Topicos
Aula Filtro de Bloom

Ivan Sendin

FACOM - Universidade Federal de Uberlândia
ivansendin@yahoo.com,sendin@ufu.br

10 de setembro de 2024

- 3 Propriedades criptográficas
- Enderecos
- Prova de Trabalho
- Commit
- Árvore de Merkle
- ECC/Assinaturas

- O MD5 é contemporaneo ao DES
- DES:56 bits e MD5 128 bits
- Pq?
- Talvez o ataque de força bruta ao DES/56 seja *equivalente* ao ataque de força bruta do MD5/128...

Paradoxo do aniversario

- Qual é a probabilidade de 2 pessoas fazerem aniversario no mesmo dia??
- $\frac{1}{365}$
- (Considerando uma distribuição uniforme...)
- Aniversario no mesmo dia $\equiv$ colisao
- Qual o tamanho minimo um grupo deve ter para que a probabilidade seja $\geq \frac{1}{2}$

Paradoxo do aniversario

- O grupo deve ter 23 pessoas...
- Como?
  1. **O número de pares cresce de forma quadrática**

Paradoxo do aniversario- Contas

- P(haver pelo menos um par)
- (dois pares, um trio)
- P(haver...) = 1 -P(Não haver...)
- P(Não haver...) = $\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots$
- P(Não haver...) = 0.49 para 23 termos
- P(haver...) = 0.51 para 23 termos

Paradoxo do aniversario- Contas

- Encontrar um colisão é mais fácil do que a nossa intuição diz
- É mais fácil do que inverter
- (Assintóticamente)

Paradoxo do aniversario - Forja....

- Produzir uma colisão
- $M_i \neq M_j$
- $\mathcal{H}(M_i) == \mathcal{H}(M_j)$
- Produzir n sequencias aleatorias
- Calcular o hash de cada uma delas
- Para $n = 2^{h/2}, p() \approx 0.5$
- Em resumo/prática: o tamanho do hash seguro é o dobro do tamanho da chave segura
- Detalhe: gasta muita memoria!

- Dada uma função de hashing $\mathcal{H}$ podemo criar um número arbitrariamente grande de funções de hashing...
- $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \ldots \mathcal{H}_n$

- Definimos:

$$\mathcal{H}_i(x) := \mathcal{H}(i||x)$$

- Concatenamos o identificador de função na mensagem
- (Pode ser antes, depois,envelopando,...)
- $\mathcal{H}_i(x) \neq \mathcal{H}_j(x)$, para $i \neq j$
- (É o que esperamos....)

- Bloom, H. (1970), "Space/Time Trade-offs in Hash Coding with Allowable Errors"
- Um filtro de Bloom é uma ED com a seguinte interface:
- *void insert(k)*
- *contains(k)*
  - probabilistico
  - não é não
  - sim é talvez
  - (não melhora se perguntar de novo!)
- Parece um *set*...sem a possibilidade de um iterator

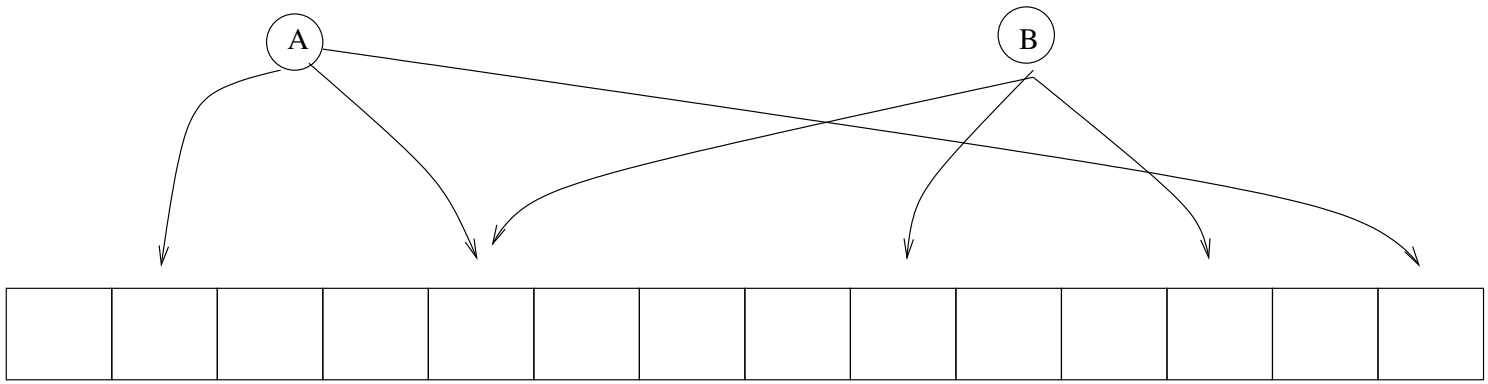
- O *construtor* do um FB recebe dois parâmetros:
- n que determina o tamanho de um vetor V de booleanos
- Inicializado em **False**
- h que determina o numero de funções de hashing utilizadas: $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \ldots \mathcal{H}_h$

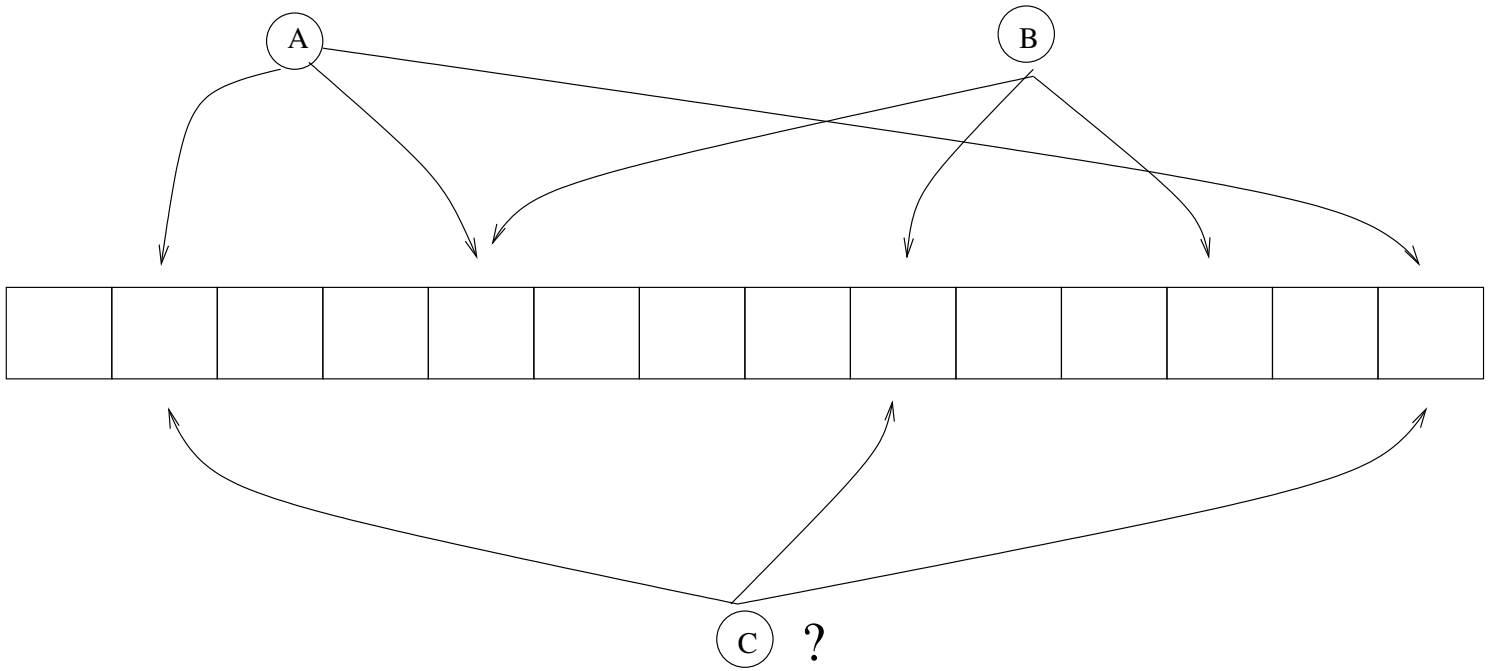
```
void insert(k):

    for i in 1..h:
        t = Hi(k)
        V[ t % n ] = 1
```

$O(1)$ (tamanho da entrada: numero de elementos...)

```
void contem(k):

    for i in 1..h:
        t = Hi(k)
        Se V[ t % n ] == 0
            return False
    return True
```

$O(1)$


A
B


A
B
C
?

- É possível calcular a probabilidade de falsos positivos
- É possível estimar o tamanho do conjunto
- União e intersecção (probabilistica) de conjuntos

- URLs Maliciosas no Google Chrome
- Consulta *confidencial* a uma base de dados **grande!!** Bitcoin/Cliente Leve

- Towards to fair trade with item validation using Bloom Filters and Smart Contracts
  Private Set Intersection
  IC/TCC
- ED Hibrida: Árvore + FB (BigData)