Las Matemáticas en la evolución de la Criptografía

Santos González y
Consuelo Martínez

#CyberCamp17

Matemáticas y Criptografía

- **El objetivo de la charla es mostrar el papel esencial que han jugado las matemáticas en la evolución de la criptografía.**

- **¿Pueden seguir jugando un papel esencial en el futuro?**

- **¿Cuales son las líneas actuales que parecen llamadas a desempeñar un papel clave en la criptografía del siglo XXI?**

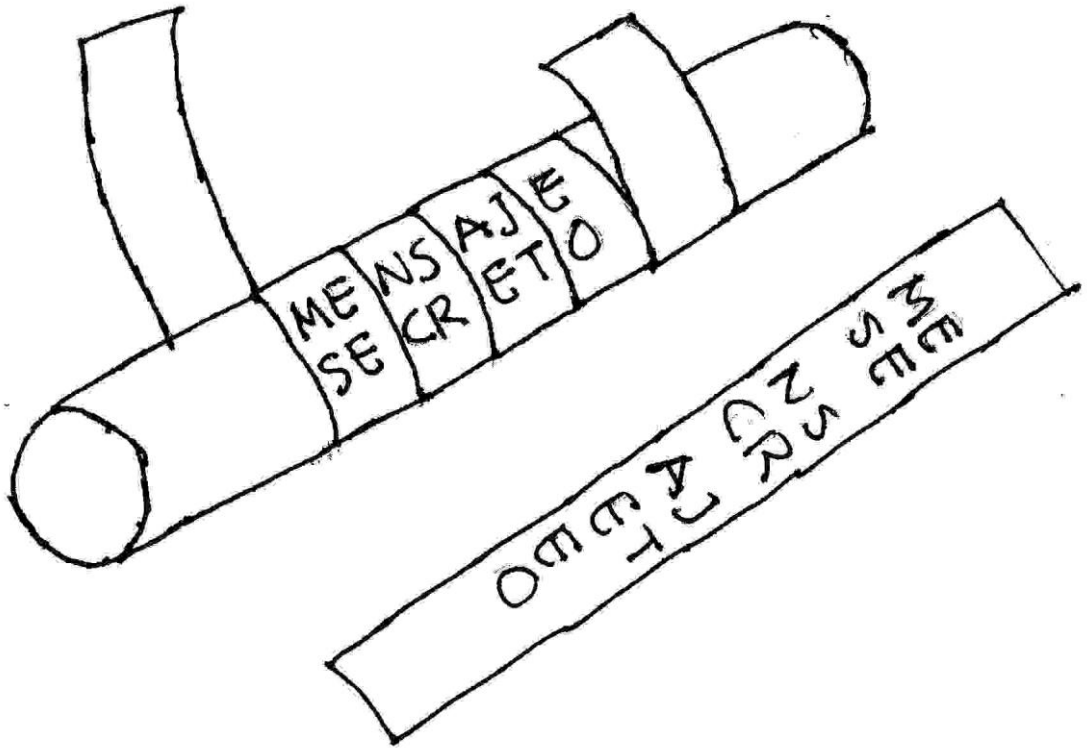
Necesidad de la criptografía

- La criptografía es un pilar básico de la ciberseguridad.
- *“La enorme utilidad de las matemáticas en la seguridad de la información está bordeando el misterio y no existe aplicación racional para ella” (Koblitz)*
- *La criptografía se remonta a varios siglos antes de Cristo.*
- *En su primera etapa tenia mas de arte o ingenio que de ciencia.*

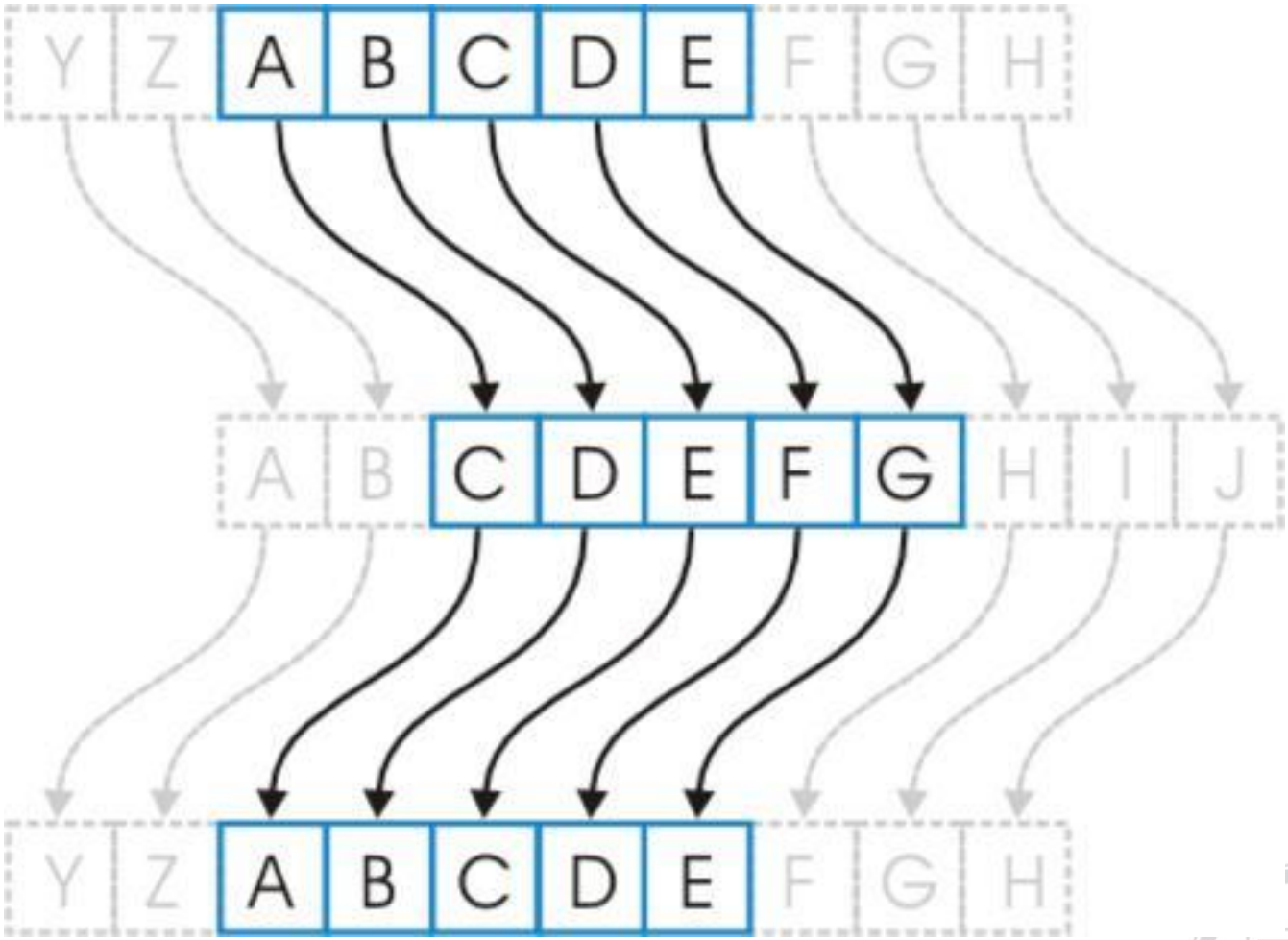
Breve recorrido histórico

- Escítala lacedemonia

Esparta / Persia



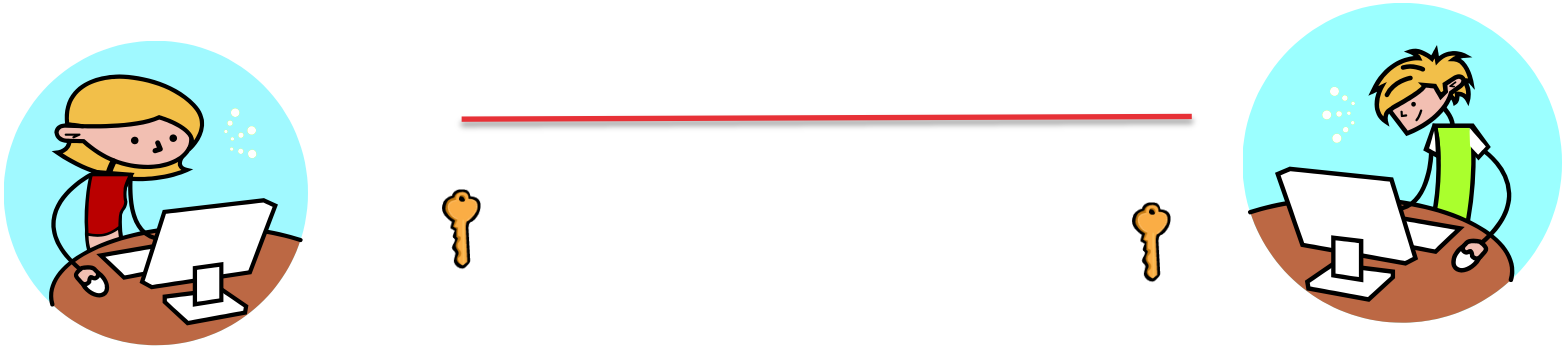
Cifrado de César



La seguridad absoluta no existe.

Criptografía simétrica (clave privada)

- **La criptografía usada hasta mediados del siglo XX.**
- **Fines militares o políticos.**



- **Principal problema la compartición de clave.**

Clave pública (criptografía asimétrica)

Revolución en el ámbito de la criptografía: internet, correo electrónico seguro, e-comercio, e-banca, …

Diffie y Hellman (1976)

Cada usuario A tiene dos claves:

clave pública f (cp_A)

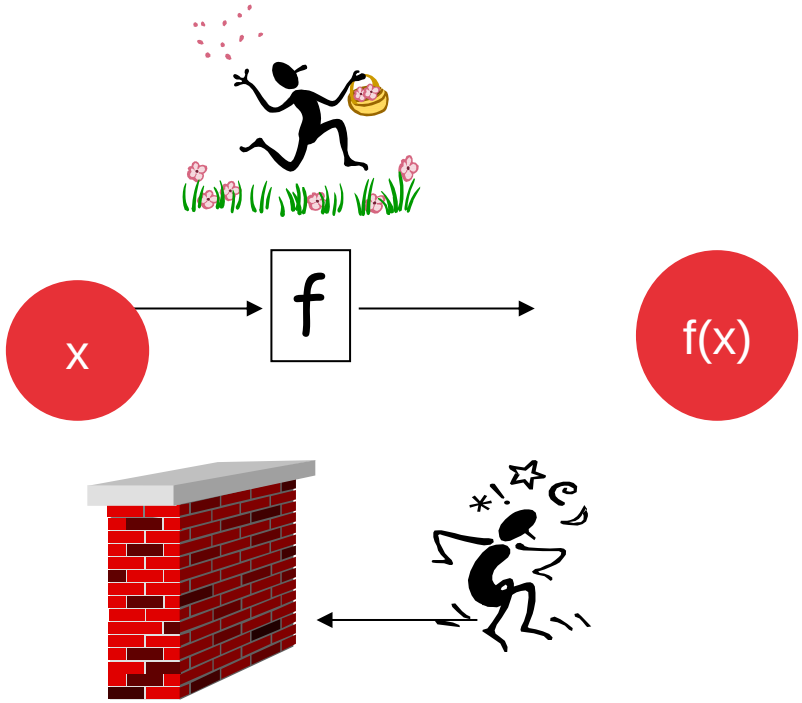
clave privada f^{-1} (dp_A)

La clave pública (conocida por todos) se usa para cifrar.

La clave privada (conocida sólo por A) se usa para descifrar.

Funciones de una via

Las matemáticas suministran funciones de una via, posibilitando la criptografía de clave pública.



Intercambio de clave de Diffie-Hellman

G=<g>

g^a

g^b

$(g^b)^a$ = g^{ab} = $(g^a)^b$

RSA

- **Su seguridad se basa en la dificultad de factorizar un número grande.**
- **Generación de Claves:**
  - Elegir primos grandes y distintos *p, q* y computar *n=pq.*
  - Elegir *e* que sea primo con $\phi(n) = (p-1)(q-1)$
  - Computar *d* tal que $de = 1 \bmod \phi(n)$
  - Clave pública (n,e)
  - Clave privada d
- **Cifrado: texto x se cifra como $c = x^e \bmod n$**
- **Descifrado: $x = E(x)^d = x^{de} \bmod n$**

Números de Fermat

- **F(n) = 2^(2^n) + 1**
- **F(0) = 3, F(1) = 5, F(2) = 17, F(3) = 257, F(4)= 65537**
- **Fermat conjeturó que todos estos números son primos.**
- **F(5) = 4294967297 = (641)(6700417) Euler**
- **F(6) = 18.446.744.073.709.551.617 = (59.649.589.127.497.217)(5.704.689.200.685.129.054.721),**
- **F(7) tiene 39 dígitos (se factorizó en 1970) = p(17).q(22),**
- **F(8) = p(16).q(62) tiene 78 dígitos,**
- **F(9) = p(49).q(99) tiene 155 dígitos,**
- **F(10) = p(8).q(10).r(40).s(252).**

El Gamal

- **Dificultad del problema del logaritmo discreto en algunos grupos.**
- **ElGamal: el grupo multiplicativo de un cuerpo finito.**
- **Criptografía de curva elíptica: grupo de una curva elíptica con coeficientes en un cuerpo finito. Permite claves mucho mas cortas.**
- **Tanto RSA como ElGamal tienen esquemas de firma asociados.**

La criptografía es básica para la seguridad ??

- **Páginas https – TLS (Transport Layer Security) que combina operaciones criptográficas --- confidencialidad, integridad y autenticidad.**

- **Servidor A envia un mensaje seguro al usuario B**
- **(Clave privada) Clave compartida, MAC(código de autenticación del mensaje) y etiqueta de autenticación.**
- **(Clave pública) B contacta con A que le envia la clave pública y cifra con ella una clave simétrica de un solo uso. Continúa como antes.**

Firmas digitales

- **¿Como se asegura B de que ha obtenido realmente la clave de A?**
- **Se usan firmas digitales emitidas por una tercera parte de confianza.**
- **En el caso de MAC, la etiqueta sólo la podían verificar A y B. En la firma digital cualquiera la puede verificar.**
- **Necesitamos procesos criptográficos sólidos y viables.**
- **¿Se usan otros problemas matemáticos aparte de la factorización de enteros y el problema del logaritmo discreto?**

Funciones hash criptográficas

- **Son funciones que asocian a una cadena de bits de longitud arbitraria una cadena de bits de longitud fija n.**
- **Se les exigen ciertas condiciones de eficiencia y seguridad (en general contrapuestas).**
- **Resistencia a colisiones, resistencia a la segunda preimagen, resistencia a una colisión objetivo.**
- No existe actualmente una familia de funciones hash que sea eficiente y demostrablemente segura
- **Se usan funciones hash que parecen seguras:** SHA-3, RIPEMD, …

Ejemplo

- **Queremos una función hash de longitud n que sea resistente a colisiones.**
- **Usando un ataque basado en la paradoja del cumpleaños el tiempo aproximado para encontrar una colisión es 2^(n/2).**
- **Para prevenir el ataque actualmente se requeriría que n > 256.**
- **Con un ordenador cuántico se encontraría una colisión en un tiempo aproximado de 2^(n/3).**
- **Para prevenir el ataque haría falta que n > 384**

Vectores cortos en retículos

- **Un retículo L es un subgrupo aditivo del espacio vectorial real n-dimensional generado por k (k menor o igual que n) vectores linealmente independientes: $b_1, b_2, \ldots, b_k$. (base de L)**
- **Por tanto L consiste en las combinaciones lineales enteras de una base.**
- **Hay varias bases de L, todas con el mismo número de elementos = la dimensión de L.**
- **El problema del vector mas corto en L trata de encontrar el vector no nulo del retículo con menor norma (euclídea).**

Vectores cortos en retículos 2

- **Relacionado con el problema de encontrar el vector de L mas cercano a un vector del espacio euclideo t.**
- **Otro problema importante es el de reducción de base del retículo.**
- **Hay diversos tipos de reducción: LLL, Korkine-Zolotaref.**

- **Para resolver eficientemente el problema del vector mas corto de un retículo se necesita, como parte del proceso, realizar de modo eficiente una reducción de base.**

Descodificación de códigos correctores

- **La teoría de códigos correctores de errores tiene como objetivo preservar la fiabilidad de la información que circula por un canal con ruido y que es susceptible de sufrir errores.**
- **Para construir un código lineal se considera un cuerpo finito F y el espacio vectorial F^n.**
- **Un (n,k)-código lineal $\mathcal{C}$ es un subespacio vectorial de dimensión k de F^n.**
- **El código $\mathcal{C}$ *se* puede construir a partir de una matriz generadora G (matriz k x n sobre F), sin mas que multiplicar k-tuplas de F por la matriz G.**

Descodificación de códigos correctores 2

- **En el espacio F^n se puede considerar la distancia de Hamming entre vectores = número de componentes distintas.**
- **El problema general de descodificación trata de recuperar la información enviada c a partir de la información recibida v (= c + e) sabiendo que se han producido t errores y estando t dentro de la capacidad correctora del código.**
- **Para ello se trata de buscar el vector dentro del código C que está mas cercano (respecto a la distancia de Hamming) al vector v de F^n recibido.**

Descodificación de códigos correctores 3

- **Se sabe que el problema general de descodificación en códigos lineales es NP-completo.**
- **Existen códigos con matrices generadoras que permiten una codificación eficiente, como los códigos binarios de Goppa.**
- **Se pueden construir usando un polinomio g con coeficientes en el cuerpo F = GF(2^m).**
- **Si g es irreducible de grado t el código de Goppa tiene distancia mínima 2t+1, por tanto puede corregir t errores.**
- **Se puede resolver el problema general de descodificación en tiempo polinomial en t.**

Resolución de ecuaciones cuadráticas

- **El problema MQ se puede plantear de la siguiente manera:**

  Dados un cuerpo finito F, dos enteros positivos n,m **y polinomios cuadráticos** p_1**, …,** p_m **en** n **variables** x_1**, …,** x_n **y coeficientes en F encontrar raíces comunes a todos ellos.**

- **Se sabe que si m y n son comparables este problema es NP completo.**

¿Por qué estos problemas?

- **En los problemas anteriores se basan las construcciones criptográficas (esquemas de cifrado y de firma digital) que parecen actualmente mas esperanzadoras en un eventual escenario post-cuántico.**
- **RSA y EC son actualmente los sistemas criptográficos mas utilizados.**
- **A lo largo de los años se han conseguido implementaciones muy eficientes y que parecen seguras.**

Computación cuántica

- **La introducción del concepto de ordenador cuántico a principios de los 80 introdujo un nuevo escenario y abrió la puerta a una amenaza latente.**
- **En 1994 Peter Shor presentó algoritmos cuánticos que permitirían la factorización de enteros y el cálculo de logaritmos discretos en tiempo polinomial.**
- **La existencia de un ordenador cuántico potente volveria el RSA y EC inutilizables.**
- **Hace falta buscar alternativas para la eventual aparición de ordenadores cuánticos potentes.**

Algoritmo de Grover

- **En 1996 Grover introdujo un algoritmo cuántico para buscar en una base no ordenada de tamaño N usando raiz cuadrada de N preguntas cuánticas.**
- **El algoritmo de Grover afectaría la seguridad de AES (Advanced Encryption Standard), pero no del modo dramático en que el RSA y EC se ven afectados por el algoritmo de Shor.**
- **Seria necesario aumentar el tamaño de claves del DES.**

Esquemas de firma basados en hash

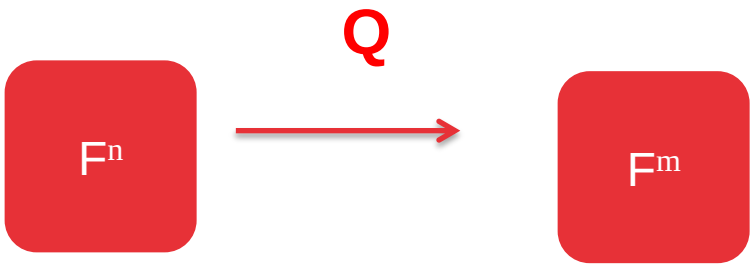
- **Merkle propuso el sistema de firma digital MSS a finales de los 70.**
- **Mientras RSA y ElGamal basan la seguridad en la dureza de un problema matemático, MSS lo hace en un hash.**
- **Genera muchos pares formados por la clave de firma y verificación de un esquema de firma de un solo uso (Lamport-Diffie) y usa un árbol hash binario para reducir la validez de 2^h claves de verificación de un solo uso a una clave pública MSS (h es la altura del árbol binario).**

hash2

- **Las claves de verificación se agrupan en parejas y las hojas del árbol son las funciones hash de ellas.**
- **La clave pública es la raiz del árbol.**
- **Inicialmente el esquema era menos eficiente que RSA o ElGamal.**
- **La versión mejorada XMSS es eficiente, flexible y de alta seguridad.**
- **Problema: Hay que mantener recuerdo de todas las claves de un solo uso utilizadas en el proceso.**
- **No existen equemas de cifrado de clave pública basados en funciones hash.**

Criptografía multivariable

- **Hay esquemas de clave pública que basan su seguridad en la dificultad de resolver un sistema no lineal de ecuaciones en varias variables.**
- **Sea F un cuerpo finito y** m, n **enteros.**



$$x = (x_1, \ldots, x_n) \longrightarrow (q_1(x), \ldots, q_m(x))$$

Es fácil obtener la preimagen de cualquier elemento por Q

Criptografía multivariable 2

- **La aplicación Q se enmascara usando dos aplicaciones afines lineales S: $F^n \longrightarrow F^n$ y T: $F^m \longrightarrow F^m$.**

La clave pública del esquema es la aplicación P = T.Q.S

(difícil de invertir)

La clave secreta está formada por Q,S,T

(permite invertir P)

Criptografía multivariable 3

- **Para firmar un documento** d **se usa una función hash**

- $H : \{0,1\}^* \longrightarrow F^m$

- **Se computa** $h = H(d)$
- **Para conseguir la firma de d se hace sucesivamente** $x = T^{-1}(h)$, $y = Q^{-1}(x)$ **(cualquier preimagen),** $z = S^{-1}(y)$.
- **Para verificar la autenticidad de la firma basta comprobar que**

$$P(z) = H(d)$$

Criptografía multivariable 4

- **Hay varios esquemas de firma multivariable prácticos: UOV, Rainbow, HFE^v (produce firmas muy cortas, Gui (prometedor), …**
- **Parecen una buena alternativa a los esquemas hash.**
- **Firmar y verificar es muy rápido.**
- **Su problema es el tamaño de la clave.**
- **Se pueden construir esquemas de cifrado cuando Q es inyectiva (m > n). Se cifra x como c = Q(x). Se descifra haciendo x = S^{-1}. Q^{-1}. T^{-1} (c).**
- **Esquema ABC, cifrados y descifrados rápidos. Claves largas. No se conoce a fondo la seguridad.**

Criptografía basada en códigos

- **Se consiguen esquemas de cifrado muy eficientes, pero la clave es muy larga.**
- **También hay esquemas de firma digital (precisan mas investigación)**
- **Primer esquema: McElice (1978). Se considera seguro, incluso frente a ataques cuánticos.**
- **Se considera un (n,k)-código lineal C con una buena matriz generatriz G y que permita corregir eficientemente un número de errores (sensiblemente) mayor que t.**
- **Se seleccionan aleatoriamente (con una distribución uniforme) matrices S,P con entradas en el cuerpo F.**

Criptografía basada en códigos 2

- **La matriz** k x k, **S, es inversible y la matriz** n x n, **P, es una matriz permutación y se usan para esconder la matriz G.**

- **La clave pública está formada por t y la matriz G' = SGP, que es una matriz generadora del código** *C'* **un código permutado de** *C*.
- **La clave secreta son las matrices S, G, P.**

- **Cifrado : Un mensaje m de F^k se codifica con G' y se le suma un vector aleatorio z de F^n de peso t**

$$c = mG' + z$$

Criptografía basada en códigos 3

- **Sea $x = c\,P^{-1}$.**
- **Como $c\,P^{-1} = mG'\,P^{-1} + z\,P^{-1} = mSG + zP^{-1}$, sabemos que x es la palabra obtenida de una palabra del código C (y = mSG) después de haberse producido un error ($z\,P^{-1}$) también de peso t.**
- **Por tanto se puede corregir la palabra x, obteniendo la palabra código y.**
- **Basta resolver el sistema lineal**

$$y = mSG$$

para recuperar el mensaje m.

Criptografía basada en códigos 4

- **El cifrado y descifrado del código McElice son eficientes, pero la clave es grande.**
- **Hay variantes que permiten implementaciones en hardware. Parecen una de las alternativas postcuánticas mas fiables.**
- **Existen también firmas en este esquema, pero son lentas y las claves son largas.**
- **Se considera que el problema de descodificar un código de Goppa usando una matriz generadora arbitraria es duro.**
- **La posibilidad de utilizar otros códigos requiere investigación adicional.**

Criptografía basada en retículos

- **Conceptualmente es similar a la basada en códigos. El conocimiento de una "buena base reducida" B permite computar el vector mas próximo a uno dado (lo que es intratable en general).**
- **Se considera un retículo L (contenido en R^n) en el que se conoce una base reducida B.**
- **Se toma B como clave privada y otra base cualquiera ("mala") como clave pública.**
- **Cifrado: Para cifrar un vector v de L se le suma un pequeño vector error e. Se tiene así $c = v + e$.**
- **Para descifrar hay que encontrar el vector v de L mas próximo a c.**

Criptografía basada en retículos 2

- **También se puede pensar en esquemas de firma digital.**
- **Se necesita una función hash**

$$h : \{0,1\}^* \longrightarrow F^m$$

- **La firma del mensaje d sería el vector s mas cercano a h(d).**
- **Para verificar la firma basta comprobar que s está en el retículo L y es cercano a h(d)**
- **¡No se producen esquemas seguros!**
- **Hay que hacer modificaciones que presentan implementaciones eficientes.**
- **¿Seguridad?**

Criptografía basada en retículos 3

- **Esquemas prometedores desde un punto de vista de investigación y desarrollo.**
- **Permiten reducción del caso peor al medio, lo que evita la necesidad de generar instancias duras**
- **Contraposición con el RSA.**
- **Hay propuestas eficientes y hay demostraciones de seguridad.**
- **Parecen objetivos contrapuestos** 

Se necesita mas investigación

Cifrado homomórfico

- **Su objetivo es permitir operar con datos cifrados, sin necesidad de descifrarlos previamente.**
- **Un cifrado homomórfico cumple respecto a dos leyes, $*$ en el conjunto de textos cifrados y . en el conjunto de mensajes, que $c_1 * c_2$ es el cifrado de $m_1.m_2$ cuando c_1 es el cifrado de m_1 y c_2 es el cifrado de m_2.**
- **El RSA es un cifrado homomórfico**
- **En la tesis de Gentry aparece la primera propuesta de cifrado homomórfico**

Idea de Gentry

- **Construir primero un esquema “algo homomórfico” (somewhat homomorphic encryption scheme) que permite realizar sumas y algunos productos.**

- **Usar un algoritmo Recrypt para construir un esquema totalmente homomórfico (fully homomorphic encryption scheme)**

- **Se conserva la seguridad en el proceso anterior**
- **Gentry utiliza criptografía basada en retículos.**

Gracias por su atención

GOBIERNO DE ESPAÑA
MINISTERIO DE ENERGÍA, TURISMO Y AGENDA DIGITAL

incibe_
INSTITUTO NACIONAL DE CIBERSEGURIDAD