GRAFOS EXPANSORES EN CRIPTOGRAFÍA - Parte II

Expansion graphs in cryptography – Part II

DARÍO ALEJANDRO GARCÍA[1]

Recibido:16 de septiembre de 2018. Aceptado:16 de octubre de 2018
DOI: http://dx.doi.org/10.21017/rimci.2019.v6.n11.a57

Resumen

Los grafos expansores son una clase de grafos que tienen dos propiedades aparentemente contradictorias: son esparsos y bien conectados. Estos grafos tienen aplicaciones importantes en ciencias de la computación, tales como la construcción de configuraciones de redes optimizadas y, como veremos más tarde, la construcción de buenas funciones de resumen basados en grafos.

Palabras clave: Criptografía, matemáticas, ciencias de la computación, seguridad.

Abstract

Expansion graphs are a class of graphs that have two apparently contradictory properties: they are sparse and well connected. These graphs have important applications in computer science, such as the construction of optimized network configurations and, as we will see later, the construction of good graph-based summary functions

Keywords: Cryptography, mathematics, computer science, security.

I. Introducción

Un grafo expansor es una clase de grafos que tiene dos propiedades aparentemente contradictorias: son esparsos y bien conectados. Estos grafos tienen aplicaciones importantes en ciencias de la computación, tales como la construcción de configuraciones de redes optimizadas y, como veremos más tarde, la construcción de buenas funciones de resumen basados en grafos.

II. Expansión en grafos

Un grafo expansor es un grafo en el que cada subconjunto S de vértices está conectado con muchos vértices en su complemento $\bar{S}$. Grafos expansores y grafos esparsos tienen muchas propiedades útiles: diámetro pequeño, alta conectividad, y alto número cromático [1].

Supongamos que un grafo G = (V,E) tiene n vértices. Considere un subconjunto S de vértices V en G, y su complemento $\bar{S}$. La frontera de S, denotada ∂S, se define como el conjunto de aristas (v, w) ∈ E tal que v ∈ S y w ∈ $\bar{S}$.

Definición 1. El coeficiente de expansión de un grafo G = (V, E) con n vertices esta dado por:

$$h(G) = \text{mín}\left\{\frac{|\partial S|}{|S|} : S \subseteq V,\ 1 \leq |S| \leq n/2\right\}$$

El coeficiente de expansión nos da una idea de qué tan conectado está el grafo. Un valor mayor h(G) significa un valor más alto entre la proporción de la frontera de S y el tamaño de S. Como definimos los conjuntos tal que $|S| \leq \frac{n}{2}$, esto implica que los conjuntos 2 de vértices que

1 Matemático de la Universidad Nacional de Colombia, Magíster en Matemáticas y Doctorado en Matemáticas de la Universidad de los Andes. Postdoctorado/Estancia postdoctoral UNIVERSITE CLAUDE BERNARD LYON 1 Mathématiques - Institut Camile Jordan. Postdoctorado/Estancia postdoctoral UNIVERSITY OF LEEDS. Marie Curie Fellowship. Correo electrónico: dagarciar@gmail.com

contienen menos de la mitad de los vértices estarán bien conectados con conjuntos grandes de vértices. En otras palabras, G es un grafo *altamente conectado*.

Por ejemplo, para un grafo completo con n vértices, cada vértice en un subconjunto $S \subseteq K_n$ está conectado con todos los vértices en su complemento, por lo cual:

$$h(G) = \min\left\{\frac{|\partial S|}{|S|} : S \subseteq V,\ 1 \le |S| \le n/2\right\}$$
$$= \min\left\{\frac{|S| \cdot (n - |S|)}{|S|} : S \subseteq V,\ 1 \le |S| \le n/2\right\}$$
$$= \min\{n - |S| : S \subseteq V,\ 1 \le |S| \le n/2\} = \lceil \frac{n}{2} \rceil.$$

Sin embargo, como mencionamos anteriormente, es importante que los grafos expansores sean también esparsos. Por ejemplo, el grafo completo K_n estará bien conectado pero no será esparso. Más aún, en vez de restringirnos a un grafo fijo tendremos que usar familias de grafos expansores construídos con el siguiente criterio:

Definición 2. Una sucesión de grafos k-regulares $\{G_i\}$ $i \in N$ cuyos tamaños crecen con i se dice una *Familia de Grafos Expansores* si existe $\varepsilon > 0$ tal que $h(G_i) \geq \varepsilon$ para todo i. [2].

Así, cuando hablamos de grafos expansores k-regulares, usualmente nos referimos a una colección infinita (o familia) de grafos k-regulares que satisfacen las propiedades de la Definición 2. La idea es que la familia de expansores nos permitirá construír grafos arbitrariamente grandes que son esparsos y bien-conectados.

Ejemplo 1. Las siguientes son familias de grafos expansores:

1. La familia de grafos 8-regulares G_m para cada entero m es una familia de grafos expansores. El conjunto de vértices es $V_m = Z_m \times Z_m$. El vértice (x,y) estará conectado con los vértices $(x+y,y)$, $(x-y,y)$, $(x,y+x)$, $(x,y-x)$, donde $(x+y+1,y)$, $(x- y+1,y)$, $(x,y+x+1)$ y $(x,y-x+1)$, todas las operaciones se realizan módulo m.

2. La família de grafos 3-regular en p vértices por cada primo p dad por el conjunto de vértices $V_p = Z_p$, y donde ponemos un vértice x conectado con $x + 1$, $x - 1$, x^{-1}, con todas las operaciones realizadas módulo p, y declaramos que el inverso de 0 es 0.

Cuando aplicamos grafos expansores en ciencias de la computación, estaremos interesados en una construcción explícita de familias de grafos expansores y en la eficiencia de dichas construcciones. El desempeño de un algoritmo que emplea grafos expansores depende – al menos parcialmente – en qué tan eficiente es la construcción de los grafos.

Existen dos niveles naturales de eficiencia a considerar en la construcción de dichos grafos. En el primero requerimos que un grafo de n vértices pueda ser generado "desde el principio" en tiempo polinomial en n. En la versión más fuerte, podríamos solicitar que el conjunto de vecinos de un vértice dado pueda ser computado en tiempo polinomial en la longitud de la descripción del vértice, que es usualmente polinomial en log n.

Definición 3. Sea $F = \{G_i\} i \in N$ una familia de grafos expansores, donde G_i es un grafo k-regular en ni vértices y los enteros $\{n_i\}$ crecen con i de tal forma que $n_i < n_{i+1} \leq n_i^2$.

1. La familia *F* se dice *ligeramente explícita* si existe un algoritmo que genera el grafo j-ésimo G_j de la familia en tiempo polinomial en j. Esto es, G_j puede computarse entiempo$<A\ j^B$ para constantes A,B>0.

2. La familia *F* es *fuertemente explícita* si existe un algoritmo tal que con un input entero i, un vértice $v \in V(G_i)$ y un entero $m \in \{1, \ldots, k\}$ calcula el m-ésimo vecino de v en el grafo Gi. Este algoritmo debería ser de tiempo polinomial en la longitud del input (el número de bits necesarios para expresar la tripla (i, v, m)).

La construcción explícita de grafos expansores es importante, pero también debemos verificar

que los grafos escogidos sean buenos expansores. Por esta razón, es importante estar en la capacidad de calcular el coeficiente de expansión de la familia de grafos escogida. Sin embargo, para la mayoría de las familias que se escojan encontrar el valor mínimo de $|\partial S|/|S|$ para todos los subconjuntos $S \subseteq V$, $1 \leq |S| \leq n/2$ puede ser una tarea difícil y demorada, ya que el número total de subconjuntos que satisfacen los criterios a verificar son

$$N = \sum_{k=1}^{n/2} \binom{n}{k}$$

Esto es, el número total de casos a verificar es aproximadamente la mitad del número total 2^n de subconjuntos de V. Como sabemos que el número total de subconjuntos de un conjunto de tamaño n es 2^n, tenemos que $N \approx 2^{n-1}$. Por lo tanto, el número de subconjuntos que hay que revisar en cada cálculo de h(G) crece exponencialmente con n, lo que hace que encontrar h(G) sea computacionalmente inviable para un entero n suficientemente grande.

A. *Teoría Espectral de grafos*

Un grafo G = (V, E) en n vértices pueden expresarse en términos de una matriz n × n con filas y columnas indexadas por V y cuyas entradas expresan el número de aristas entre los diferentes vértices. Esto motiva la siguiente definición:

Definición 4. Sea G = (V,E) un grafo no-dirigido en n vértices. La matriz de adyacencia AG es la matriz n × n cuyos elementos se determinan por $a(i, j) = |\{(v_i, v_j) \in E\}|$

Por ejemplo, las matrices de adyacencia del ciclo C6 y del grafo completo K5 son respectivamente:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ y } \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Como el grafo G es no-dirigido, las aristas en E son pares no-ordenados, por lo que tenemos a(i, j) = a(j, i). Por lo tanto, AG es una matriz simétrica, y sus valores propios son todos números reales. En otras palabras, todos los valores $\lambda_m(G)$ que satisfacen la relación $A_G v_m = \lambda_m(G) v_m$, donde $v_m \in R^V \setminus \{0\}$ son reales. Dicho vector $v_m \neq 0$ se conoce como vector propio de A_G correspondiente al valor propio λ_m.

Para la matriz simétrica A_G, los vectores propios pueden escogerse para formar una base ortonormal de R^V, y el conjunto de valores propios de A_G se conoce como el espectro del grafo G. En el caso de grafos k-regulares, los valores propios de A_G satisfacen algunas propiedades interesantes:

Teorema 1. Sea G = (V,E) un multigrafo no-dirigido k-regular, y sea A_G su matriz de adyacencia correspondiente. Sean $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda n$ los valores propios reales de A_G. Entonces,

1. $\lambda_1 = k$ y $\lambda_n \geq -k$.
2. $\lambda_2 = k$ si y sólo si G no e sconexo.
3. $\lambda_n = -k$ si y sólo si al menos una de las componentes conexas de G es un grafo bipartito.

Definición 5. Si G = (V,E) es un grafo no-dirigido con valores propios $\lambda_1 \geq \lambda_2 \geq$

$\cdots \geq \lambda_n$ definimos la brecha espectral como $\Delta(G) := \lambda_1 - \lambda_2$. Eso nos lleva a la siguiente relación entre h(G) y λ_2 [3]:

Teorema 2. (Desigualdad de Cheeger's) El coeficiente de expansión h(G) de un grafo k-regular se relaciona con su brecha espectral Δ (G) por la desigualdad

$$\frac{\Delta(G)}{2} \leq h(G) \leq \sqrt{2k \cdot \Delta(G)}$$

Como sabemos que $\lambda_1 = k$ para un grafo k-regular, también podemos escribir la desigualdad como:

$$\frac{k-\lambda_2}{2} \leq h(G) \leq \sqrt{2k \cdot (k-\lambda_2)}$$

Fig. 1. El grafo de Petersen, un ejemplo de grafo de Ramanujan.

B. Cotas en los valores propios y grafos de Ramanuja

El Teorema 2 nos lleva a plantearnos una pregunta concerniente al tamaño de la brecha espectral. Sabemos que la brecha espectral de un grafo k-regular con n-vértices depende de ambos k y n, pero ¿exactamente qué tan grande puede ser la brecha espectral? La respuesta a esta pregunta no depende únicamente de k y n, sino también de la relación entre ellos. En el contexto de grafos expansores, estaremos interesados en el caso en que k es fijo y n es un entero con n >> k.

Definición 6. Supongamos que G es un grafo k-regular en n-vértices, y sean $k = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -k$ los valores propios reales A_G. Definimos $\lambda(G) = m\acute{a}x\,|\lambda_i| < k\ |\lambda_i|$.

Dicho de otra forma, λ(G) es el máximo de los valores propios de A_G, excluyendo los valores ±k. La cota para λ(G) está dada por un resultado que se debe a N. Alon y C. Boppana:

Teorema 3. (Alon-Boppana) Sea $\{G_i\}i\in N$ una familia infinita de grafos k-regulares conexos en n-vértices, donde k es fijo y n crece con i. Entonces, para todo i, $\lambda(G) \geq 2\sqrt{k-1} - o(1)$, donde o(1) es una función que tiende a cero para cada valor fijo de k y $n \to \infty$.

Esta definición nos da una clasificación de grafos como el de la Fig. 2, donde la cota de Alon-Boppana es óptima:

Definición 7. Un grafo finito k-regular G se dice grafo de Ramanujan si $\lambda(G) \leq 2\sqrt{k-1}$.

Trabajos de Lubotzky, Phillips, y Sarnak [4] muestran que la construcción explícita de familias infinitas de *grafos de Ramanujan* k-regulares son posibles siempre que k − 1 sea primo. Ellos usaron la conjetura de Ramanujan en la demostración, lo que le dió origen al término grafos de Ramanujan. El trabajo de Morgenstern [5] extiende la construcción a familias infinitas de grafos de Ramanujan k-regulares en el caso en que k − 1 sea una potencia de primo [2]. También es posible mostrar que la cota de Alon-Boppana y la definición de grafos de Ramanujan definen una clase de grafos donde la brecha espectral es casi tan grande como es posible.

III. Expansión en grafos de Cayley

Las ideas de grafos expansores que hemos explorado anteriormente en el caso general pueden aplicarse de manera relativamente efectiva en el caso específico de grafos de Cayley. Una familia infinita de grupos $\{G_n\}$ puede convertirse en una familia de expansores: sis existe una constante k y un conjunto de generadores Sn de tamaño k para cada G_n, la familia de grafos de Cayley $\{C_{Gn}, S_n\}$ es una familia de grafos expansores k-regulares.

No todas las clases de grupos pueden convertirse en expansores razonables. Por ejemplo, los grupos abelianos no pueden volverse expansores con conjuntos de generadores de tamaño acotado. Sin embargo, muchas de las familias de grupos simples pueden convertirse en familias de expansores, así como los grupos especiales lineales $SL_d(F_{pm})$ para cada d ≥ 2, m ≥ 1, y primo p.[6].

El estudio de propiedades de expansión en un grafo de Cayley construído a partir de un grupo abeliano G es equivalente al estudio de sus caracteres, donde un *caracter* de un grupo G es un homomorfismo de grupos de G al grupo multiplicativo de un campo, que en nuestro caso es el campo de números complejos C.

Proposition 1. Sea A la matriz de adyacencia normalizada de un grafo de Cayley $C_{G,S}$. Sea χ un cáracter de G. Entonces el vector $(\chi(g))_{g\in G}$ es un valor propio de A con valor propio $\frac{1}{|S|}\sum_{s\in S}\chi(s)$

Este enfoque puede generalizarse a grupos no abelianos G mediante el estudio de las representations de grupos, donde una representación es un homomorfismo de grupos que va de G a grupos de matrices sobre C.

El concepto de expansión en grafos se traslada a teoría de grupos por medio de la *constante de Kazhdan.*

Definición 7. La representación regular r de un grupo G es la representación que a cada $g \in G$ le asocia la matriz de tamaño $|G|$ cuya entrada correspondiente a $(u, g \cdot u)$ es 1 para todo u, y 0 en otro caso.

Definición 8. La constante de Kazhdan para G y S se define como

$$K(\mathcal{G},S) = \min_{v\in\mathbb{C}^{|\mathcal{G}|},v\perp 1} \max_{s\in S} \frac{||r(s)v - v||^2}{||v||^2}$$

Para un grupo G y un conjunto simétrico $S \subseteq G$ de tamaño k, la constante de Kazhdan K(G,S) se relaciona con la brecha espectral del grafo de Cayley $C_{G,S}$: [6]

C. Grafos expansores y grafos aleatorios

Los grafos expansores pueden ser estudiados desde una perspectiva algebraica o combinatórica, pero también pueden estudiarse de forma probabilística y estadística. Por ejemplo, como veremos en la siguiente sección, caminatas aleatorias en grafos expansores tienen importantes aplicaciones en criptografía.

Un grafo aleatorio en n vértices se construye a partir de un conjunto de n vértices aislados, y se desarrollan sucesivamente adquiriendo aristas de forma aleatoria. El principal objetivo al construír dichos grafos es determinar en qué momento de la evolución de un grafo aleatorio una propiedad específica se cumplirá con alta probabilidad ([9]). El modelo de grafos aleatorios más estudiado es el modelo de Erdős-Rényi (cf. [8]), que puede presentarse de dos maneras: la primera variante es la G(n,m), donde se escoge aleatoriamente un grafo del conjunto de todos los grafos con n vértices y m aristas. La segunda variante, G(n, p), construye el grafo en n vértices escogiendo las aristas independientemente con probabilidad p. En esta segunda variante, un grafo con n vértices tendrá exactamente r aristas con probabilidad $p^r(1-p)^{(n/2)-r}$.

El siguiente resultado conecta grafos expansores y grafos aleatorios:

Lema 1. Sea G = (V, E) un grafo k-regular con n vertices. Entonces para todo $S, T \subseteq V$ tenemos que

$$\left||E(S,T)| - \frac{k|S||T|}{n}\right| \leq \lambda(G)\sqrt{|S|\cdot|T|}$$

Este resultado muestra que un valor pequeño en el segundo valor propio de un grafo implica que sus aristas están "dispersas". La parte izquierda de la anterior desigualdad mide la desviación del número de aristas entre S y T con respecto al número de aristas esperadas entre S y T en un grafo aleatorio con densidad de aristas k/n. Un valor pequeño de $\lambda(G)$ (o una brecha espectral alta) –que es el caso en grafos expansores– implica que la desviación entre estos dos valores es pequeña, lo que convierte a G en un grafo casi-aleatorio [2].

Existe también un converso de este resultado. Cuando la brecha espectral de un grafo k-regular G es mucho más pequeña que k, las cotas superior e inferior del Teorema 2 difieren considerablemente. Un converso del Lema 1 captura mejor la brecha espectral [2]:

Lema 2. Sea G = (V,E) un grafo k-regular con n vértices, y supongamos que la desigualdad

$$\left|E(S,T)| - \frac{k \cdot |S| \cdot |T|}{n}\right| \leq p\sqrt{|S| \cdot |T|}$$

se cumple para cualesquiera conjuntos disjuntos S, T y algún valor positivo p. Entonces $\lambda(G) \leq O(p \cdot (1 + \log(k/p)))$. Esta cota es óptima.

IV. Caminatas aleatorias y Hashes Expansores

Muchas de las aplicaciones de grafos expansores se concentran en caminatas aleatorias en grafos. Para un grafo expansor k-regular G, una caminata aleatoria consiste en empezar en un vértice y moverse a uno de sus k-vecinos de forma aleatoria. Este proceso se repite en el nuevo vértice, y cada una de los movimientos se escoge de forma independiente a las decisiones anteriores.

Un aspecto interesante es que las caminatas aleatorias de longitud t en un grafo expansor es casi equivalente a escoger aleatoria e independentemente un conjunto de cardinalidad t, esto debido al hecho que los grafos expansores son altamente conectados. Computacionalmente, esto significa que se requieren pocos bits aleatorios para escoger una caminata aleatoria de longitud t, por lo que aplicaciones que requieran escoger muestras aleatorias pueden realizarse en muy corto tiempo [2].

A. *Introducción a las caminatas aleatorias*

Un vector $p \in R^v$ se denomina un vector de distribución de probabilidad si sus coordenadas son no-negativas y se cumple

$$\sum_{i=1}^{n} p_i = 1$$

Por ejemplo, el vector de probabilidad que corresponde a la distribución uniforme sobre {1,...,n} se denota como u = (1/n,..., 1/n). Una *caminata aleatoria* en un grafo G = (V,E) se inicia seleccionando un vértice v_1 con una distribución de probabilidad inicial p_1 sobre V . Esto induce una sucesión de distribuciones de probabilidad pi sobre V tal que la probabilidad de que $v_i = x$ es justamente $p_i(x)$. Para cada grafo finito conexo no-bipartito G, las distribuciones $\{p_i\}$ convergen a una distribución límite. Más aún, si el grafo G es k-regular, la distribución es simplemente u, [2].

Cuando realizamos un paso de la caminata aleatoria en un grafo k-regular G = (V, E), moviendónos del vértice i-ésimo al vértice (i + 1)-ésimo del camino, la distribución de probabilidad se actualiza a

$$p_{i+1} = \hat{A}_G p_i$$

donde

$$\hat{A}_G = \frac{1}{k}$$

A_G es la matriz de adyacencia normalizada. En general, decimos que la distribución de probabilidad luego de realizar una caminata de longitud t está dada por

$$p_t = (\hat{A})^t_G p_1$$

De forma equivalente, podríamos definir una caminata aleatoria en G = (V,E) es una cadena de Markov con conjunto de estado V y matriz de transición $\hat{A}$. Si estamos interesados en qué tan rápido dicha cadena de Markov converge a su distribución límite, el siguiente resultado establece que cuando la distribución límite es uniforme, la cadena de Markov converge a una tasa exponencial determinada por $\lambda_2(G)$ [3].

Teorema 4. Sea M una matriz de transición normal de una cadena de Markov con n estados, que tiene la distribución uniforme u como un punto estacionario, $M_u = u$. Para cualquier distribución inicial p tenemos que

$$||M^t \boldsymbol{p} - \boldsymbol{u}||_1 \leq \sqrt{n} \cdot \lambda_2(M)^t$$

La expresión $||\cdot||_1$ denota la norma L^1. Recordemos que la norma L^p determina la longitud de un vector en el espacio de Lebesgue Lp y se define como

$$||x||_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Dado que M puede ser cualquier matriz simétrica, podemos en particular tomar la matriz de adyacencia normalizada de un grafo expansor y considerar el caso cuando $M = \hat{A}_G$. Para un grafo k-regular G, tendremos

$$||\hat{A}_G^t \mathbf{p} - \mathbf{u}||_1 \leq \sqrt{n}\left(\frac{\lambda_2(G)}{k}\right)^t$$

Por lo tanto, usando el Teorema 2, podemos concluír que la tasa de convergencia está relacionada con el coeficiente de expansión de G:

$$||\hat{A}_G^t \mathbf{p} - \mathbf{u}||_1 \leq \sqrt{n}\left(1 - \frac{h(G)^2}{2k^2}\right)^t$$

Esto es, para una familia de grafos expansores, la tasa de convergencia de la caminata aleatoria correspondiente es exponentialmente rápida en la longitud de la caminata t.

Otra propiedad interesante de las caminatas aleatorias en grafos expansores es que la probabilidad de que permanezca en un subconjunto pequeño dado de vértices decrece exponencialmente con cada paso. El resultado general de cadenas de Markov necesario para esto es el siguiente:

Teorema 5. Sea X_0 una variable aleatoria uniformemente distribuida en n estados, y supongamos que existe una cadena de Markov $X_0, \ldots, X_t$ con matriz de transición M. Supongamos que la distribución uniforme $\cong$ es un punto estacionario de M. Sea B el conjunto de estados, y B(j) el evento que $X_j \in B$ para j = 0,1,...,t. Entonces la probabilidad de B(t) satisface

$$\Pr(B(t)) \leq \left(\lambda_2(M) + \frac{|B|}{n}\right)^t$$

Cuando aplicamos caminatas aleatorias en grafos expansores, X_0 será algún vértice del grafo escogido aleatoriamente con distribución uniforme. Usamos X_0 como punto inicial para la caminata aleatoria $X_0, \ldots, X_t$ donde X_t es el vértice en el t-ésimo paso. Dado un conjunto de vértices B, B(t) denotará el evento de que todos los puntos $X_0, \ldots, X_t$ están en B. Usando el resultado anterior tendremos que

$$\Pr(B(t)) \leq \left(\frac{\lambda_2(G)}{k} + \frac{|B|}{n}\right)^t$$

Esta probabilidad decrece exponencialmente siempre que

$$\frac{\lambda_2(G)}{k} + \frac{|B|}{n} < 1$$

Para una familia de grafos expansores, existe una constante $\varepsilon > 0$ tal que se obtendrá un decrecimiento exponencial para cualquier B que satisfaga $|B|/n < \varepsilon$. [3].

B. Funciones de resumen expansoras

Una función de resumen expansora es una función de resumen donde el input se usa como direcciones de una caminata aleatoria en un grafo expansor sin retorno, y el output de la función es el vértice final de la caminata. Para una función de resumen fija, la caminata comenzará en un vértice fijo en el grafo expansor determinado, y una familia de funciones de resumen podría ser definida si permitimos variar el vértice de inicio.

1) Construcción General

Para llevar a cabo una caminata en un grafo expansor no dirigido k-regular, el input de la función de resumen debe ser dividido en pedazos de tamaño c, de tal forma que $2^c = k - 1$. Comenzando en el primer vértice (que es seleccionado aleatoriamente), cada paso de la caminata escoge una arista saliendo del vértice al siguiente vértice, donde la elección de la arista a seguir está determinada por los siguientes c bits del input. Como la caminata es sin retorno, quedan $k - 1$ opciones para la siguiente arista en cada paso. Por otra parte, los grafos k-regulares no pueden tener retorno a menos que se admitan

aristas múltiples, así que el input de la función de resumen debe ser dividido en pedazos de tamaño c tal que $2^c = k$, pues hay k opciones para la siguiente arista en cada paso.

Como ya hemos explicado anteriormente, caminatas aleatorias en grafos expansores convergen de forma rápida a la distribución uniforme, por lo que el output de la función de resumen usada para la caminata será uniforme suponiendo que el input es uniformemente aleatorio. También sabemos que caminatas aleatorias en grafos expansores no tienden a quedarse en conjuntos pequeños de vértices por mucho tiempo, lo que implica que la construcción de la función de resumen tendrá un codominio significativo, de tamaño comparable al tamaño del grafo.

2) Resúmenes de Cayley

Así como hemos usado grafos expansores para la construcción de funciones resumen, podemos también construír funciones de resumen basadas en grafos de Cayley. Para construír un *resumen de Cayley* a partir de un grafo de Cayley $C_{G,S}$, sea $\{1,...,k\}^*$ un conjunto de secuencias $(m_1,m_2,...,m_l)$ de longitud arbitraria consistiendo de elementos de $\{1,...,k\}$. Fijando un valor inicial g_0 en G y un ordenamiento $\sigma : \{1,...,k\} \rightarrow S$, podemos determinar una función de resumen de Cayley $H : \{1, \ldots, k\}^* \rightarrow G$ definido por $H(\Lambda) = g_0$, $H(m_l) = g_0 * \sigma(m_l)$ and $H(m_l,m_2,...,m_l) = H(m_1,m_2,...,m_{l-1}) * \sigma(m_l)$. Los cómputos sucesivos de $g_0 * \sigma(m_l)$, $g_0 * \sigma(m_1) * \sigma(m_2)$, ... corresponden a la caminata v_{g0}, $v_{g0} * \sigma(m1)$, $v_{g0}\sigma(m1) * \sigma(m2)$, ..., $v_{g0\sigma}(m1) * \cdots\sigma(ml)$ en el grafo de Cayley $C_{G,S}$.

Si S es invariante bajo inversos, el grafo de Cayley correspondiente $C_{G,S}$ será no dirigido, y aplicaremos la construcción dividiendo el mensaje en pedazos de (k − 1) bits, definiendo H como $H(m_1,m_2,...,m_l) = H(m_1,m_2,...,m_{l-1})\sigma l$ con $\sigma 1 = \sigma(m1)$ y $\sigma i = (\sigma i-1,mi)$: $=\sigma(\sigma-1(\sigma-1) + mi \text{ mod} k)$, para$i \in \{2,...,l\}$.

C. Consideraciones de seguridad

Basados en nuestras consideraciones sobre grafos expansores, concluímos que los grafos usados en resúmenes expansores satisfacen los siguientes requerimientos:

(a) Gran expansión: Este requerimiento garantiza que los valores resumen de mensajes relativamente cortos están bien distribuidos en el conjunto de salida.

(b) Diámetro pequeño: Diámetro pequeño implica que cada vértice es el valor de salida de un mensaje corto.

(c) Girth grande: un girth grande garantiza que no hay colisiones en mensajes cortos y acota la distancia entre dos mensajes que colisionan. Este requerimiento puede no ser necesario si el vértice inicial se escoge aleatoriamente, pero para resúmenes de Cayley tener un girth grande es necesario.

(d) Eficiencia: Calcular los vecinos de cada vértice dado es eficiente.

(e) Resistencia a preimágenes y colisiones: Los siguientes problemas deben ser difíciles:

- Problema restringido de dos caminos: dado un vértice inicial seleccionado aleatoriamente en un grafo G, encontrar dos caminos de longitud a lo más l que comiencen en v_0 y terminen en el mismo vértice.

- Problema restringido de ciclos: Dado un vértice inicial v_0 seleccionado aleatoriamente en un grafo G, encontrar un ciclo en G de longitud a lo más l que pase por v_0.

- Problema del camino: Dado un vértice inicial v_0 y un vértice final v seleccionados aleatoriamente en un grafo G, encontrar un camino en G de longitud a lo más l que comience en v0 y termine en v.

- Problema de dos caminos: Dado un grafo G escogido aleatoriamente, encontrar dos caminos en G de longitud a lo más l

que comienzan y terminan en los mismos vértices.

- Problema del ciclo: Dado un grafo G escogido aleatoriamente, encontrar un ciclo en G de tamaño a lo más l.

Es interesante destacar que los problemas anteriores pueden trasladarse a problemas en teoría de grupos. Primero, definimos la longitud de un producto de elementos $g_0 g_1 \cdots g_{\mu-1}$ como μ. Este producto se dice un producto reducido si $g_i g_{i+1} \neq e$ para $0 \leq i \leq \mu - 2$. En particular, el problema del ciclo, el problema de los dos caminos y el problema del camino son equivalentes (respectivamente) a los siguientes tres problemas en teoría de grupos para resúmenes de Cayley [6]:

- Problema de factorización: Dado un grupo G, un subconjunto S de G, y un elemento g del grupo, encontrar un producto reducido de a lo más l elementos de S que sea igual a g, esto es, encontrar elementos $s_i \in S$ tales que

$$\prod_{0 \leq i < \mu} s_i = g$$

Con $s_i \in S$, $s_i s_{i+1} \neq 1$, y $\mu \leq l$.

- Problema del equilibrio: Dado un grupo G y un subconjunto S de G, encontrar dos productos reducidos de longitud a lo más l de elementos en S que sean iguales, esto es,

$$\prod_{0 \leq i < \mu} s_i = \prod_{0 \leq i < \mu'} s'_i$$

con $s_i , s'_i \in S$, $s_i s_i + 1$, $s'_i s'_{i+1} \neq 1$, y $\mu , \mu' \leq l$.

- Problema de representación: Dado un grupo G, un subconjunto S de G, encontrar un producto reducido de elementos de S de longitud a lo más l que sea igual al elemento unidad, esto es,

$$\prod_{0 \leq i < \mu} s_i = e$$

Con $s_i \in S$, $s_i s_{i+1} \neq 1$ y $\mu \leq l$.

D. Ejemplos de resúmenes expansores

1) Expansores de Zémor

El primer estudio de resúmenes expansores fue realizado por Gilles Zémor a comienzos de la década de los 90. Su esquema era un resumen en un grafo dirigido de Cayley, construído en el grupo $G = SL_2(F_p)$, el conjunto de matrices 2 × 2 sobre el campo matrices sobre el campo F_p con determinante 1, donde las operaciones de grupos son respectivamente la multiplicación y la inversión de matrices. El conjunto generador para el grafo es:

$$S_1 = \{A_1, B_1\} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\}$$

El cómputo de un resumen para un mensaje de longitud μ para C_G,S_1 requiere μ multiplicaciones por A ó B, y cada una de estas multiplicaciones requiere 2 sumas módulo p, lo que hace este esquema razonablemente eficiente en términos computacionales.

Sin embargo, este esquema ya no es tan seguro debido a un ataque que utiliza una estrategia de levantamiento: el problema se levanta de $SL_2(F_p)$ a $SL_2(Z)$ explotando el hecho de que A y B generan el conjunto $SL_2(Z+)$ en $SL_2(Z)$. Esto permite que el problema de representación pueda ser resuelto usando el algoritmo de Euclides. Para rectificar esto, se proponen otros dos conjuntos de generadores [6]:

$$S_2 = \{A_2, B_2\} = \{A_1^2, B_1^2\} = \left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\}$$

$$S_3 = \{A_3, B_3\} = \{A_1, A_1 B_1\} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

Estos cambios en A y B afectan la eficiencia en el cómputo del mensaje. Para C_G,S_2, cada bit del mensaje requiere dos sumas y dos multiplicaciones módulo p. Para C_G,S_3, cada bit requiere en promedio tres sumas módulo p. Aún no se conocen ataques para los esquemas que usan los conjuntos S_2 y S_3. [7].

2) *Funciones de resumen de Zémor-Tillich*

Jean-Pierre Tillich y Gilles Zémor proponen una nueva familia de funciones de resumen basadas en cálculos en un campo finito de característica 2 [9]. En vez de construír un resumen expansor usando el grupo $SL_2(F_p)$, decidieron usar $SL_2(F_{2n})$, el grupo de matrices 2 × 2 de determinante 1 sobre el campo finito $K := F_{2n}$, que también puede representarse como el campo cociente $K = F_2[X]/\langle P_n(X)\rangle$ donde $P_n(X)$ es un polinomio binario irreducible de grado n [6].

La función resumen de Zémor-Tillich tiene como parámetro de definición un polinomio irreducible $P_n(X)$ de grado n, tiene como punto inicial la matriz identidad de tamaño 2×2, y el siguiente conjunto generador:

$$S = \{A_0, A_1\} = \left\{ \begin{pmatrix} X & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} X & X+1 \\ 1 & 1 \end{pmatrix} \right\}$$

Ahora, consideremos el mapa

$$\begin{aligned} \pi : \{0,1\} &\longrightarrow \{A_0, A_1\}, \\ A_0 &\longmapsto 0, \\ A_1 &\longmapsto 1. \end{aligned}$$

Así, el código de resumen de un mensaje binario $m = m_0 m_1 ... m_{\mu-1}$ es el producto de matrices $H_{ZT}(P_n(X),m) := \pi(m_0)\pi(m_1)...\pi(m_{\mu-1})$, y por lo tanto es un elemento del grupo $SL_2(F_{2n})$ [9]. De forma más compacta, podemos calcular el valor resumen de Zémor-Tillich de una cadena de bits $m=m_0m_1 \ldots m_{\mu-1}$ como

$$H_{ZT}(P_n(X), m) := \prod_{i=0}^{\mu-1} \begin{pmatrix} X & 1 + m_i X \\ 1 & m_i \end{pmatrix}$$

Como estas funciones de resumen usan únicamente operaciones en un campo de característica 2, este proceso es más eficiente que la propuesta original de Zémor. A pesar de que existen ataques para estas funcione de resumen, el proceso permanece fundamentalmente fuerte si se escogen parámetros suficientemente grandes y genéricos.

3) *Funciones de resumen LPS*

Dados dos números primos l y p, donde l es un primo pequeño y p es relativamente grande tal que p, $l \equiv 1 \pmod 4$ y l es un residuo cuadrático módulo p (esto es, existe un entero x tal que $x^2 \equiv l \pmod p$). Denotaremos el grafo LPS con parámetros l, p como $X_{l,p}$.

Se utilizan grafos LPS para este esquema, ya que estos son grafos de Ramanujan, de gran girth y diámetro pequeño:

$$\begin{aligned} &\mathbf{girth}(\mathfrak{X}_\ell, p) \geq 4 \log_\ell p - \log_\ell 4, \\ &\mathbf{diam}(\mathfrak{X}_\ell, p) \leq 2 \log_\ell \left(\frac{p(p-1)(p+2)}{2} \right) + 2 \log_\ell 2 + 1. \end{aligned}$$

Los vértices de $X_{l,p}$ son matrices en $PSL_2(F_p)$, que es el grupo cociente de $SL_2(F_p)$ bajo la relación de equivalencia $M \sim -M$ para toda matriz M. Equivalentemente, podemos definir $PSL_2(F_p)$ como el grupo de matrices 2 × 2 sobre F_p con determinante cuadrado no nulo, modulo la relación de equivalencia $M_1 \sim \lambda M_2$, $\lambda \in F_p^*$ [6]. Una matriz M está conectada con las matrices gM, donde g's son las siguientes matrices definidas explícitas: dado un entero j tal que $j^2 \equiv -1 \pmod p$, hay exactamente 8(l+1) soluciones (g_0, g_1, g_2, g_3) para la ecuación

$$g_0^2 + g_1^2 + g_2^2 + g_3^2 = \ell.$$

Entre estas soluciones, podemos considerar aquellas para las cuales g_0 es impar y g_1, g_2, g_3 son pares. A cada una de estas soluciones, asociamos la matriz

$$g = \begin{pmatrix} g_0 + j g_1 & g_2 + j g_3 \\ -g_2 + j g_3 & g_0 - j g_1 \end{pmatrix}$$

Esto nos da un conjunto S de l + 1 matrices en $PGL_2(F_p)$, cuyos determinantes son cuadrados módulo p y por lo tanto yacen en el subgrupo de $PGL_2(F_p)$ de índice 2, que es precisamente $PSL_2(F_p)$. Así, podemos ver que los grafos LPS son grafos de Cayley, y $X_{l,p} := C_{SL2(Fp),S}$. Adicionalmente, el grafo no dirigido construído a partir de S es invariante bajo inversos. Como l es pequeño, el conjunto de matrices en S puede encontrarse de forma rápida usando una búsqueda exhaustiva. El grafo $X_{l,p}$ tiene $p(p^2 - 1)/2$ vértices y es (l + 1)-regular.

4) Funciones de resumen de Morgenstern

Los grafos de Ramanujan usados por Morgenstern [5] generalizan los grafos LPS para un primo impar $p \equiv 1 (\text{mod } 4)$ a cualquier q que sea una potencia de 2 a un exponente que sea otro primo. La aritmética en campos de característica 2 normalmente es más eficiente de implementar que aritmética en campo primo grande. Esto nos lleva a la propuesta de Morgenstern para funciones resumenes, que usa valores pequeños de q. [6].

Los grafos de Morgenstern para un número par q se define como sigue: sea q una potencia de 2 y sea $\varepsilon \in Fq$ tal que $f(x) := x^2 + x + \varepsilon$ es irreducible en $F_q[X]$. Sea. $P_n(X) \in F_q[X]$ un polinomio irreducible de grado n y considere el campo F_{qn} representado como $F_q[X]/P_n(X)$. Denotamos el grafo de Morgenstern como $\Gamma_{q,n}$. Así como los grafos LPS, los grafos Morgenstern son grafos de Ramanujan, tienen girth grande y diámetro pequeño [6]:

$$\mathbf{girth}(\Gamma_{q,n}) \geq \frac{2}{3}\log_q(q^n(q^{2n}-1)),$$
$$\mathbf{diam}(\Gamma_{q,n}) \leq 2\log_q(q^n(q^{2n}-1)) + 2.$$

Los vértices del grafo son elementos del grupo $PSL_2(F_{qn})$, que es el grupo de matrices 2×2 sobre F_{qn} con determinante cuadrado no nulo, modulo la relación de equivalencia $M_1 \sim \lambda M_2$, $\lambda \in F^*_{q\,n}$. Sea $j \in F^*_{q\,n}$ una raíz de f(x). El conjunto S se toma como $S = \{s_k\}_{k=0,\ldots,q,}$ donde

$$s_k = \begin{pmatrix} 1 & \gamma_k + \delta_k j \\ (\gamma_k + \delta_k j + \delta_k)X & 1 \end{pmatrix}, j = 0, \ldots, q;$$

Y $\gamma k, \delta k \in Fq$ son todas las q+1 soluciones en F_q de la ecuación $\gamma k^2 + \gamma k \delta k + \delta k^2 \varepsilon = 1$. Los grafos de Cayley $\Gamma_{q,n} := C_{PSL2(Fqn),}\, S$ son no-dirigidos puesto que cada sk tiene orden 2.

Referencias

[1] Terence Tao. Expansion in Finite Simple Groups of Lie Type. Graduate Studies in Mathematics. Vol. 164, American Mathematical Society. 2015.

[2] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their appli- cations. Bulletin of the American Mathematical Society 43, no. 4, 439-561. 2006.

[3] Emanuel Kowalski. An introduction to expander graphs. ETH Zúrich. Noviembre 2017.

[4] Alexander Lubotzky, Ralph Phillips, Peter Sarnak. Ramanujan graphs. Combinatorica 8, No. 3, 261-277. 1988.

[5] Moshe Morgenstern. Existence and explicit constructions of (q+1)-regular Ramanujan graphs for every prime power q. Journal of Combinatorial Theory. Series B 62, No. 1, 44-62. 1994.

[6] Cristophe Petit. On graph-based cryptographic hash functions Ph.D. thesis, Catholic University of Louvain, May 2009.

[7] Béla Bollobás. Random graphs. Cambridge Studies in Advance Mathematics. Cam- bridge University Press, 1985.

[8] Darío García. Grafos aleatorios y sus aplicaciones. Corporación Universitaria Repu- blicana. 2016.

[9] Jean-Pierre Tillich and Gilles Zémor, Hashing with SL2. Reporte técnico, Ecole Na- tionale Supérieure des Telecommunications. 1994.