ACH2076 – Segurança da Informação

Aula 03: Criptoanálise

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2023

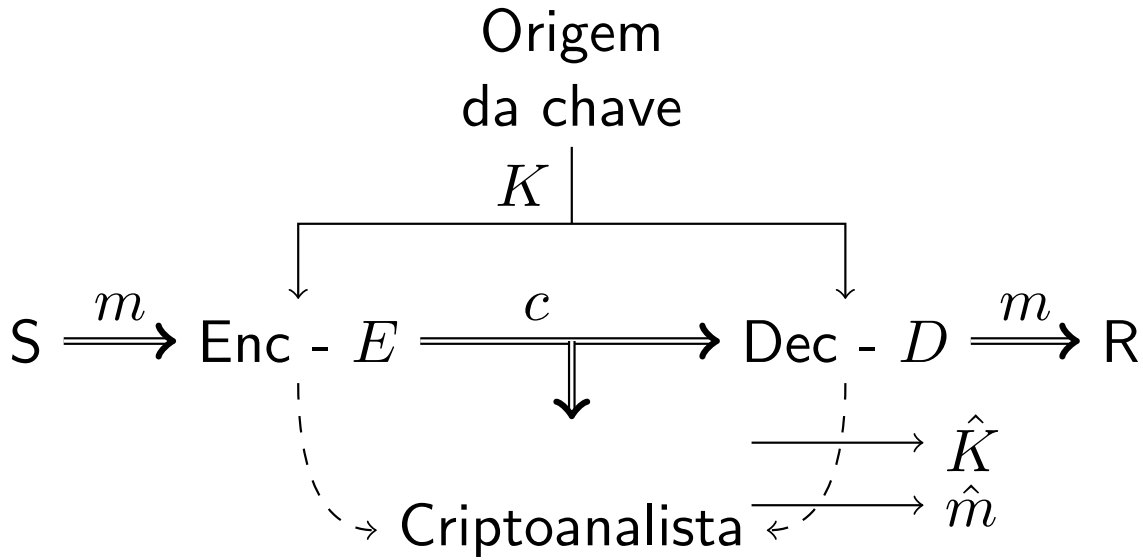
Livros Textos

1. William Stallings. Criptografia e Segurança de redes: princípios e práticas, 4ª edição. **Capítulo 2**
2. Marcio Moretto Ribeiro. Segurança da Informação (apostila).

- Força Bruta
  - Máxima Verossimilhança
- Criptoanálise
  - Texto cifrado
  - Pares texto cifrado / texto aberto
  - Pares texto cifrado / texto aberto arbitrários

Criptoanálise

Estudo de métodos e princípios de transformar mensagens ininteligíveis de volta a mensagens inteligíveis, sem conhecimento da chave.



Criptoanálise vs Força Bruta

- Criptoanálise
  - Contam com a natureza do algoritmo
  - Características gerais do texto claro
  - Pares de (texto claro, texto cifrado)

- Força Bruta
  - Conhece algoritmo de decriptografia
  - Experimenta cada chave possível
  - Testa se é uma tradução inteligível
  - Na média, metade de todas chaves precisam ser experimentadas

Criptoanálise

- Incondicionalmente seguro
  - one-time pad (chave de uso único)
  - Texto com n caracteres
  - Vigènere e chave com n caracteres aleatórios

- Computacionalmente Seguro
  - Custo para quebrar cifra superior ao valor da informação codificada
  - Tempo exigido para quebrar a cifra superior ao tempo de vida útil da informação

Força Bruta

Tamanho da chave	Quantidade de Chaves	Tempo Necessário (1 decrip/μs)	Tempo Necessário (10^6 decrip/μs)
2 caracteres (Cifra de César Afim)	312	156μs	$< 1\mu$s
26 caracteres (monoalfabética)	26!=4×10^{26}	$6,4 \times 10^{12}$ anos	$6,4 \times 10^{6}$ anos
32 bits	2^{32}=$4,3 \times 10^{9}$	$35,8$ minutos	$2,15$ms
56 bits (DES)	2^{56}=$7,2 \times 10^{16}$	1142 anos	$10,01$ horas
128 bits (AES)	2^{128}=$3,4 \times 10^{38}$	$5,4 \times 10^{24}$ anos	$5,4 \times 10^{18}$ anos
168 bits (3DES)	2^{168}=$3,7 \times 10^{50}$	$5,9 \times 10^{36}$ anos	$5,9 \times 10^{30}$ anos

Força Bruta

- Teste de texto inteligível
  - Texto em língua conhecida: português, inglês, etc.

- Exemplo:
  - "Drosb mykmr, Wkckiycrs Wkxklo, cksn: Go my-ybnsxkdon bokvvi govv kxn dro zvkiobc gybuon bokvvi govv dyqodrob. Lydr dokwc gobo cy xobfyec sx dro psxkv cod, led Tkzkx rkn tecd k vsddvo lsd wybo."

Força Bruta

- Cesar Afim (1,8): Vjgkt eqcej, Ocucaqujk Ocpcdg, uckf: Yg eq-qtfkpcvgf tgcnna ygnn cpf vjg rncagtu yqtmgf tgcnna ygnn vqigvjgt. Dqvj vgcou ygtg uq pgtxqwu kp vjg hkpcn ugv, dwv Lcrcp jcf lwuv c nkvvng dkv oqtg.
- Cesar Afim (1,9): Uifjs dpbdi, Nbtbzptij Nbobcf, tbje: Xf dp-psejobufe sfbmmz xfmm boe uif qmbzfst xpslfe sfbmmz xfmm uphfuifs. Cpui ufbnt xfsf tp ofswpvt jo uif gjobm tfu, cvu Kbqbo ibe kvtu b mjuumf cju npsf.
- Cesar Afim (1,10): Their coach, Masayoshi Manabe, said: We co-ordinated really well and the players worked really well together. Both teams were so nervous in the final set, but Japan had just a little bit more.

Força Bruta

- Como testar automaticamente?

- Exemplo: Análise de Frequência
  - Língua apenas com 3 símbolos: `a`, `b` e `c`
  - Considere textos com 100 caracteres
  - Distribuições de probabilidades (monogramas, digramas, trigramas)

$$\Pr(N_a = \alpha, N_b = \beta, N_c = \theta)$$
$$\Pr(N_{aa} = \alpha, N_{ab} = \beta, N_{ac} = \theta, N_{ba} = \gamma, \ldots)$$
$$\Pr(N_{aaa} = \alpha, N_{aab} = \beta, N_{aac} = \theta, N_{aba} = \gamma, \ldots)$$

Força Bruta

- Como obter distribuições de probabilidades?
- Simplificação: assume independência

$$\begin{aligned}
&\Pr(N_a = \alpha, N_b = \beta, N_c = \theta) = \\
&\quad = \Pr(N_a = \alpha | N_b = \beta, N_c = \theta) \times \Pr(N_b = \beta | N_c = \theta) \times \Pr(N_c = \theta) \\
&\quad \approxeq \Pr(N_a = \alpha) \times \Pr(N_b = \beta) \times \Pr(N_c = \theta)
\end{aligned}$$

Força Bruta

Distribuição Binomial (Bernoulli)

- Distribuição de probabilidades discreta
- Número de sucessos em uma sequência de N tentativas
- p é chance de obter sucesso e mantém-se constante

$$\Pr(N_a = \alpha) = \binom{N}{\alpha} (p_a)^{\alpha} (1 - p_a)^{N-\alpha}$$

$$\binom{N}{\alpha} = \frac{N!}{(N-\alpha)!\alpha!}$$

Força Bruta

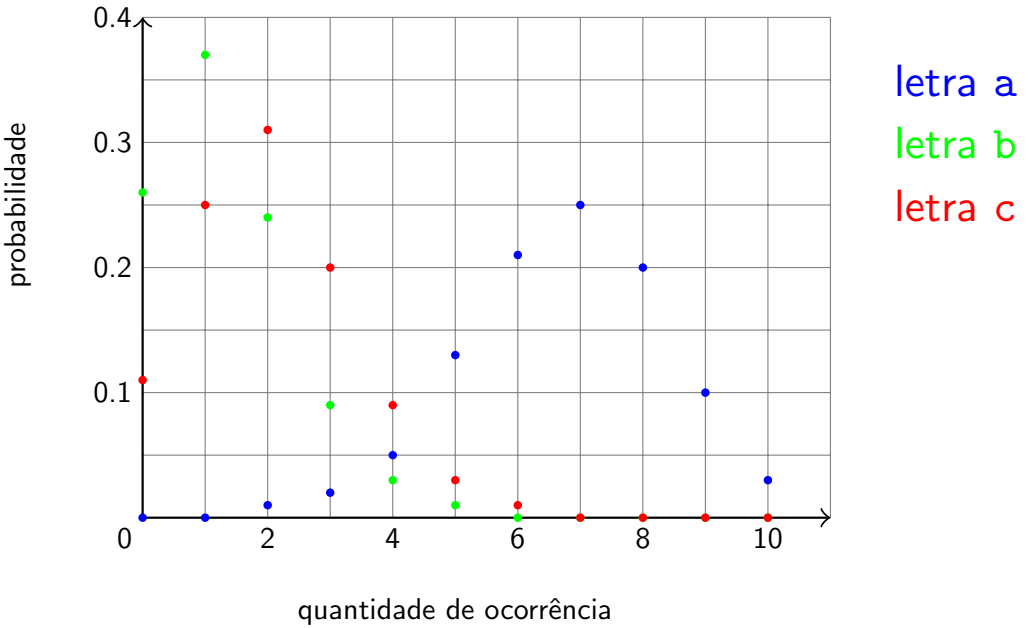
Máxima Verossimilhança (exemplo com monogramas)

- Dado um texto cifrado c, testa-se todas chaves K
- Para cada chave, conte as quantidades $\alpha_a^K, \alpha_b^K, \alpha_c^K, \ldots, \alpha_z^K$ para cada letra no texto "aberto" m obtido ao aplicar a chave específica K no texto cifrado c.
- Maximize

$$\Pr\left(\text{texto em inglês} \mid N_a = \alpha_a^K, N_b = \alpha_b^K, N_c = \alpha_c^K, \ldots, N_z = \alpha_z^K\right) \propto$$
$$\propto \Pr\left(N_a = \alpha_a^K, N_b = \alpha_b^K, N_c = \alpha_c^K, \ldots, N_z = \alpha_z^K \,\middle|\, \text{texto em inglês}\right)$$

Força Bruta - Exemplo

- Considere um alfabeto com apenas 3 letras: a, b e c; e a distribuição abaixo para textos com 10 caracteres
- Utilizando verossimilhança, encontre as chaves nos casos a seguir:
  - $c =$ "bbabcbbabb"; cifra monoalfabética
  - $c =$ "acaaaaaacc"; cifra de Hill (chave 2×2 e não contém zeros)
  - $c =$ "bbbccbbbba"; cifra Afim



Criptoanálise vs Força Bruta

- Criptoanálise
  - Contam com a natureza do algoritmo
  - Características gerais do texto claro
  - Pares de (texto claro, texto cifrado)

- Força Bruta
  - Conhece algoritmo de decriptografia
  - Experimenta cada chave possível
  - Testa se é uma tradução inteligível
  - Na média, metade de todas chaves precisam ser experimentadas

Criptoanálise: força do adversário

- Apenas Texto Cifrado: o adversário conhece um texto cifrado c
- Texto claro conhecido: o adversário conhece um texto claro m e o texto cifrado correspondente c
- Texto claro escolhido: o adversário pode escolher um texto claro m e obter o texto cifrado correspondente c
- Texto cifrado escolhido: o adversário pode escolher um texto cifrado c e obter o texto claro correspondente m

Criptoanálise: Cifra de Hill e Afim

- Texto claro/cifrado escolhido:
  - Qual texto escolher?
  - Qual tamanho mínimo?
- Texto claro conhecido: esperar a condição acima acontecer.
- Apenas texto cifrado: dá pra fazer melhor que força bruta?

Criptoanálise: Cifra de Vigènere

- Texto claro/cifrado escolhido:
  - Qual texto escolher?
  - Qual tamanho mínimo?
- Texto claro conhecido: esperar a condição acima acontecer.
- Apenas texto cifrado: dá pra fazer melhor que força bruta?

Criptoanálise: Cifra Monoalfabética

- Texto claro/cifrado escolhido:
  - Qual texto escolher?
  - Qual tamanho mínimo?
- Texto claro conhecido: esperar a condição acima acontecer.
- Apenas texto cifrado: dá pra fazer melhor que força bruta?

Criptoanálise: Cifra Monoalfabética

Apenas Texto Cifrado: análise de frequência

- Solução Força Bruta
  - Considero um critério para indicar o quanto um texto é "inglês"
  - Testo todas as chaves possíveis (26!)
  - Verifico a chave que produziu maior aderência ao "inglês"
  - Problema: custoso computacionalmente

Criptoanálise: Cifra Monoalfabética

Apenas Texto Cifrado: análise de frequência

- Solução Ingênua
  - Solução Ingênua
  - Considera ordem da frequência de ocorrência de letras E, T, A, O, I, N, S, H, R, D, L, C, U, M, W, F, G, Y, P, B, V, K, J, X, Q e Z
  - Realizo a contagem de ocorrência de letras no texto cifrado
  - Atribuo conforme a ordem acima
  - Problema: qualidade varia bastante com o tamanho do texto

Criptoanálise: Cifra Monoalfabética

Apenas Texto Cifrado: análise de frequência

- Solução "Greedy" (gulosa)
  - Em cada iteração preencho uma letra ou um conjunto de letras na chave-solução
  - Considera um teste de aderência para cada letra
  - Teste de aderência considera solução parcial
  - Teste de aderência considera digrama, trigrama, etc.

Criptoanálise: Cifra Monoalfabética

Solução "Greedy" (gulosa)

- Solução Parcial Δ: chave para encriptar

$$\begin{array}{cccccccccccccccccccccccc} a & b & c & d & e & f & g & h & i & j & k & l & m & n & o & p & q & r & s & t & u & v & w & x & y & z \\ . & . & . & . & . & . & . & . & . & . & X & . & . & . & . & K & . & . & . & L & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & T & . & . & . & . & F & . & . & . & G & . & . & . & . & . & . \end{array}$$

- Notação
  - conjunto $\mathcal{P}$: conjunto de letras no texto aberto $\{a, b, \ldots, y, z\}$
  - conjunto $\mathcal{C}$: conjunto de letras no texto cifrado $\{A, B, \ldots, Y, Z\}$
  - conjunto $\mathcal{R} \subset \mathcal{P}$: se $\theta \in \mathcal{R}$, então letra θ não foi atribuída a uma letra do texto cifrado
  - conjunto $\mathcal{T} \subset \mathcal{C}$: se $\Theta \in \mathcal{T}$, então letra Θ ainda não foi atribuída a uma letra do texto aberto
- Solução Geral: repete enquanto $|\mathcal{T}| > 0$
  - Escolha uma permutação Φ com N_C letras em $\mathcal{T}$
  - Para cada permutação ϕ de N_C letras em $\mathcal{R}$ avalie sua aderência com a permutação Φ
  - Escolha a permutação ϕ^* com melhor aderência

Criptoanálise: Teste de aderência

Considere as permutações ϕ e Φ, dada uma medida $\mathcal{X}$ do texto cifrado e solução parcial Δ, deseja-se calcular a probabilidade:

$$Pr(\phi = \Phi | \mathcal{X}, \Delta) = Pr(\phi_1 = \Phi_1, \phi_2 = \Phi_2, \ldots, \phi_k = \Phi_k | \mathcal{X}, \Delta)$$

Posso medir no texto cifrado:

- $\mathcal{X}_{\Phi_i}$ para $1 \leq i \leq N_C$: quantas vezes aparece a letra Φ_i
- $\mathcal{X}_{\Phi_i\Phi_j}$ para $1 \leq i, j \leq N_C$: quantas vezes aparece o digrama $\Phi_i\Phi_j$
- $\mathcal{X}_{\Phi_i\Delta_j}$ para $1 \leq i \leq N_C$ e $1 \leq j \leq |\Delta|$: digramas $\Phi_i\Delta_j$ com elementos em Φ e Δ
- $\mathcal{X}_{\Delta_j\Phi_i}$ para $1 \leq i \leq N_C$ e $1 \leq j \leq |\Delta|$: digramas $\Delta_j\Phi_i$ com elementos em Φ e Δ
- $\mathcal{X}_{\Phi_i\Phi_j\Phi_k}$, $\mathcal{X}_{\Phi_i\Phi_j\Phi_k\Phi_l}$, $\mathcal{X}_{\Phi_i\Phi_j\Phi_k\Phi_l\Phi_m}$, etc.

Criptoanálise: Probabilidades

Como calcular (estimar) $Pr(\phi = \Phi | \mathcal{X}, \Delta)$?
Regra de Bayes:

$$Pr(\phi = \Phi | \mathcal{X}, \Delta) \propto Pr(\mathcal{X} | \phi = \Phi, \Delta) Pr(\phi = \Phi | \Delta)$$

Aproximação (exemplo apenas com digramas):

$$\begin{aligned} Pr(\mathcal{X} | \phi = \Phi, \Delta) &= Pr(\mathcal{X}_{\Phi_i}, \mathcal{X}_{\Phi_i \Phi_j}, \mathcal{X}_{\Phi_i \Delta_j}, \ldots | \phi_1 = \Phi_1, \ldots, \phi_k = \Phi_k, \Delta) \\ &\approx \prod_{\Phi_i} Pr(\mathcal{X}_{\Phi_i} | \phi_i = \Phi_i) \prod_{\Phi_i, \Phi_j} Pr(\mathcal{X}_{\Phi_i \Phi_j} | \phi_i = \Phi_i, \phi_j = \Phi_j) \\ &\quad \prod_{\Phi_i, \Delta_j} Pr(\mathcal{X}_{\Phi_i \Delta_j} | \phi_i = \Phi_i) \prod_{\Phi_i, \Delta_j} Pr(\mathcal{X}_{\Delta_j \Phi_i} | \phi_i = \Phi_i) \end{aligned}$$

Finalmente, as distribuições $Pr(\mathcal{X}_{\Phi_i} | \phi_i = \Phi_i)$ e
$Pr(\mathcal{X}_{\Phi_i \Phi_j} | \phi_i = \Phi_i, \phi_j = \Phi_j)$ são aproximadas por distribuições binomiais.

Criptoanálise: Exemplo

Considere um alfabeto com apenas 5 letras ($\mathcal{P} = \{a, b, c, d, e\}$ e $\mathcal{C} = \{A, B, C, D, E\}$) e o texto cifrado:

`ACDEADECADCEDACDEDACDEDCBABCDBBCDEADBECBACDEBAACDE`
`DEABBBADEADEDABDDAAABDEBDEABDEACDEACDEADACBDEBCADE`

Ocorrem as seguintes medidas para cada letra: $\mathcal{X}_A = 23$, $\mathcal{X}_B = 16$, $\mathcal{X}_C = 15$, $\mathcal{X}_D = 27$ e $\mathcal{X}_E = 19$

Ocorrem as seguintes medidas para cada digrama: $\mathcal{X}_{AA} = 3$, $\mathcal{X}_{AB} = 5$, $\mathcal{X}_{AC} = 8$, $\mathcal{X}_{AD} = 7$, $\mathcal{X}_{AE} = 0$, $\mathcal{X}_{BA} = 3$, $\mathcal{X}_{BB} = 3$, etc.

Solução Inicial Δ vazia $\begin{matrix} a & b & c & d & e \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{matrix}$, $\mathcal{R} = \mathcal{P}$ e $\mathcal{T} = \mathcal{C}$

Criptoanálise: Exemplo

- Escolha uma permutação Φ com 2 letras em $\mathcal{T}$: $\Phi = AD$
- Permutações em $\mathcal{R}$: $ab, ac, ad, ae, ba, bc, bd, be, ca, cb, \ldots$
- Para cada permutação ϕ calcule sua probabilidade (exemplo $\phi = ab$)

$$Pr(\mathcal{X}_A = 23|a = A)Pr(\mathcal{X}_D = 27|b = D)Pr(\mathcal{X}_{AA} = 3|a = A)$$
$$Pr(\mathcal{X}_{AD} = 7|a = A, b = D)Pr(\mathcal{X}_{DA} = 5|a = A, b = D)Pr(\mathcal{X}_{DD} = 1|b = D)$$

- Escolha a permutação ϕ^* com melhor aderência (suponha $\phi^* = ce$)

Criptoanálise: Exemplo

Passo 2 ($N_P = 2$)

- $\Delta = \begin{matrix} a & b & c & d & e \\ . & . & . & . & . \end{matrix} \begin{matrix} \\ A \\ \\ \\ D \end{matrix}$, $\mathcal{R} = \{a, b, d\}$ e $\mathcal{T} = \{B, C, E\}$
- Escolha uma permutação Φ com 2 letras em $\mathcal{T}$: $\phi = BC$
- Permutações em $\mathcal{R}$: ab, ad, ba, bd, da, db
- Para cada permutação ϕ calculo a probabilidade (exemplo $\phi = ab$)

$$Pr(\mathcal{X}_B = 16|a = B)Pr(\mathcal{X}_C = 15|b = C)Pr(\mathcal{X}_{BB} = 3|a = B)$$
$$Pr(\mathcal{X}_{BC} = 3|a = B, b = C)Pr(\mathcal{X}_{CB} = 2|a = B, b = C)Pr(\mathcal{X}_{CC} = 0|b = C)$$
$$Pr(\mathcal{X}_{BA} = 4|c = A, a = B)Pr(\mathcal{X}_{AB} = 4|c = A, a = B)$$
$$Pr(\mathcal{X}_{BD} = 5|e = D, a = B)Pr(\mathcal{X}_{DB} = 2|e = D, a = B)$$
$$Pr(\mathcal{X}_{CA} = 2|c = A, b = C)Pr(\mathcal{X}_{AC} = 8|c = A, b = C)$$
$$Pr(\mathcal{X}_{CD} = 9|e = D, b = C)Pr(\mathcal{X}_{DC} = 2|e = D, b = C)$$

- Escolha a permutação ϕ^* com melhor aderência (suponha $\phi^* = ad$)

Criptoanálise: Melhorias

Problema: Se N_C é muito grande existem muitas permutações
Solução: Considerar letras em $\mathcal{R}$ que atendem algum requisito
Exemplo: $Pr(\mathcal{X}_A|a = A) > 0.1 \max_{\theta \in \mathcal{R}} Pr(\mathcal{X}_A|\theta = A)$
Problema: Quais letras em Δ preencher primeiro?
Solução: alfabética, aleatória, maior frequência, análise otimizada