Esquemas de identificação baseados no problema da decodificação de síndromes

André Jucovsky Bianchi ajb@ime.usp.br

Departamento de Ciência da Computação Instituto de Matemática e Estatística Universidade de São Paulo

5 de outubro de 2011

Objetivo do seminário

Apresentar esquemas de identificação baseados no problema da decodificação de síndromes:

- Identificação baseada na decodificação de síndromes em códigos binários [Stern, 1994].
- ► Um melhoramento do esquema anterior utilizando códigos q-ários [Cayrel et al., 2011].

Estrutura da apresentação

Provas interativas com conhecimento zero

Identificação baseada no problema da decodificação de síndromes

Identificação baseada no problema da decodificação de síndromes q-árias

Estrutura da apresentação

Provas interativas com conhecimento zero

Identificação baseada no problema da decodificação de síndromes

Identificação baseada no problema da decodificação de síndromes q-árias

Motivação

Não se conhece algoritmo probabilístico de tempo polinomial para as seguintes linguagens:

- ▶ QR = $\{(x, n) \mid \exists y \text{ t.q. } x = y^2 \pmod{n}\}.$
- ▶ QNR = $\{(x, n) \mid \nexists \ y \ \text{t.q.} \ x = y^2 \ (\text{mod } n)\}.$

Onde:

 \triangleright $x, n \in \mathbb{N}$, com 0 < x < n e mdc(x, n) = 1.

Motivação

Não se conhece algoritmo probabilístico de tempo polinomial para as seguintes linguagens:

- ▶ QR = $\{(x, n) \mid \exists y \text{ t.q. } x = y^2 \pmod{n}\}.$
- ▶ QNR = $\{(x, n) \mid \nexists \ y \ \text{t.q.} \ x = y^2 \ (\text{mod } n)\}.$

Onde:

 \triangleright $x, n \in \mathbb{N}$, com 0 < x < n e mdc(x, n) = 1.

Observação: O problema QR é *NP*—completo (e portanto QNR é *coNP*—completo).

Sistemas de prova interativos Intuição

Intuitivamente, um procedimento de prova de teoremas eficiente deve possuir as seguintes características:

- 1. Deve ser possível "provar" um teorema verdadeiro.
- 2. Deve ser impossível "provar" um teorema falso.
- 3. A comunicação e verificação da "prova" devem ser eficientes.

O verificador pode ser convencido de forma errônea da veracidade de uma proposição apenas com uma probabilidade bem baixa.

Formalização

Um sistema de prova interativo para um conjunto S é um jogo de dois participantes: um verificador que executa um algoritmo probabilístico de tempo polinomial V e um provador que executa um algoritmo P, satisfazendo duas condições:

- 1. Completeness: Para todo $x \in S$, $P\{V \text{ aceita a prova de } P\} = 1$.
- 2. Soundness: Para todo $x \notin S$ e toda estratégia \tilde{P} , $\mathbf{P}\{V \text{ aceita a prova de } \tilde{P}\} \leq \frac{1}{2}$.

Considerações

IP é a classe das linguagens que possuem um sistema de prova interativo:

- Existem outras definições equivalentes com valores diferentes para as probabilidades.
- ▶ NP \subset IP.
- ► IP = PSPACE.

Sistemas de prova interativos $QNR \in IP$

Um sistema de prova interativo para QNR:

- 1. Dado x, \bigvee escolhe $i \xleftarrow{\$} \{0,1\}$ e $z \in \mathbb{Z}_n^*$
 - Se i = 0, V computa $w = z^2 \pmod{n}$.
 - ▶ Se i = 1, \bigvee computa $w = xz^2 \pmod{n}$.
 - \triangleright V envia w para P e pergunta o valor de i.
- 2. Se $x \in QNR$, w é resíduo quadrático se e só se i = 0. P pode então calcular j e enviar o resultado para V.
- 3. V aceita se e só se i = j.

Provas com conhecimento zero

A transcrição de V de um sistema de prova interativo é o conjunto das seguintes informações:

- A(s) entrada(s) do sistema.
- ▶ As mensagens trocadas por P e V.
- Os números aleatórios escolhidos por V.

Dada uma entrada x para um sistema de provas, definimos:

- T(x): o conjunto de todas as possíveis transcrições verdadeiras.
- $ightharpoonup \mathcal{F}(x)$: o conjunto de todas as possíveis transcrições forjadas.
- ▶ $\mathbf{P}_{\mathcal{T}}$ { T}: probabilidade de que T ocorra como transcrição de uma prova interativa real conduzida por P e V com entrada x.
- ▶ P_F{ T}: probabilidade de que T ocorra como transcrição de uma prova interativa forjada criada por um forjador F com entrada x.



Provas com conhecimento zero perfeito

Dizemos que um sistema de prova interativo para uma linguagem L é de conhecimento zero perfeito se existe um forjador F tal que para qualquer entrada x valem:

- 1. $\mathcal{T}(x) = \mathcal{F}(x)$.
- 2. $\mathbf{P}_{\mathcal{T}}\{T\} = \mathbf{P}_{\mathcal{F}}\{T\}$, para todo $T \in \mathcal{T}(x)$.

Provas com conhecimento zero computacional

Dizemos que um sistema de prova interativo para uma linguagem L é de conhecimento zero computacional se existe um forjador F tal que para qualquer entrada x valem:

- 1. $\mathcal{T}(x) = \mathcal{F}(x)$.
- 2. $\mathbf{P}_{\mathcal{T}}\{T\}$ e $\mathbf{P}_{\mathcal{F}}\{T\}$ são computacionalmente indistinguíveis, para todo $T \in \mathcal{T}(x)$.

Autenticação com conhecimento zero: o esquema de identificação Fiat-Shamir

Preâmbulo:

- 1. Uma autoridade publica n = pq, com $p \in q$ primos, mantendo $p \in q$ secretos.
- 2. P escolhe $s \stackrel{\$}{\leftarrow} \{1, \dots, n-1\}$ e computa sua chave pública $v = s^2 \pmod{n}$.

Protocolo de identificação:

- 1. P escolhe $r \stackrel{\$}{\leftarrow} \mathbb{Z}_{n-1}^*$, computa $x = r^2 \pmod{n}$ e envia para V.
- 2. **V** escolhe $e \stackrel{\$}{\leftarrow} \{0,1\}$ e envia para P.
- 3. P computa $y = rs^e \pmod{n}$ e envia para V.
- 4. Se y = 0 ou $y^2 \neq xv^e \pmod{n}$, V rejeita. Caso contrário, V aceita.

Autenticação com conhecimento zero: o esquema de identificação Fiat-Shamir

Claramente, vale que:

▶ $P{V \text{ aceita a prova de } P} = 1.$

Como forjar o esquema:

- $ightharpoonup ilde{P}$ poderia enviar r^2 ou r^2v^{-1} no passo 1.
- ▶ $\mathbf{P}\{V \text{ aceita a prova de } \tilde{P}\} \leq \frac{1}{2}$.

Além disso, o esquema é de conhecimento zero.

Estrutura da apresentação

Provas interativas com conhecimento zero

Identificação baseada no problema da decodificação de síndromes

Identificação baseada no problema da decodificação de síndromes q-árias

Um (n, k, t)-código é um subespaço k-dimensional de um espaço vetorial n-dimensional sobre um corpo finito \mathbb{F}_q , capaz de corrigir um número máximo de t erros.

O peso de Hamming de um vetor x é o número de coordenadas não nulas de x, e é representado por wt(x).

Uma Matriz Geradora G de um código linear C é uma matriz cujas linhas formam uma base de C:

$$\mathcal{C} = \{xG : x \in \mathbb{F}_q^{n-k}\}.$$

Uma *Matriz de Paridade H* de um código linear \mathcal{C} é uma matriz de dimensão (k,n) cujas linhas formam uma base para o espaço ortogonal complementar do subespaço \mathcal{C} :

$$\mathcal{C} = \{ x \in \mathbb{F}_q^n : Hx^T = 0 \}.$$

Mais definições: o problema SD

O Problema da Decodificação de Síndromes Binárias (SD).

Entrada: $H \xleftarrow{\$}$ Binárias(k, n), $y \xleftarrow{\$} \mathbb{F}_2^k$ e um inteiro $\omega > 0$. Saída: Uma palavra $s \in \mathbb{F}_2^n$ tal que $wt(s) \leq \omega$ e $Hs^T = y^T$.

Mais definições: o problema SD

O Problema da Decodificação de Síndromes Binárias (SD).

Entrada: $H \xleftarrow{\$}$ Binárias(k, n), $y \xleftarrow{\$} \mathbb{F}_2^k$ e um inteiro $\omega > 0$. Saída: Uma palavra $s \in \mathbb{F}_2^n$ tal que $wt(s) \leq \omega$ e $Hs^T = y^T$.

Observação: O problema SD é NP-difícil.

Mais definições: o problema G-SD

O Problema Geral da Decodificação Binária (G-SD).

Entrada: $G \stackrel{\$}{\leftarrow} \text{Binárias}(n-k,n)$, $y \stackrel{\$}{\leftarrow} \mathbb{F}_2^n$ e um inteiro $\omega > 0$. Saída: Palavras $x \in \mathbb{F}_2^{n-k}$, $e \in \mathbb{F}_2^n$ tais que $wt(e) \leq \omega$ e xG + e = y.

Mais definições: o problema G-SD

O Problema Geral da Decodificação Binária (G-SD).

Entrada: $G \stackrel{\$}{\leftarrow} \text{Binárias}(n-k,n)$, $y \stackrel{\$}{\leftarrow} \mathbb{F}_2^n$ e um inteiro $\omega > 0$. Saída: Palavras $x \in \mathbb{F}_2^{n-k}$, $e \in \mathbb{F}_2^n$ tais que $wt(e) \leq \omega$ e xG + e = y.

Observação: O problema G-SD é NP-difícil.

Identificação através da decodificação de síndrome [Stern, 1994] - Parâmetros do sistema

Parâmetros públicos:

- ▶ $H \in \mathbb{F}_2^{k \times n}$, matriz de paridade.
- $\omega \in \mathbb{N}$, o peso de Hamming da chave privada.
- ▶ Uma função de hash h.

Chave privada:

▶ $s \in \mathbb{F}_2^n$, com peso ω .

Identificação pública:

$$i^T = Hs^T \in \mathbb{F}_2^n$$
.

Identificação através da decodificação de síndrome [Stern, 1994] - Protocolo de identificação

O protocolo consiste de r rodadas iguais:

- 1. P escolhe $u \stackrel{\$}{\leftarrow} \mathbb{F}_2^n$ e uma permutação $\sigma \stackrel{\$}{\leftarrow} S_n$, e envia três comprometimentos para V:
 - $ightharpoonup c_1 \leftarrow h(\sigma, Hu^T).$
 - $ightharpoonup c_2 \leftarrow h(u\sigma).$
 - ▶ $c_3 \leftarrow h((u \oplus s)\sigma)$.
- 2. V envia $b \stackrel{\$}{\leftarrow} \{0,1,2\}$ para P.
- 3. P revela para V alguns valores dependendo de b:
 - ▶ Se b = 0, P revela $u \in \sigma$, e V verifica $c_1 \in c_2$.
 - ▶ Se b = 1, P revela $u \oplus s$ e σ , e V verifica c_1 e c_3 .
 - ▶ Se b = 2, P revela $u\sigma$ e $s\sigma$, e V verifica c_2 , c_3 , e o fato de que $s\sigma$ possui peso ω .

Identificação através da decodificação de síndrome [Stern, 1994] - Falsificando a identificação

Estratégias para falsificar a identificação:

- ▶ Utilizar $\tilde{s} \neq s$ com $wt(\tilde{s}) = wt(s) = \omega$.
- ▶ Utilizar $\tilde{s} \neq s$ com $wt(\tilde{s}) \neq wt(s)$ tal que $H\tilde{s}^T = i^T$.

Em ambos os casos, a probabilidade de sucesso é $\left(\frac{2}{3}\right)^r$.

[Stern, 1994] - Segurança do esquema (1/3): prova interativa

A segurança do esquema é baseada nos seguintes fatos:

- ▶ Determinar se um código possui uma palavra s de peso ω , cuja imagem é uma palavra i de k bits é NP-completo.
- ▶ Inverter $s \mapsto Hs^T$ é um problema NP-difícil.
- ▶ $P{V \text{ aceita a prova de } P} = 1$, se P é honesto.
- ▶ $\mathbf{P}\{V \text{ aceita a prova de } \tilde{P}\} \leq \left(\frac{2}{3}\right)^r$ para todo \tilde{P} desonesto.

E nas seguintes suposições:

- Problemas NP-difíceis são intratáveis.
- Funções de hash criptográficas seguras existem.

[Stern, 1994] - Segurança do esquema (2/3): conhecimento zero

As únicas informações reveladas por P são:

- ightharpoonup b = 0: uma palavra aleatória u e uma permutação aleatória σ .
- ▶ b = 1: uma palavra aleatória $u \oplus s$ e uma permutação aleatória σ .
- ▶ b=2: uma palavra aleatória $u\sigma$ e uma palavra aleatória $s\sigma$ de peso ω .

O esquema não revela nenhuma informação sobre s (*Zero-Knowledge proof-system*).

[Stern, 1994] - Segurança do esquema (3/3): valores dos parâmetros

Para um nível de segurança igual a 2⁶¹, o autor sugere utilizar:

- ightharpoonup n=2k.
- n = 512, k = 256.
- $\sim \omega \approx 56.$

O valor de ω deve ser escolhido com cuidado:

- Chaves secretas com peso baixo demais podem ser encontradas com probabilidade alta.
- ➤ O limite de Gilbert-Varshamov é um limite teórico para o peso mínimo d de um (n, k)—código aleatório:

 - $H_2(x) = -x \log_2(x) (1-x) \log_2(1-x).$
- ightharpoonup O autor sugere utilizar ω um pouco abaixo do valor d.

Identificação através da decodificação de síndrome [Stern, 1994] - Desempenho do esquema

Considerações sobre o desempenho do esquema:

- Não ocupa muita memória.
- ▶ A complexidade de comunicação é comparável ao esquema Fiat-Shamir.
- ▶ A segurança pode ser aumentada escolhendo n, k e ω maiores.
- ► A dificuldade de computar Hu^T pode ser reduzida.
- O nível de segurança depende de r.
- O esquema não é baseado em identidade.

[Stern, 1994] - Variante 1: computacionalmente mais leve

Introduzimos um novo parâmetro q e as seguintes alterações no protocolo:

- ▶ No passo 1, $c_1 \leftarrow h(\sigma)$.
- ▶ Após o passo 1, V envia q índices de $\{1, ..., k\}$.
- ▶ P envia os bits q_1, \ldots, q_b correspondentes aos índices de Hu^T .
- O resto do protocolo é similar.
- Aumentos na probabilidade de falsificar a prova podem ser compensados aumentando o número de rodadas.
- ▶ n = 512, k = 256, $\omega = 56$, q = 64 e um round a mais para um nível de segurança parecido.

[Stern, 1994] - Variante 2: número menor de rodadas

Substituímos a chave secreta s por um código simplex gerado por s_1, \ldots, s_m . A chave pública é Hs_1^T, \ldots, Hs_m^T .

- Cada rodada possui 5 passagens de mensagem.
- ▶ m = 7, n = 576 e k = 288 para um nível de segurança parecido.
- ▶ $\mathbf{P}\{V \text{ aceita a prova de } \tilde{P}\} = \frac{1}{2}$.

[Stern, 1994] - Variante 3: baseada em identidade

Nesta variante, a chave pública pode ser derivada da identidade do usuário.

- É usado um conjunto de t códigos simplex.
- ▶ n = 864, k = 432, $\omega = 95$ e t = 56.

[Stern, 1994] - Um esquema análogo baseado no problema (modular) da mochila

Esta variante do esquema utiliza a matriz H sobre um corpo finito pequeno. A restrição do peso é substituída pela restrição de que a solução s de $Hs^T = i^T$ deve ser constituída apenas de zeros e uns.

- ▶ O tamanho do corpo é *m*.
- n = 128, k = 64, e m = 5.
- n = 196, k = 128, e m = 3.
- n = 192, k = 96, e m = 6.
- n = 384, k = 256, e m = 3.

Estrutura da apresentação

Provas interativas com conhecimento zero

Identificação baseada no problema da decodificação de síndromes

Identificação baseada no problema da decodificação de síndromes q-árias

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Motivação

Duas grandes desvantagens do esquema anterior em relação ao esquema Fiat-Shamir:

- 1. A probabilidade de sucesso de um provador \tilde{P} é alta, e portanto demanda um maior número de rodadas para atingir um nível de segurança razoável.
- 2. Os dados compartilhados por todos os usuários são muito grandes (66 Kbits contra 1024 bits).

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Ataques a esquemas de identificação baseados no problema SD

Existem duas famílias principais de algoritmos que decodificam um código linear binário aleatório:

- ▶ Information Set Decoding (ISD).
- Generalized Birthday Algorithm (GBA).

Os parâmetros apresentados são baseados na família ISD:

► WF_{ISD} é a complexidade do algoritmo ISD.

[Cayrel et al., 2011] - Mais definições: o problema qSD

O Problema da Decodificação de Síndromes q-árias (qSD).

Entrada: $H \xleftarrow{\$} q$ -árias(k, n), $y \xleftarrow{\$} \mathbb{F}_q^r$ e um inteiro $\omega > 0$. Saída: Uma palavra $s \in \mathbb{F}_q^n$ tal que $wt(s) \leq \omega$ e $Hs^T = y^T$.

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Mais definições: o problema qSD

O Problema da Decodificação de Síndromes q-árias (qSD).

Entrada: $H \xleftarrow{\$} q$ -árias(k, n), $y \xleftarrow{\$} \mathbb{F}_q^r$ e um inteiro $\omega > 0$. Saída: Uma palavra $s \in \mathbb{F}_q^n$ tal que $wt(s) \leq \omega$ e $Hs^T = y^T$.

Observação: O problema qSD é NP-difícil.

O limite Gilbert-Varshamov q-ário.

Seja $H_q(x)$ a função de entropia q-ária:

$$H_q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x).$$

Suponha $0 \le \xi \le (q-1)/q$. Então existe uma sequência infinita de (n,k,d)-códigos lineares q-ários com $d/n=\xi$ e taxa R=k/n satisfazendo a desigualdade:

$$R \geq 1 - H_q(\xi), \ \forall n.$$

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Mais definições:

Sejam $\sigma \in S_n$ e $\gamma \in (\mathbb{F}_q^*)^n$. Definimos a seguinte transformação:

$$\pi_{\gamma,\sigma}: \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^n$$

$$v \longmapsto (\gamma_{\sigma(1)}v_{\sigma(1)}, \dots, \gamma_{\sigma(n)}v_{\sigma(n)})$$

Note que:

- $\qquad \qquad \pi_{\gamma,\sigma}(\alpha \mathbf{v}) = \alpha \pi_{\gamma,\sigma}(\mathbf{v}), \ \forall \alpha \in \mathbb{F}_q, \ \forall \mathbf{v} \in \mathbb{F}_q^n.$
- $wt(\pi_{\gamma,\sigma}(v)) = wt(v), \forall v \in \mathbb{F}_q^n$.

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Geração de chaves

Algoritmo de geração de chaves:

Escolha aleatoriamente n, $r \in \omega$, tais que $WF_{ISD}(n, r, \omega, q) \ge 2^{\kappa}$.

- $H \xleftarrow{\$} \mathbb{F}_q^{k \times n}.$
- $s \stackrel{\$}{\leftarrow} \mathbb{F}_q^n$, tal que $wt(s) = \omega$.
- ▶ $y \leftarrow Hs^T$.

Saída: $(s, (y, H, \omega))$.

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Protocolo de identificação

O protocolo consiste de r rodadas iguais:

- 1. P escolhe $u \stackrel{\$}{\leftarrow} \mathbb{F}_2^n$, uma permutação $\sigma \stackrel{\$}{\leftarrow} S_n$ e $\gamma \stackrel{\$}{\leftarrow} (\mathbb{F}_q^*)^n$, e envia dois comprometimentos para V:
 - $ightharpoonup c_1 \leftarrow h(\sigma, \gamma, Hu^T).$
 - $c_2 \leftarrow h(\pi_{\gamma,\sigma}(u),\pi_{\gamma,\sigma}(s)).$
- 2. V envia $\alpha \stackrel{\$}{\leftarrow} \{0,1\}$ para P.
- 3. *P* envia $\beta \leftarrow \pi_{\gamma,\sigma}(u + \alpha s)$ para V.
- 4. V envia $b \stackrel{\$}{\leftarrow} \{0,1\}$ para P.
- 5. P revela para V alguns valores dependendo de b:
 - ▶ Se b = 0, P revela γ e σ , e V verifica c_1 .
 - Se b=1, P revela $\pi_{\gamma,\sigma}(s)$ e V verifica c_2 e o fato de que $\pi_{\gamma,\sigma}(s)$ possui peso ω .

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Propriedades do sistema

- Completeness: P{ V aceita a prova de P} = 1, se P é honesto.
- ► Soundness: $\mathbf{P}\{V \text{ aceita a prova de } \tilde{P}\} \leq \left(\frac{q}{2(q-1)}\right)^r$ para todo \tilde{P} desonesto.
- Zero-Knowledge proof-system.

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Segurança do esquema

A segurança do esquema se baseia em três propriedades dos códigos q-ários lineares:

- Códigos aleatórios satisfazem o limite inferior de Gilbert-Varshamov.
- 2. Para *n* grande, quase todos os códigos lineares estão acima do limite.
- 3. Resolver o problema da decodificação de síndrome q-ária para códigos aleatórios é NP-difícil.

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Comparação entre esquemas SD e qSD

Para um nível de segurança de 2^{87} e probabilidade de enganar igual a 2^{-16} , temos:

	SD	G-SD	Stern5	qSD
Rodadas	28	28	16	16
Tamanho da matriz	122500	122500	122500	32768
Identificação	350	700	2450	512
Chave secreta	700	1050	4900	1024
Comunicação	42019	35486	62272	31888
Computação de <i>P</i>	2 ^{22,7}	$2^{22,7}$	$2^{21,92}$	$2^{16}m + 2^{16}a$
Domínio	\mathbb{F}_2	\mathbb{F}_2	\mathbb{F}_2	F ₂₅₆

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Parâmetros

Para um nível de segurança de 2¹²⁸, os parâmetros indicados são:

$$q = 256$$
, $n = 208$, $k = 104$, $wt(s) = 78$.

	qSD	
Rodadas	16	
Tamanho da matriz	86528	
Identificação	832	
Chave secreta	1664	
Comunicação	47248	
Computação de <i>P</i>	$2^{17,4}m + 2^{17}a$	
Domínio	\mathbb{F}_{256}	

Identificação através da decodificação de síndrome q—ária [Cayrel et al., 2011] - Reduzindo o tamanho da chave pública

Utilizando matrizes duplamente circulantes, é possível diminuir o tamanho da chave consideravalmente.

Referências bibliográficas



Cayrel, P.-L., Véron, P., and El Yousfi Alaoui, S. M. (2011). A zero-knowledge identification scheme based on the q-ary syndrome decoding problem.

In Proceedings of the 17th international conference on Selected areas in cryptography, SAC'10, pages 171–186, Berlin, Heidelberg. Springer-Verlag.



Stern, J. (1994).

A new identification scheme based on syndrome decoding. In *Proceedings of the 13th annual international cryptology conference on Advances in cryptology*, pages 13–21.

Obrigado pela atenção.

Dados de contato:

- ► Meu email: ajb@ime.usp.br
- ► Esta apresentação: http://www.ime.usp.br/~ajb/