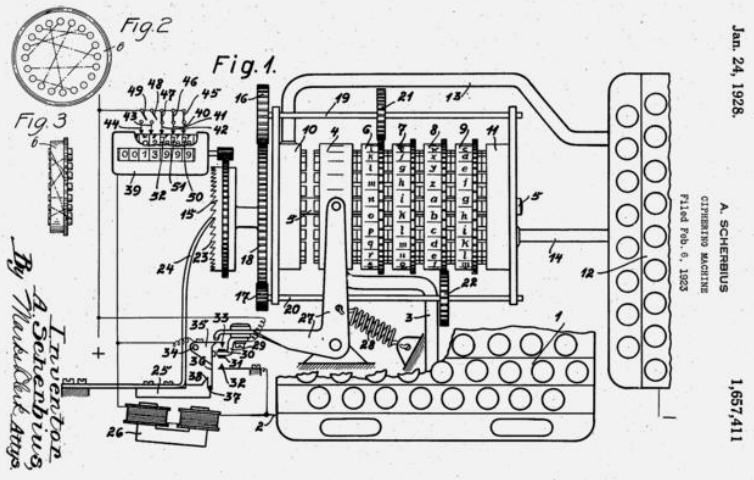
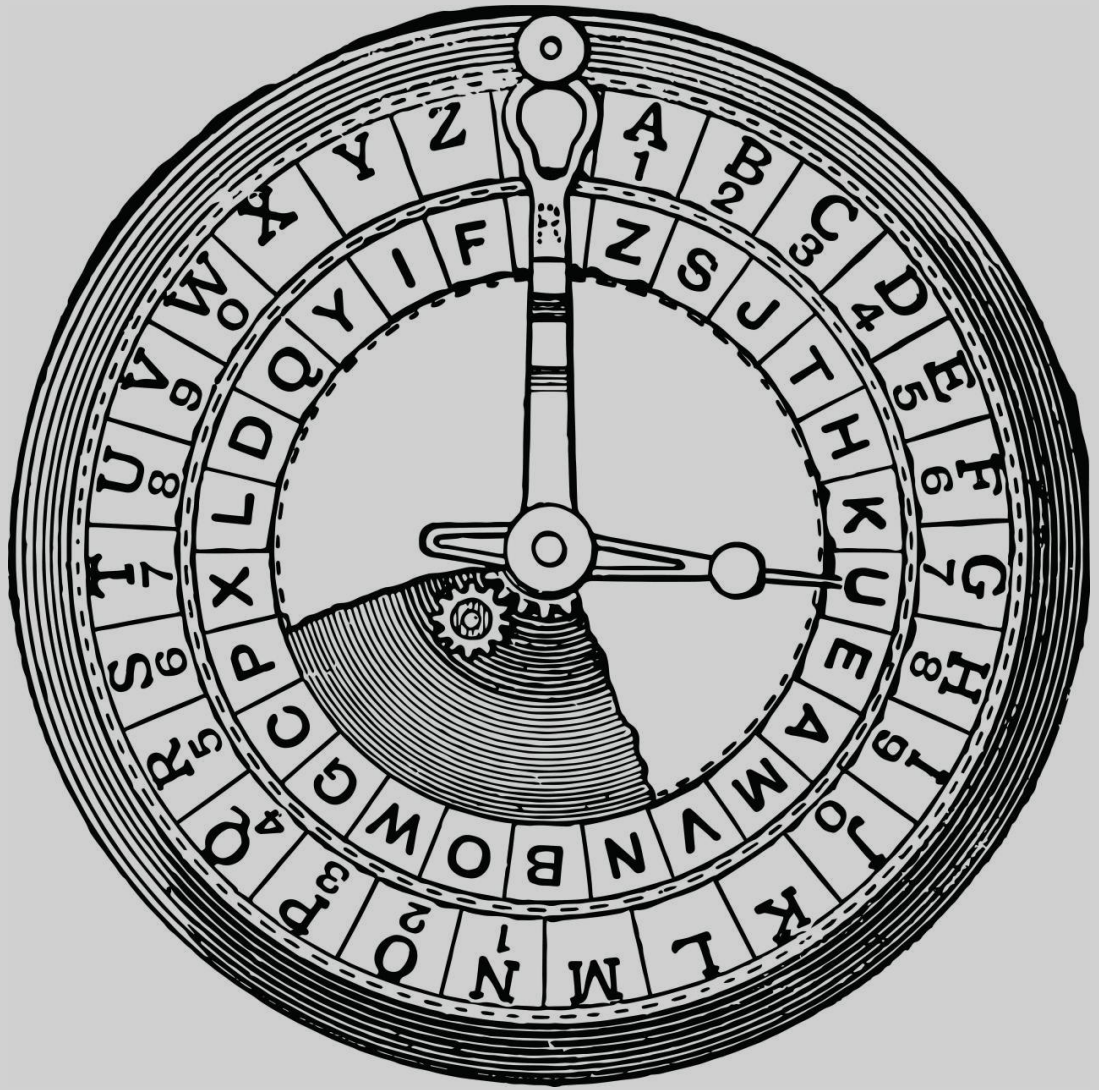
Criptografia Perfeitamente Secreta

Fundamentos da Criptografia

Nicollas Mocelin Sdroievski



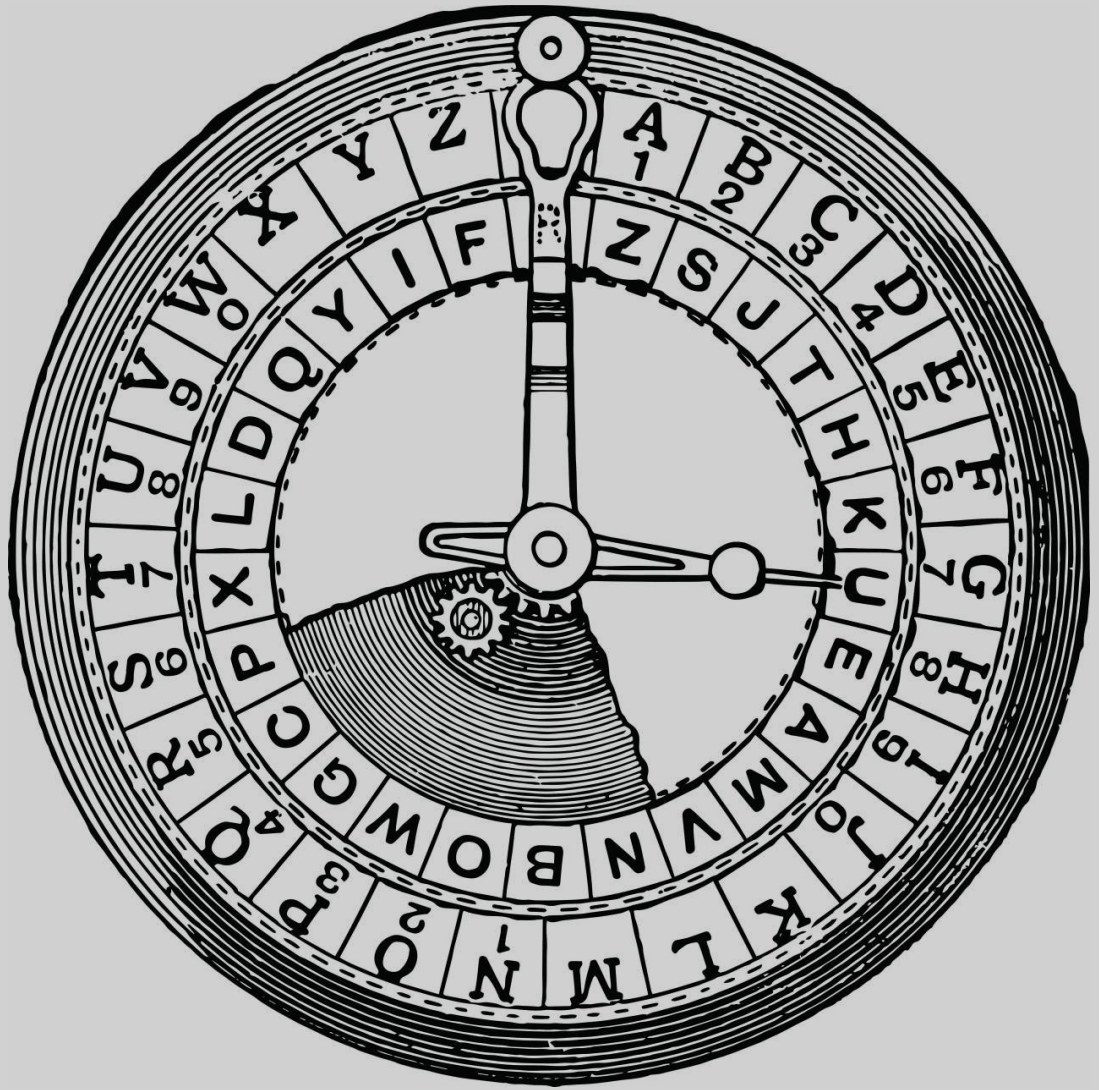
Na última aula



Cifras clássicas e seus ataques

- César
  - Força bruta
  - Análise de frequência
- Monoalfabética
  - Não mais vulnerável à força bruta
  - Análise de frequência
- Vigenère
  - Ataques apenas para tamanhos de chave “pequenos”

Um pouco de Probabilidade



Definições básicas

- Temos um universo contável Ω de elementos
- Associamos a cada elemento $x \in \Omega$ uma probabilidade $p(x)$, respeitando às seguintes regras
  - $p(x) \in [0,1]$ para todo $x \in \Omega$
  - $\sum_{x \in \Omega} p(x) = 1$

- Um evento E é um subconjunto de Ω, e a probabilidade de E é definida como:
  - $\Pr[E] = \sum_{x \in E} p(x)$

Exemplos

- Lançamento um dado com 6 lados
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $p(x) = \frac{1}{6}$ para cada $x \in \Omega$
  - Para o evento $E =$ números pares $= \{2, 4, 6\}$, $\Pr[E] = \frac{1}{2}$
- Lançamento de duas moedas “justas” (com lados 0 e 1)
  - $\Omega = \{00, 01, 10, 11\}$
  - $p(x) = \frac{1}{4}$ para cada $x \in \Omega$
  - Para $E =$ algum $0 = \{00, 01, 10\}$, $\Pr[E] = \frac{3}{4}$

Mais definições

- $\overline{E}$ é o evento de que E não ocorre, e $\Pr\left[\,\overline{E}\,\right] = 1 - \Pr[E]$
- $E_1 \wedge E_2$ é o evento de que ambos E_1 e E_2 ocorrem
  - Por definição, $\Pr[E_1 \wedge E_2] \leq \Pr[E_1]$
  - E_1 e E_2 são **independentes** sse $\Pr[E_1 \wedge E_2] = \Pr[E_1] \cdot \Pr[E_2]$

- $E_1 \vee E_2$ é o evento de que algum dentre E_1 e E_2 ocorre
  - Por definição, $\Pr[E_1 \vee E_2] \geq \Pr[E_1]$

Limitante da união

- **Proposição A.7 (Limitante da união).** Sejam E_1 e E_2 eventos, então:

$$\Pr[E_1 \vee E_2] \leq \Pr[E_1] + \Pr[E_2]$$

Aplicando várias vezes

$$\Pr\left[\bigvee_{1 \leq i \leq k} E_i\right] \leq \sum_{1 \leq \mathrm{i} \leq k} \Pr[E_i]$$

Probabilidade condicional

- A *probabilidade condicional de* E_1 *dado* E_2 é definida a seguir:

$$\Pr[\, E_1 \mid E_2 \,] = \frac{\Pr[E_1 \wedge E_2]}{\Pr[E_2]}$$

- Segue que $\Pr[E_1 \wedge E_2] = \Pr[\, E_1 \mid E_2 \,] \cdot \Pr[E_2]$

- **Teorema A.8 (Teorema de Bayes).** Se $\Pr[E_2] \neq 0$, então

$$\Pr[\, E_1 \mid E_2 \,] = \frac{\Pr[\, E_2 \mid E_1 \,] \cdot \Pr[E_1]}{\Pr[E_2]}$$

Lei da probabilidade total

- Sejam $E_1, \ldots, E_n$ eventos tal que $\Pr[E_1 \vee \cdots \vee E_n] = 1$ e $\Pr[E_i \wedge E_j] = 0$ para $i \neq j$
  - $E_1, \ldots, E_n$ formam uma partição do espaço de todos os eventos possíveis
- Então para qualquer evento F

$$\Pr[F] = \sum_{i=1}^{n} \Pr[F \wedge E_i]$$

Usaremos bastante o formato

$$\Pr[F] = \sum_{i=1}^{n} \Pr[F \mid E_i] \cdot \Pr[E_i]$$

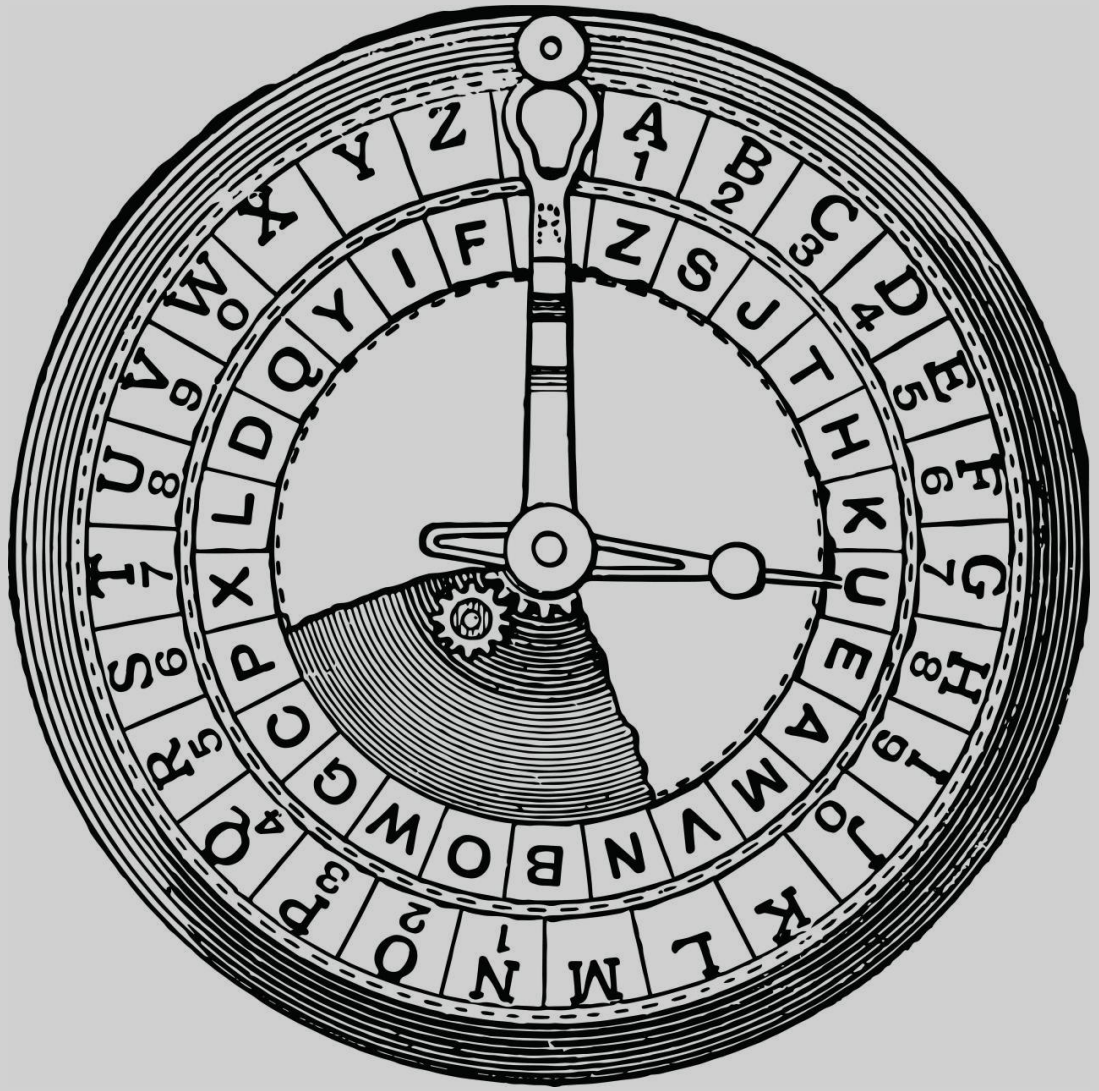
Variáveis aleatórias

- Uma **variável aleatória (discreta, real)** X é uma função que mapeia elementos de Ω para números reais

- Na prática, pode-se imaginar que X é uma variável cujo valor é escolhido de maneira probabilística de um conjunto S de números reais

- **Exemplo:** Se $\Omega =$ o resultado de 100 lançamentos de moeda, X pode ser o número de lançamentos que resultaram em cara
  - $\Pr[X = 100] = \frac{1}{2^{100}}$

Exemplos com cifras simétricas



Definições

- Uma **cifra simétrica** é definida por três algoritmos Gen, Enc, e Dec, e um espaço de mensagens $\mathcal{M}$
- Gen: algoritmo probabilístico que obtém uma chave k
  - $\mathcal{K}$ é o conjunto de todas as chaves possíveis
- Enc: algoritmo **probabilístico** que recebe $k \in \mathcal{K}$ e $m \in \mathcal{M}$ e retorna um texto cifrado c
  - Notação: $c \leftarrow \mathrm{Enc}_k(m)$
  - $\mathcal{C}$ é o conjunto de todos os textos cifrados que podem ser produzidos por Enc
- Dec: algoritmo **determinístico** que recebe k e c e retorna a mensagem usada para obter c
  - Com probabilidade 1, vale que $\mathrm{Dec}_k(c) = m$ para $c \leftarrow \mathrm{Enc}_k(m)$

Algumas variáveis aleatórias

- K denota a chave produzida por Gen
  - $\Pr[K = k]$ é a probabilidade de que a chave $k \in \mathcal{K}$ é produzida por Gen
- M denota a mensagem que será cifrada
  - $\Pr[M = m]$ é a probabilidade de que a mensagem assume o valor $m \in \mathcal{M}$
- C denota o texto cifrado produzido por Enc
  - $\Pr[C = c]$ é a probabilidade de que o texto cifrado assuma o valor $c \in \mathcal{C}$
- Assumimos que K e M são independentes
- C é independente de K e/ou M?
  - Não!

Exemplo (Cifra de César)

- $\mathcal{K} = \{0,1, \dots, 25\}$ e $\Pr[K = k] = \frac{1}{26}$ para cada $k \in \mathcal{K}$

- Dada a seguinte distribuição sobre $\mathcal{M}$
  - $\Pr[M = a] = 0.7$ e $\Pr[M = z] = 0.3$

- Qual o valor de $\Pr[C = b]$?
  - Ou $M = 0 = a$ e $K = 1$ ou $M = 25 = z$ e $K = 2$

$$\Pr[C = b] = \Pr[M = a \wedge K = 1] + \Pr[M = z \wedge K = 2]$$

$$= \Pr[M = a] \cdot \Pr[K = 1] + \Pr[M = z] \cdot \Pr[K = 2]$$

$$= 0.7 * \left(\frac{1}{26}\right) + 0.3 * \left(\frac{1}{26}\right) = \frac{1}{26}$$

Probabilidades condicionais

- Podemos calcular probabilidades condicionais também
- Qual a probabilidade de que a mensagem a foi cifrada, se o texto cifrado é b?
- Pelo Teorema de Bayes

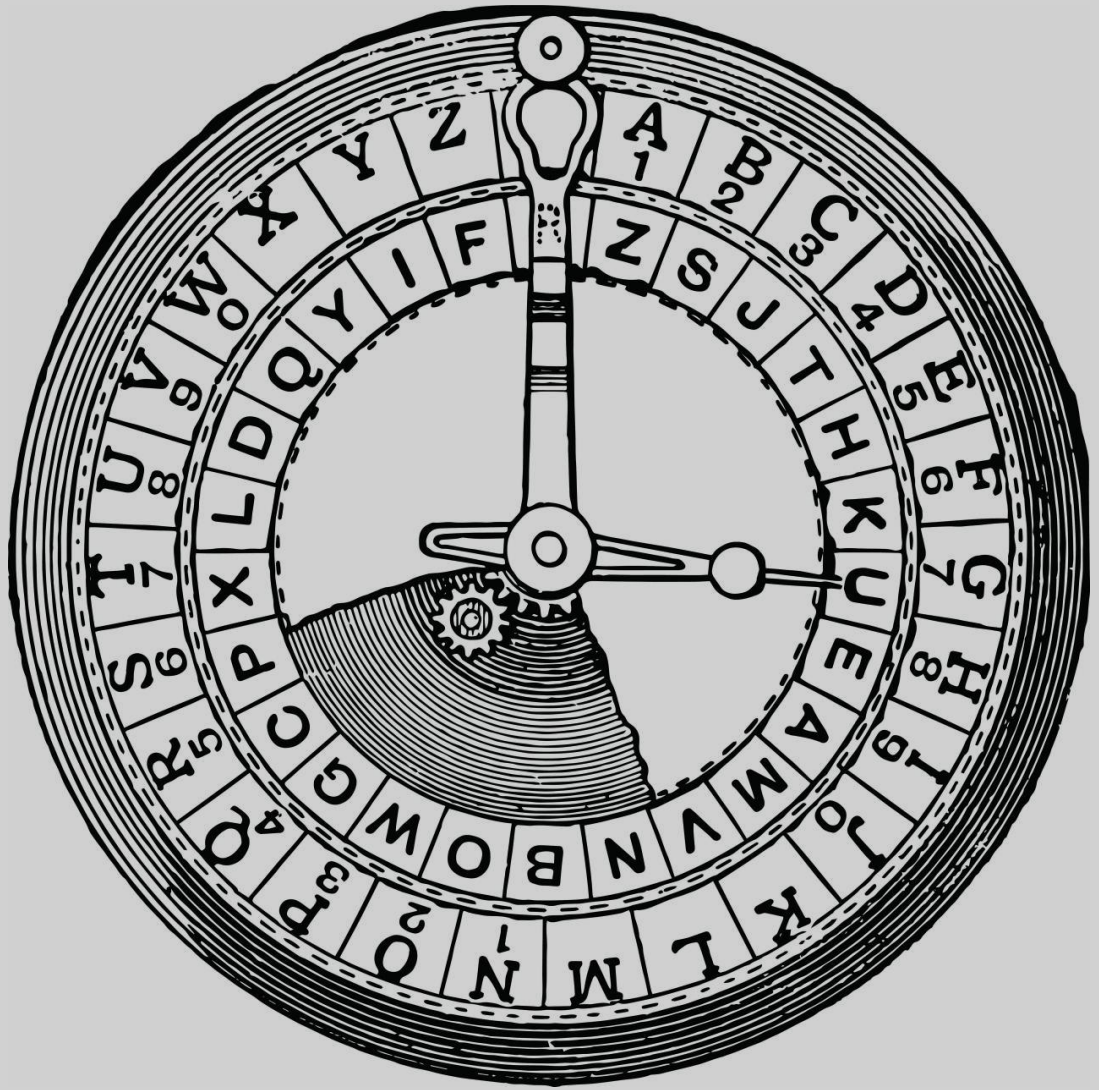
$$\Pr[\, M = a \mid C = b \,] = \frac{\Pr[\, C = b \mid M = a \,] \cdot \Pr[M = a]}{\Pr[C = b]}$$

$$= 0.7 \cdot \frac{\Pr[\, C = b \mid M = a \,]}{1/26}$$

Note que $\Pr[\, C = b \mid M = a \,] = 1/26$ pois se $M = a$, a única maneira de $C = b$ é se $K = 1$ (o que ocorre com probabilidade $1/26$

Concluímos que $\Pr[\, M = a \mid C = b \,] = 0.7$

Segredo perfeito



Segredo perfeito

- **Definição.** Uma cifra $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ com espaço de mensagens $\mathcal{M}$ é **perfeitamente secreta** se para toda distribuição de probabilidade sobre $\mathcal{M}$, toda mensagem $m \in \mathcal{M}$, e todo texto cifrado $c \in \mathcal{C}$ para o qual $\Pr[C = c] > 0$:

$$\Pr[\, M = m \mid C = c \,] = \Pr[M = m]$$

- **Definição alternativa:** para todos $m, m' \in \mathcal{M}$, e todo $c \in \mathcal{C}$

$$\Pr[\mathrm{Enc}_K(m) = c] = \Pr[\mathrm{Enc}_K(m') = c]$$

Um exemplo

- $\mathcal{M} = \{\text{compre, venda}\}$
- $\Pr[\, M = \text{compre}\,] = 0.6$
- $\Pr[\, M = \text{venda}\,] = 0.4$

- Alice cifrou uma mensagem m e obteve c
- Antes de ver c, já sabíamos das probabilidades de $m = \text{compre}$ e $m = \text{venda}$
- Depois de ver c, algo mudou?
  - As probabilidades continuam as mesmas de antes!
- Não aprendemos nada novo sobre a mensagem

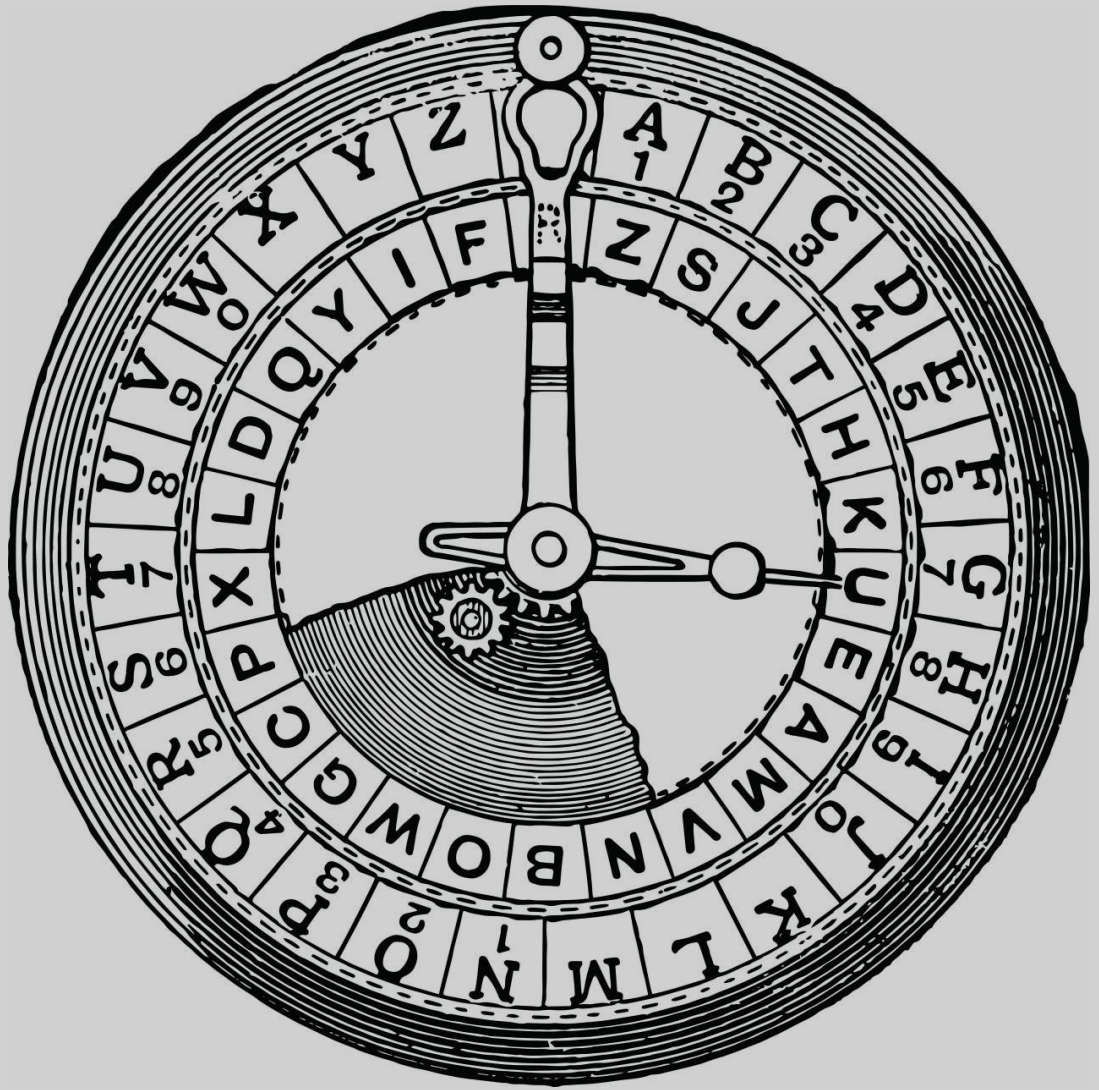
Vigenère não(?) é perfeitamente secreto

- Considere usar a cifra de Vigenère com chave de tamanho 2 e textos de tamanho 3
- Como provar que não é perfeitamente secreto?
- Considere as possíveis mensagens (cada uma com probabilidade $1/2$)
  - $m_1 = aaa$
  - $m_2 = abc$
- Qual é o valor da seguinte probabilidade?

$$\Pr[m = m_1 \mid C = bcd]$$

- Zero, pois para qualquer escolha de chave k, teríamos que $c \leftarrow \mathrm{Gen}_k(m_1)$ teria o primeiro caractere igual ao terceiro
- Conhecer o texto cifrado revela informação sobre a mensagem original

One-time pad



One-time pad

- Fixe um inteiro $\ell > 0$
- $\mathcal{M} = \mathcal{K} = \mathcal{C} = \{0,1\}^{\ell}$ (strings binárias de tamanho ℓ)
- Gen: escolhe uma chave em $\{0,1\}^{\ell}$ de maneira uniforme (cada uma com probabilidade $2^{-\ell}$
- Enc: dada uma chave $k \in \{0,1\}^{\ell}$ e uma mensagem $m \in \{0,1\}^{\ell}$, tem como saída $c := k \oplus m$
- Dec: dada uma chave $k \in \{0,1\}^{\ell}$ e um texto cifrado $c \in \{0,1\}^{\ell}$, tem como saída $m := k \oplus c$
- $a \oplus b$ denota o XOR bit a bit entre as strings a e b

O one-time pad é perfeitamente secreto

- **Teorema 2.9.** O one-time pad é perfeitamente secreto
- **Demonstração:** pelo Teorema de Bayes, para todos $m \in \mathcal{M}$ e $c \in \mathcal{C}$:

$$\Pr[\, M = m \mid C = c \,] = \frac{\Pr[\, C = c \mid M = m \,] \cdot \Pr[M = m]}{\Pr[C = c]}$$

- Vamos calcular as probabilidades separadamente

O one-time pad é perfeitamente secreto

Para qualquer valor de $m \in \mathcal{M}$

$$\begin{aligned}\Pr[C = c \mid M = m] = \Pr[Enc_k(m) = c] &= \Pr[m \oplus K = c] \\ &= \Pr[K = m \oplus c] \\ &= 2^{-\ell}\end{aligned}$$

E então

$$\begin{aligned}\Pr[C = c] &= \sum_{m \in \mathcal{M}} \Pr[C = c \mid M = m] \cdot \Pr[M = m] \\ &= 2^{-\ell} \cdot \sum_{m \in \mathcal{M}} \Pr[M = m] \\ &= 2^{-\ell}\end{aligned}$$

O one-time pad é perfeitamente secreto

- Finalmente

$$\Pr[\, M = m \mid C = c \,] = \frac{\Pr[\, C = c \mid M = m \,] \cdot \Pr[M = m]}{\Pr[C = c]}$$

$$= 2^{-\ell} \cdot \frac{\Pr[M = m]}{2^{-\ell}}$$

$$= \Pr[M = m]$$

- Como desejado

Teorema de Shannon

- **Teorema 2.11.** Seja $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ uma cifra simétrica com espaço de mensagens $\mathcal{M}$, para o qual $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$. A cifra é perfeitamente secreta se e somente se:

1. Cada chave $k \in \mathcal{K}$ é escolhida com (a mesma) probabilidade $1/|\mathcal{K}|$ por Gen.
2. Para todo $m \in \mathcal{M}$ e $c \in \mathcal{C}$, existe uma única chave $k \in \mathcal{K}$ tal que $\mathrm{Enc}_k(m)$ tem como saída c.

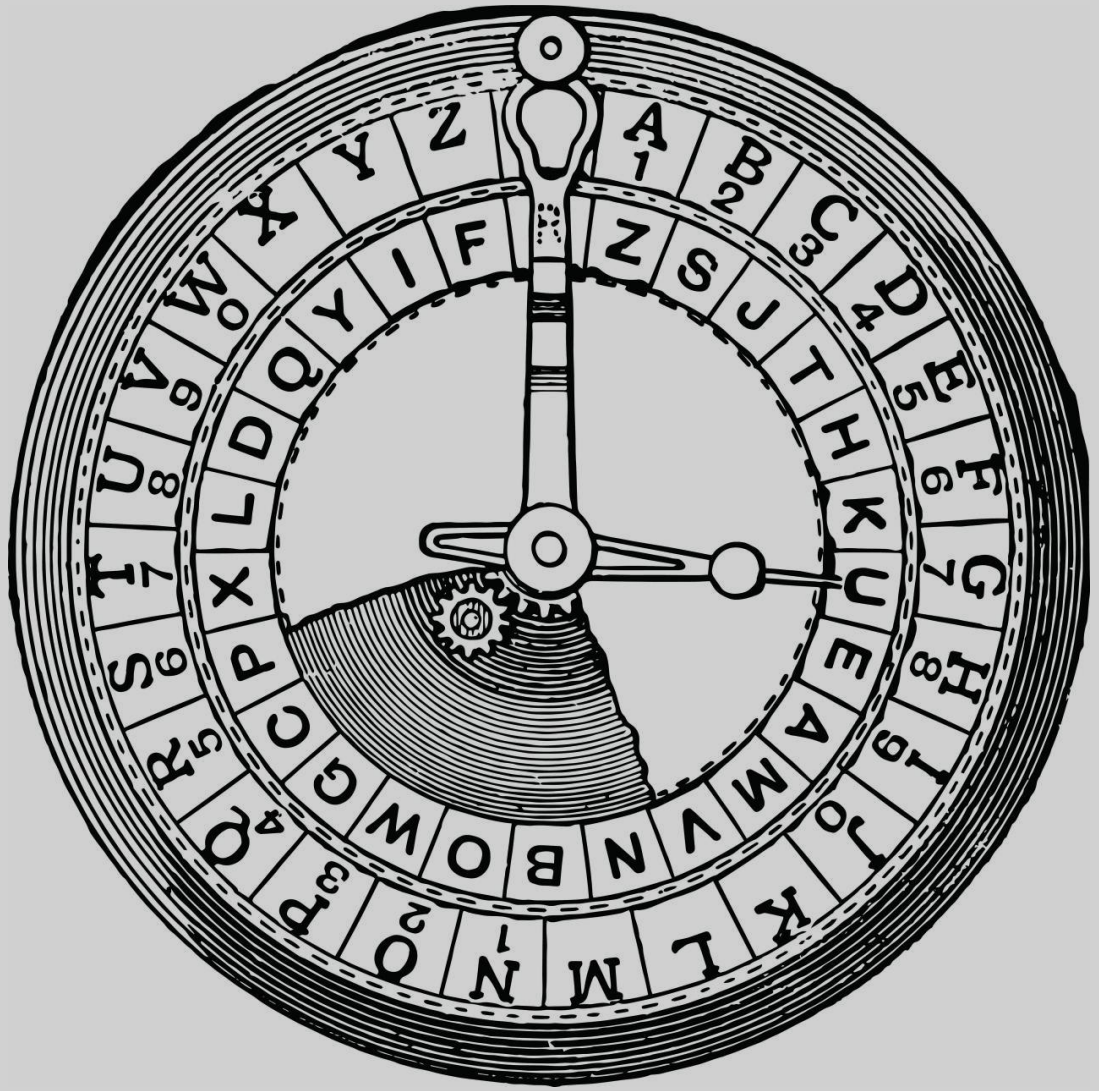
Ideia da prova (detalhes no livro)

- ($\Leftarrow$) A condição 2 implica que qualquer texto cifrado c poderia ser obtido a partir de uma mensagem m.
  - Como existe apenas uma chave que mapeia m para c,
  - e cada chave é escolhida com a mesma probabilidade,
  - o segredo perfeito da cifra segue como no one-time pad.
- ($\Rightarrow$) Se uma cifra é perfeitamente secreta, então para todos m, c deve existir alguma chave mapeando m para c.
  - Como$|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$, só pode existir uma chave com essa propriedade.
  - Se cada chave não for escolhida com a mesma probabilidade que as outras, a cifra deixa de ser perfeitamente secreta.

Algumas observações/perguntas

- A segurança não depende da distribuição sobre as mensagens
  - O esquema serve para qualquer tipo de aplicação
- Utilizado para criptografar a comunicação através do “telefone vermelho” durante a guerra fria
  - As chaves eram trocadas usando entregadores de confiança carregando maletas cheias de papel com caracteres aleatórios
- O que acontece ao reutilizar a mesma chave para criptografar uma segunda mensagem?

Limitações



Chave de uso único

- Reutilizar a mesma chave revela informação sobre as mensagens
- $c = m \oplus k$ e $c' = m' \oplus k$
- Que informação alguém que intercepte c e c' consegue sobre as mensagens m e m'?
  - $c \oplus c' = (m \oplus k) \oplus (m' \oplus k) = m \oplus m'$
  - Descobre exatamente onde as duas mensagens diferem
  - Os EUA conseguiram usar este tipo de análise para recuperar cerca de 3000 mensagens da União Soviética criptografadas com o one-time pad (https://en.wikipedia.org/wiki/Venona_project)
- Essa limitação vale para **toda** cifra perfeitamente secreta

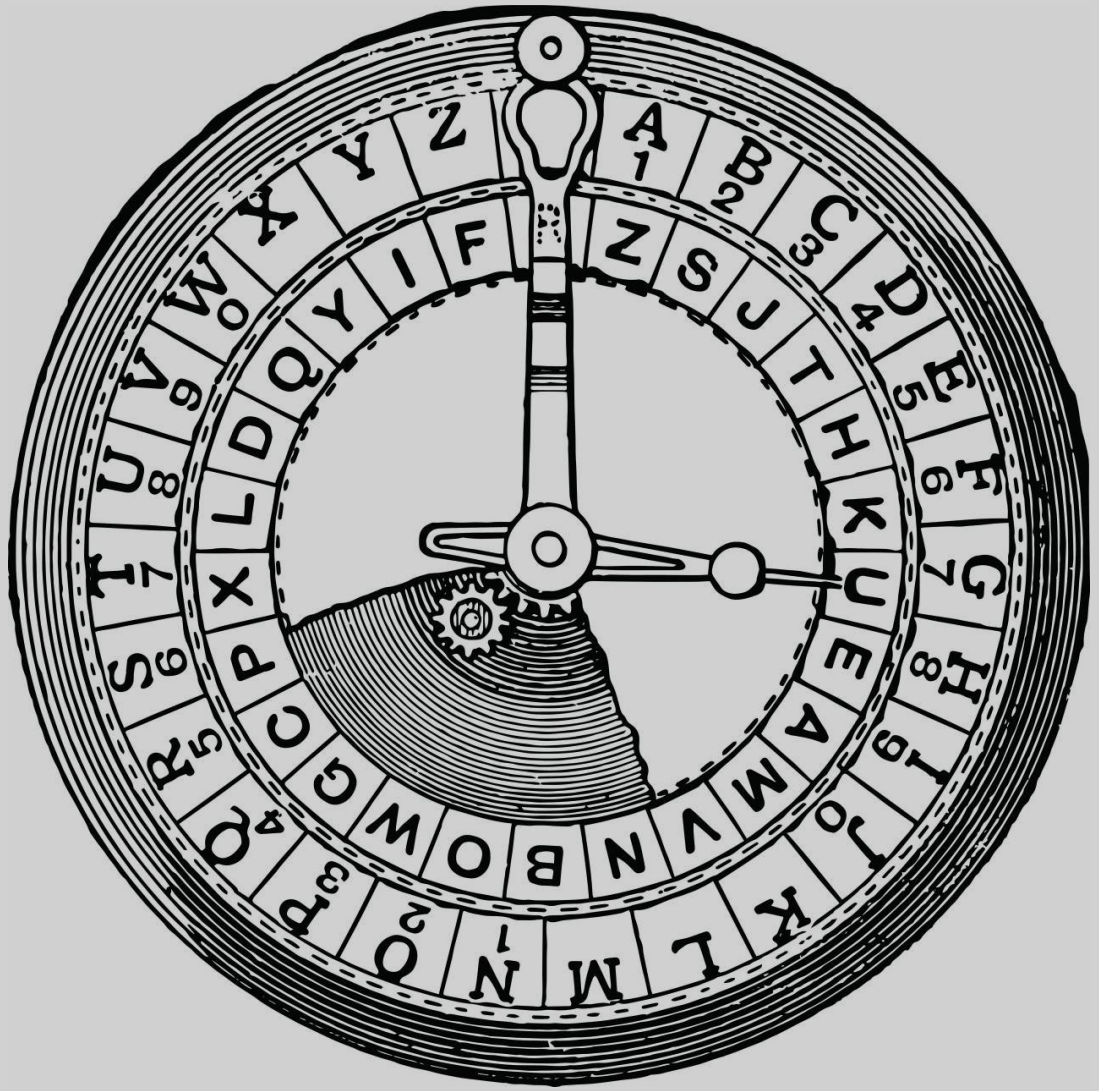
Tamanho da chave

- **Teorema.** Se $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ é uma cifra perfeitamente segura com espaço de mensagens $\mathcal{M}$ e espaço de chaves $\mathcal{K}$, então $|\mathcal{K}| \geq |\mathcal{M}|$
- **Demonstração:** assuma que $|\mathcal{K}| < |\mathcal{M}|$, distribuição uniforme sobre $\mathcal{M}$ e seja $c \in \mathcal{C}$ um texto cifrado que ocorre com probabilidade não-nula
- Seja $\mathcal{M}(c)$ as mensagens para as quais c pode ser decifrado

$$\mathcal{M}(c) = \{m \mid m = \mathrm{Dec}_k(c) \text{ para algum } k \in \mathcal{K}\}$$

- Segue que $|\mathcal{M}(c)| \leq |\mathcal{K}|$ (por que?)
- Nesse caso, deve haver uma mensagem $m' \in \mathcal{M}$ que não está em $\mathcal{M}(c)$, e o esquema não é perfeitamente secreto (por que?)

Para fechar



Para fechar

- Hoje
  - Probabilidade
  - Segredo perfeito
  - One-time pad
  - Limitações
- Próxima aula
  - Contornar estas limitações
  - Segurança **computacional**
    - Relaxamento de segredo perfeito