas. Un *sistema de pruebas interactivas* para L es un protocolo interactivo en el cual P ayuda a V a decidir si $w \in L$. Requerimos que con alta probabilidad el verificador esté en lo correcto cuando acepte o rechace la membresía de w en L. Más precisamente, para toda constante $c > 0$, para $w \in L$ (con $|w|$ suficientemente grande), la probabilidad (sobre todos los lanzamientos de una moneda) que V termine y acepte debe de ser al menos de $1 - |w|^{-c}$. Si $w \notin L$ entonces requerimos que ningún probador P^* sea capaz de convencer a V de lo contrario: esto es, para toda constante $c > 0$ y w suficientemente grande, y para cualquier protocolo interactivo (P^*, V), V rechace con probabilidad de al menos $1 - |w|^{-c}$.

En un sistema de prueba interactiva de *conocimiento cero*, cuando $w \in L$, P no revela ningún conocimiento adicional mas allá del hecho de membresía. Informalmente, "ningún conocimiento adicional" significa que el poder computacional de cualquier verificador V^* después de haber participado en el protocolo no es más de lo que V^* pudo haber ganado con simplemente haber supuesto que $w \in L$.

Para resolver el dilema del espía, usamos un sistema de prueba interactiva de conocimiento cero. Un auténtico probador P transmitirá partes del secreto y así V aceptará $w \in L$, pero la transmisión será inútil para verificadores falsos. Un probador impostor P^* causaría que V rechazará w con una alta probabilidad.

Veamos un ejemplo de como funciona este sistema. Sea L_3 el conjunto de cadenas que representan gráficas 3-coloreables, bajo un esquema fijo de representación de gráficas. Una gráfica es 3-coloreable si a sus vertices pueden ser asignados colores tal que dos vertices adyacentes no tengan el mismo color y no más de 3 colores son usados. Denotaremos por C al conjunto de colores. Ambos P y V tienen acceso a una *función de encripción* $f : C \times R \longrightarrow C^e$, donde R contiene cadenas largas (representando secuencias largas de lanzamientos de una moneda) y C^e es el conjunto de colores encriptados, esto es, $f(c_1, r) = c_1^e$ encripta el color c_1 mediante una cadena aleatoria de lanzamientos de monedas r dando como color encriptado c_1^e. La cadena de entrada w_G representa una 3-coloración particular de la gráfica G (G conexa con n vértices y m aristas). El secreto, sólo conocido por P, es una 3-coloración correcta de G; sea c_i el color de el vértice i bajo esta coloración (nótese que V y P conocen la gráfica G y el conjunto C).

Una prueba interactiva de conocimiento cero para $w_G \in L_3$ es esquematizada como sigue:

(1) P aplica una permutación π al azar a los colores (digamos que P cambia todos los vertices rojos a azul, todos los vertices azules a amarillo y todos los amarillos a rojo): Ahora cada vértice i tiene color $\pi(c_i)$. Para cada $i = 1, \ldots, n$, el probador forma al azar cadenas r_i a partir de muchos lanzamientos de una moneda y calcula $c_i^e = f(\pi(c_i), r_i)$. El color del vértice encriptado c_i^e es enviado a V.

(2) V guarda $c_1^e, \ldots, c_n^e$ y escoje al azar dos vértices adyacentes x y y y se los envia a P.

(3) P checa que (x, y) sea realmente una arista de G. Si no, P se detiene habiendo detectado un impostor V que no sabe el protocolo (no conoce la gráfica G). Si (x, y) es una arista de G, el probador manda los colores $\pi(c_x)$ y $\pi(c_y)$ y los valores r_x y r_y a V.

(4) V calcula $c'_x = f(\pi(c_x), r_x)$ y $c'_y = f(\pi(c_y), r_y)$; V checa que $c'_x = c_x^e$ y $c'_y = c_y^e$ y también que $\pi(c_x), \pi(c_y) \in C$ y que $\pi(c_x) \neq \pi(c_y)$. Si alguno de estos chequeos falla, entonces V se detiene y rechaza.

(5) Si el chequeo en (4) fué bueno, entonces P y V empiezan nuevamente en el paso (1). Si m^2 iteraciones de este protocolo es completado sin rechazo de V, entonces V se detiene y acepta w_G.

El anterior protocolo muestra un sitema interactivo de prueba para L_3. Si $w_G \in L_3$ entonces V acepta con probabilidad 1 después de m^2 iteraciones (ya que si $w_G \in L_3$ entonce V nunca va a detectar una falla de P). Si $w_G \notin L_3$ entonces el probador debe de mandar una coloración inválida (vertices adyacentes con el mismo color o bien que use más de tres colores) en el paso (1), la cual será detectada en el paso (4) con probabilidad al menos de $1/m$ en una iteración. La probabilidad que V se detenga y acepte después de m^2 pruebas al azar (sin detectar alguna falla de P) es de a lo más $(1 - 1/m)^{m^2}$. Dado que $(1 - 1/m)^{m^2} \approx exp(-m)$, se puede mostrar que dicha probabilidad es suficiente.

Para una prueba de interacción sería suficiente que P simplemente mandara los colores de los vértices a V; los demás pasos son necesarios para asegurar el conocimiento cero. En el paso (4) V "aprende" que los vértices x y y tienen diferente color, lo cual no es más que lo que aprendería si simple-

mente supusiera que $w_G \in L_3$. El verificador no gana información adicional en cada etapa porque los colores son permutados al azar y encriptados en cada iteración. Aún después de muchas pruebas V no tendrá idea de como es la 3-coloración de la gráfica (ver [1] y [2] para una prueba mas detallada).

3 Conclusiones.

La anterior prueba de conocimiento cero depende fuertemente en suponer que $f(\cdot,\cdot)$ es un esquema de *encripción segura*, en el sentido que no es factible desencriptarla derivando cada $\pi(c_i)$ a partir de c_i^e. Hasta el momento no es conocida la existencia de dicha función; sin embargo, son usadas funciones que se "creen" que son de encripción segura, es decir, funciones encriptadoras que no han podido ser "atacables". Por ejemplo, una función encriptadora muy conocida, y que no ha sido atacable, es la que multiplica dos números primos (grandes) y se basa en el hecho que no se conoce una forma de factorizar un número "eficientemente".

Sin embargo, si no queremos comprometer el protocolo, existen pruebas de conocimiento cero que no requieren de funciones de encripción en absoluto.

Por último diré que la noción de sistemas de prueba interactiva y complejidad de conocimiento en pruebas fueron creadas por S.Goldwasser, S.Micali y C.Rackoff (ver [3] y [4]).

Bibliografía

[1] O. Goldreich, S. Micali y A. Widgerson, *Proofs that yield nothing but their validity and a methodology of cryptographic protocol design*, Proceedings of the 27th Symposium on Foundations of Computer Science, 1986, 174-187.

[2] O. Goldreich, S. Micali y A. Widgerson, *Proofs that yield nothing but their validity or all lenguages in NP have zero- knowledge proofs*, Journal of Association for Computing Machinery, (38) 1 1991, 691-729.

[3] S. Goldwasser, S. Micali y C. Rackoff, *The knowledge complexity of interactive proof systems*, Proceedings of the 17th Symposium on Theory of Computing, 1985, 291-304.

[4] O. Goldwasser, S. Micali y C. Rackoff, *The knowledge complexity of interacticve proof systems*, SIAM Journal of Computing (18) 1, 1989, 186-208.