POLITECNICO DI MILANO

Dipartimento di
Elettronica, Informazione
e Bioingegneria

Stato dell'arte della Crittografia basata su Codici

Gerardo Pelosi gerardo.pelosi@polimi.it

PQCifris, Roma, 9 Maggio 2019

Crittosistemi post-quantum

Problemi computazionali facenti uso di sistemi lineari

Sia H una matrice $n-k \times n$ a coefficienti in un campo finito $\mathbb{F}_q$

Dato un vettore $\mathbf{s} \in \mathbb{F}_q^{n-k}$, è possibile calcolare un vettore $\mathbf{e} \in \mathbb{F}_q^n$ tale che $\mathbf{He}^T = \mathbf{s}$?

- Scegliendo $n-k$ colonne di H è possibile individuare una sua sottomatrice $A_{n-k \times n-k}$
- con buona probabilità, la matrice A risulterà invertibile
- calcolare un vettore $e \in \mathbb{F}_q^n$ diventa immediato: $H = [A \mid B] \Rightarrow e = \left[A^{-1}s \mid 0_{k \times k}\right]$

Crittosistemi post-quantum

Problemi computazionali difficili con sistemi lineari

Sia H una matrice $n-k \times n$ a coefficienti in un campo finito $\mathbb{F}_q$

Dato un vettore s $\in \mathbb{F}_q^{n-k}$, è possibile **calcolare un vettore e** $\in \mathbb{F}_q^n$ corto in una metrica designata **tale che** $\mathbf{He}^T = \mathbf{s}$ **?**

- metrica $=$ distanza di Hamming $\Rightarrow$ crittosistemi basati su codici di correzione errori
- metrica $=$ distanza Euclidea $\Rightarrow$ crittosistemi basati su reticoli
- metrica $=$ distanza basata sul rango $\Rightarrow$ crittosistemi *rank-based*

La sicurezza e l'efficienza di ogni crittosistema derivano dalla metrica considerata!

Se si considerasse un sistema non-lineare (quadratico, cubico, etc...) $\Rightarrow$ crittosistemi basati su sistemi multivariati

Crittosistemi post-quantum

Vantaggi e svantaggi dei crittosistemi post-quantum

- progettati per rendere impraticabile il calcolo dei migliori algoritmi di crittonanalisi matematica, anche se accelerati tramite l'uso di un computer quantistico
- esibiscono un'efficienza di calcolo delle funzioni di cifratura e decifrazione uguale o superiore a quello dei crittosistemi correnti
- è possibile esibire efficaci ed efficienti contromisure ad attacchi conosciuti di crittoanalisi applicata
- la dimensione delle chiavi crittografiche è superiore a quelle dei crittosistemi correnti a pari garanzie di sicurezza!

Crittosistemi post-quantum basati su codici

Codici lineari di correzione errori

Con un codice di correzione $\mathcal{C}$ applicabile a parole x composte da k simboli consecutivi ognuno in $\mathbb{F}_q$, $x \in \mathbb{F}_q^k$, per ottenere una parola di codice $y \in \mathbb{F}_q^n$, $n > k$, con distanza minima d, in grado di correggere $t \leq \lfloor \frac{d-1}{2} \rfloor$ simboli errati è necessario disporre di

- matrice generatrice $G \in \mathbb{F}_q^{k \times n}$ tale che $\mathcal{C} = \{y = xG \mid x \in \mathbb{F}_q^k\}$
- *t-bounded decoder*: $\forall y \in \mathcal{C}, e \in \mathbb{F}_q^n, wt(e) \leq t \Rightarrow \Phi_{\mathcal{C}}(y + e) = x$

- matrice di parità $H \in \mathbb{F}_q^{(n-k) \times n}$ tale che $\mathcal{C} = \{y \in \mathbb{F}_q^n \mid Hy^{\mathrm{T}} = 0\}$
- *t-bounded syndrome decoder*: $\forall y \in \mathcal{C}, e \in \mathbb{F}_q^n, wt(e) \leq t, s = H(y + e)^{\mathrm{T}} \Rightarrow \Psi_{\mathcal{C}}(s) = e$

Crittosistemi post-quantum basati su codici

Crittosistema di McEliece

Procedura di generazione delle chiavi

- Sia $G' \in \mathbb{F}_q^{k \times n}$ la matrice generatrice di un codice di correzione per al più t errori
- Selezionare casualmente una matrice di permutazione $P \in \mathbb{F}_q^{n \times n}$ e una matrice invertibile $S \in \mathbb{F}_q^{k \times k}$
- chiave pubblica: $\mathtt{k}_{\mathtt{pub}} = (G, t)$, con $G = SG'P$; chiave privata: $\mathtt{k}_{\mathtt{priv}} = (S^{-1}, G', P^{-1})$

Procedura di cifratura

- Dato un *plaintext* $u \in \mathbb{F}_q^k$, si sceglie casualmente in modo uniforme $e \in \mathbb{F}_q^n$ con $wt(e) \leq t$ e usando la chiave pubblica si calcola il *ciphertext* $c = uG + e$

Procedura di decifrazione

- Dato un *ciphertext* $c \in \mathbb{F}_q^n$, usando la chiave privata, si applica il decodificatore a cP^{-1} ottenendo $\widetilde{u} = \Phi_{\mathcal{C}}(cP^{-1}) = \Phi_{\mathcal{C}}(uSG' + eP^{-1}) = uS$ e si calcola $u = \widetilde{u}S^{-1}$

Crittosistemi post-quantum basati su codici

Crittosistema di Niederreiter

Procedura di generazione delle chiavi

- Sia $H' \in \mathbb{F}_q^{(n-k)\times n}$ la matrice di parità di un codice di correzione per al più t errori
- Selezionare casualmente una matrice di Permutazione $P \in \mathbb{F}_q^{n\times n}$ e una matrice invertibile $S \in \mathbb{F}_q^{(n-k)\times(n-k)}$
- chiave pubblica: $\mathtt{k}_{\mathtt{pub}} = (H, t)$, con $H = SH'P$; chiave privata: $\mathtt{k}_{\mathtt{priv}} = (S^{-1}, H', P^{-1})$

Procedura di cifratura

- Dato un *plaintext* $u \in \mathbb{F}_q^n$ con $wt(u) \leq t$, usando la chiave pubblica, si calcola il *ciphertext* come la sindrome del codice $c = Hu^{\mathrm{T}}$

Procedura di decifrazione

- Dato un *ciphertext* $c \in \mathbb{F}_q^n$, usando la chiave privata, si decodifica $S^{-1}c$ ottenendo $\widetilde{u} = \Psi_{\mathcal{C}}(S^{-1}c) = \Psi_{\mathcal{C}}(H'P^{-1}u^{\mathrm{T}}) = \Psi_{\mathcal{C}}(H'(uP)^{\mathrm{T}}) = uP$, e si calcola $u = \widetilde{u}P^{-1}$

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli crittosistemi di McEliece/Niederreiter

Problema di decodifica a minima distanza di codici lineari (P-DMD)

- Dato un codice lineare casuale $\mathcal{C} \subseteq \mathbb{F}_q^n$ con potere correttivo t e una parola di codice $y \in \mathcal{C}$, trovare un'altra parola di codice $y' \in \mathcal{C}$ tale che $wt(y - y') \leq t$, oppure dichiarare la non esistenza di una tale y'

Problema di decodifica della sindrome di codici lineari (P-DS)

- Dato un codice lineare casuale $\mathcal{C} \subseteq \mathbb{F}_q^n$ con potere correttivo t, una matrice di parità $H \in \mathbb{F}_q^{(n-k)\times n}$ e una sindrome $s \in \mathbb{F}_q^{n-k}$, trovare una parola di codice $e \in \mathcal{C}$ con $wt(e) \leq t$ tale che $He^{\mathrm{T}} = s$, oppure dichiarare la non esistenza di una tale e

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter

- I due problemi P-DMD e P-SD sono classificabili computazionalmente in NP e sono equivalenti
- La versione decisionale del P-DMD è stata provata essere NP-*complete* per codici binari da Berlekamp, McEliece e van Tilborg nel 1978, mostrando una riduzione al problema del *Three-Dimensional Matching*.
  Per codici in $\mathbb{F}_q$ la stessa prova è stata formalizzata da Barg nel 1994.

- Risolvere i problemi di ricerca rende risolvibili anche i problemi di decisione $\Rightarrow$
  I problemi di ricerca sono difficili quanto o più dei corrispondenti problemi di decisione
- Per problemi decisionali NP-*complete*, i corrispondenti problemi di ricerca sono anche'essi in NP-*complete*. I due problemi di ricerca P-DMD e P-SD sono classificabili certamente come NP-*hard* e in particolare come NP-*complete*

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter

- Fattorizzazione di interi e estrazione di log. discreti sono stati provati essere risolvibili con un algoritmo per macchina quantistica in tempo polinomiale che termina non correttamente un numero limitato di volte (classe di complessità BQP)
- È congetturato che BQP $\cap$ NP-*complete* $= \emptyset$
  - se per violare un cifrario devo risolvere un problema in NP-*complete*, allora una macchina quantistica non è in grado di volare il cifrario in tempo polinomiale
- Se ne deduce che le versioni decisionali di P-DMD e P-CS non sono risolvibili da macchina quantistica, e dunque anche P-DMD e P-CS non lo sono!

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter

McEliece propose di utilizzare codici di Goppa binari ($n = 1024$, $k = 524$, $t = 101$) fondando la sicurezza dello schema su

- difficoltà di decodificare un codice casuale
- difficoltà di derivare un algoritmo di decodifica a partire dalla matrice generatrice pubblica del codice di Goppa selezionato

Sebbene la seconda assunzione non sia provata essere NP-*hard*, il sistema originale ha resistito a ben oltre 40 anni di crittanalisi algebrica

- solo per codici di Goppa con $k/n \to 1$ è nota la costruzione di un attacco strutturale

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter

Il crittosistema di McElice originale non è usato in pratica

- a causa delle dimensioni di chiave molto grandi!
- relativamente bassa efficienza delle implementazioni dati i parametri a disposizione!

Security level (bit)	(n, k)	t	chiave pub. (kB)	chiave pub. RSA (kB)
128	(2960, 2288)	56	1502	3
256	(6624, 5129)	115	7489	15

Soluzione: cambiare codice!

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter – attacchi passivi

Attacchi strutturali l'attaccante prova a distinguere/ricostruire la struttura del codice segreto a partire dalle informazioni pubbliche del crittosistema per invalidare l'assunzione che esse equivalgano a esibire un codice casuale

Message Recovery Attack dato un *ciphertext* e le informazioni pubbliche, l'attaccante tenta di calcolare il *plaintext* corrispondente

Key Recovery Attack date le informazioni publiche del crittosistema e una chiave pubblica, l'attaccante mira a derivare la chiave privata corrispondente

Crittosistemi post-quantum basati su codici

Sicurezza computazionale degli schemi di McEliece/Niederreiter – attacchi passivi

- Codici insicuri ad **attacchi strutturali**
  - GRS (Niederreiter 1986 - broken in 1992)
  - Reed-Muller (b. Borodin 2013)
  - Convolutional codes (b. Landais 2013), Algebraic geometry (b. Pellikaan 2014)
  - q-ary Goppa (b. Faugère 2015, Otmani 2015),
  - Polar codes (b. Bardet 2016), QD-Srivastava codes (signature b. Faugère 2016)
- **Message recovery attack**
  - algoritmi di enumerazione o ricerca esaustiva (*bruteforcing*)
  - algoritmi *Information Set Decoding* (ISD)
- **Key recover attack**
  - gli algoritmi ISD sono applicabili se il codice segreto sottostante ammette parole con peso molto basso

Crittosistemi post-quantum basati su codici

Decodificare un codice casuale: *bruteforcing*

Consideriamo per comodità il P-SD con un codice binario $\mathcal{C}(n, k, t)$

Data la matrice di parità $H \in \mathbb{F}_2^{(n-k)\times n}$, $s \in \mathbb{F}_2^{n-k}$, trovare e con $wt(e) = t$ tale che $He^{\mathrm{T}} = s$

$$H = \left[\right]_{(n-k)\times n} \qquad s = \left[\right]_{(n-k)\times 1}$$

$$e = \left[\right]_{1\times n}$$

Trovare t colonne di H che sommate tra loro uguagliano s ha complessità computazionale $O\left(\binom{n}{t}\right)$ operazioni tra colonne della matrice H

Crittosistemi post-quantum basati su codici

Decodificare un codice casuale: *bruteforcing*

Data la matrice di parità $H \in \mathbb{F}_2^{(n-k)\times n}$, pensata come divisa a metà $H = [H_1 H_2]$ e la sindrome $s \in \mathbb{F}_2^{n-k}$, trovare $e = [e' e'']$ con $wt(e') = t/2$, $wt(e'') = t/2$, tale che $He^{\mathrm{T}} = s$

$$H = \left[\begin{array}{c|c} & \\ H_1 & H_2 \\ & \end{array}\right]_{(n-k)\times n} \qquad s = \left[\begin{array}{c} \\ \\ \end{array}\right]_{(n-k)\times 1}$$

$$e = \left[\begin{array}{c|c} e' & e'' \end{array}\right]_{1\times n}$$

Trovare $t/2$ colonne di H_1 che sommate tra loro uguagliano s sommata a $t/2$ colonne da H_2 ha complessità $O\left(2\binom{n/2}{t/2}\right)$ con probabilità di successo $\frac{\binom{n/2}{t/2}^2}{\binom{n}{t}}$

costo = num. iter. $\times$ costo 1 iter. $= 2\frac{\binom{n}{t}}{\binom{n/2}{t/2}} \approx (8\pi t)^{\frac{1}{4}}\sqrt{\binom{n}{t}}$ op. tra colonne di H

Crittosistemi post-quantum basati su codici

Decodificare un codice casuale: Information Set Decoding (Prange 1962)

Data la matrice di parità $H \in \mathbb{F}_2^{(n-k)\times n}$, si selezionano casualmente $n-k$ colonne permutandole (P) per portarle a sinistra della matrice e si applica la riduzione di Gauss-Jordan ottenendo $H' = UHP = [I_{n-k} V_{k\times k}]$ e $s' = Us$

$$UHP = \left[\begin{array}{c|c} I_{n-k} & \end{array}\right]_{(n-k)\times n} \qquad Us = \left[\begin{array}{c} \\ \end{array}\right]_{(n-k)\times 1}$$

$$eP = \left[\begin{array}{c|c} e' & e'' \end{array}\right]_{1\times n}$$

Obiettivo: ottenere $wt(s') = t$ **!** che è equivalente a $wt(e'') = 0 \Rightarrow eP = [Us\, 0_{k\times k}]$

Probabilità di successo $\frac{\binom{n-k}{t}}{\binom{n}{t}} \approx (\frac{n-k}{n})^t$

costo = num. iter. $\times$ costo 1 iter. $= (\frac{n}{n-k})^t (n-k)n = 2^{ct}(n-k)n \quad$ con $c = -\log_2(1-\frac{k}{n})$

Crittosistemi post-quantum basati su codici

Decodificare un codice casuale: Information Set Decoding – miglioramenti

- Lee Brickell 1988
- Leon 1988
- Stern 1989 & Dumer 1991
- May, Meurer, Thomae 2011;
- Becker, Joux, May, Meurer, 2012;
- May, Ozerov, 2015.

- Molti dei miglioramenti alla complessità dell'ISD riguardano solo codici binari
- complessità temporale asintotica per un codice (n, k, t) è: $2^{ct(1+o(1))}$ dove $c = -\log_2(1 - \frac{k}{n})$, indipendentemente dalla variante algoritmica dell'ISD

Per risolvere il P-SD di un crittosistema con livello di sicurezza 2^{128}

- da Prange 1962 a May Ozerov 2015, l'ISD è migliorato di un fattore $\approx 2^{30}$
- da Stern 1989 a May Ozerov 2015, l'ISD è migliorato di un fattore $\approx 2^4$

Crittosistemi post-quantum basati su codici

Sicurezza dei crittosistemi di McEliece/Niederreiter – attacchi attivi
Chosen Ciphertext Attack

- L'attaccante sceglie due messaggi u_0, u_1, lo sfidante crea una coppia di chiavi pubblica/privata e cifra uno dei due scegliendo a caso e sottopone il *ciphertext* all'attaccante
- L'attaccante interrogando un oracolo di decifrazione per tutti i *ciphertext* di sua scelta (prima e dopo aver interagito con lo sfidante) tenta di indovinare il testo in chiaro scelto dallo sfidante con una probabilità migliore del 50%
  - l'attaccante non può interrogare l'oracolo con la sfida ricevuta
  - l'attaccante se usa chiave pubblica per cifrare i due messaggi, non ottiene informazioni confrontando il valore ricevuto dallo sfidante

Nota:

- L'attaccante può chiedere la decifrazione di qualunque stringa di lunghezza adeguata
- L'oracolo deve rifiutare sistematicamente la decifrazione di un cifrato non ammissibile

Crittosistemi post-quantum basati su codici

Attacchi attivi e contromisure

Alcuni crittosistemi non sono sempre in grado di decifrare correttamente un cifrato ammissibile! (DFR $\neq 0$).
Un attaccante CCA può:

ability-1 chiedere la decifrazione di cifrati non ammissibili

ability-2 interrogare l'oracolo con cifrati ammissibili fino ad osservare un fallimento della decifrazione (a.k.a. *reaction attacks*)

Dunque, in un crittosistema resistente ad attacchi CCA è necessario che l'attaccante non distingua tra una decifratura a buon fine e una situazione di errore dovuta a una delle cause precedenti (accidentali e/o indotte):

counter-1 al termine del processo di decifrazione si controlli se quanto calcolato è valido (es. il mittente accoda al messaggio una informazione di conferma – come un hash dello stesso – il ricevente controlla l'uguaglianza delle hash).
Se non lo è, si risponde al mittente con una stringa casuale estratta a partire da un valore segreto del ricevitore.

- partendo da cifrari con DFR $= 0$ sicuri contro attacchi passivi (a.k.a. OW-CPA), ci sono costruzioni standard per ottenere schemi resistenti a CCA

counter-2 avere una DFR trascurabile nel livello di sicurezza del crittosistema

Crittosistemi post-quantum basati su codici

KEM CCA ammessi al Round 2 dell'iniziativa di standardizzazione U.S. NIST
Security level: sforzo computazionale di rompere AES-128 con quantum computer

schema	codice	t	QC	DFR$=0$	Public key size
Classic McEliece	Goppa(3488, 2720)	64	$\times$	✓	261 kB
NTS-KEM	Goppa(4096, 3328)	64	$\times$	✓	319 kB
HQC*	BCH tensor ripetiz.	60	✓	✓	3125 B
BIKE2	MDPC(23558, 11779)	134	✓	$\times$	1472 B
BIKE2-CPA	MDPC(20326, 10163)	134	✓	$\times$	1270 B
LEDAcrypt	LDPC(104294, 52147)	136	✓	$\times$	6518 B
LEDAcrypt-CPA	LDPC(29878, 14939)	136	✓	$\times$	1867 B

Conclusioni

Vantaggi

- Difficoltà computazionale del problema di decodifica di un codice casuale ammette una riduzione a un problema NP-complete
- Algoritmi efficienti: funzione di cifratura molto veloce

Limiti

- Dimensione delle chiavi pubbliche (non necessariamente un problema)
- Relativamente poche famiglie di codici adatti

Conclusioni

Problema aperto - schema di firma con chiavi ridotte

CFS (Courtois, Finiasz & Sendrier, 2001): stesse garanzie McEliece/Niederreiter
Codice di Goppa $n = 2^m$, con t errori; pub.: $H \in \mathbb{F}_2^{mt \times n}$; priv.: (S, H', P), $H = SH'P$

Firma

- Firma su $\mathtt{msg}$ composta come $e \in \mathbb{F}_2^n$ di peso minimo w tale che $He^{\mathrm{T}} = \mathtt{hash}(\mathtt{msg}) \in \mathbb{F}_2^{mt}$, avendo a disposizione un *syndrome decoder* H'
  - nota w può risultare minore o maggiore di t; $w \leq t$ con prob. $\frac{1}{t!}$

Verifica

- Dati $\mathtt{msg}$, e, controllo che $He^{\mathrm{T}} \stackrel{?}{=} \mathtt{hash}(\mathtt{msg})$

- il miglior attacco ha un costo $= O(n^{t(0.5+o(1))}) \Rightarrow$ esponenziale in mt
- costo della firma = $O(t!t^2(\log n)^3) \Rightarrow$ esponenziale in t
- dimensioni della chiave pubblica $= mtn \Rightarrow$ esponenziale in m
- per sicurezza equiv. 2^{80} pre-quantum: $n = 2^{16}$, $m = 16$, $t = 9$, $\mathtt{size}(H) = 1$ Mb

Grazie per l’attenzione!