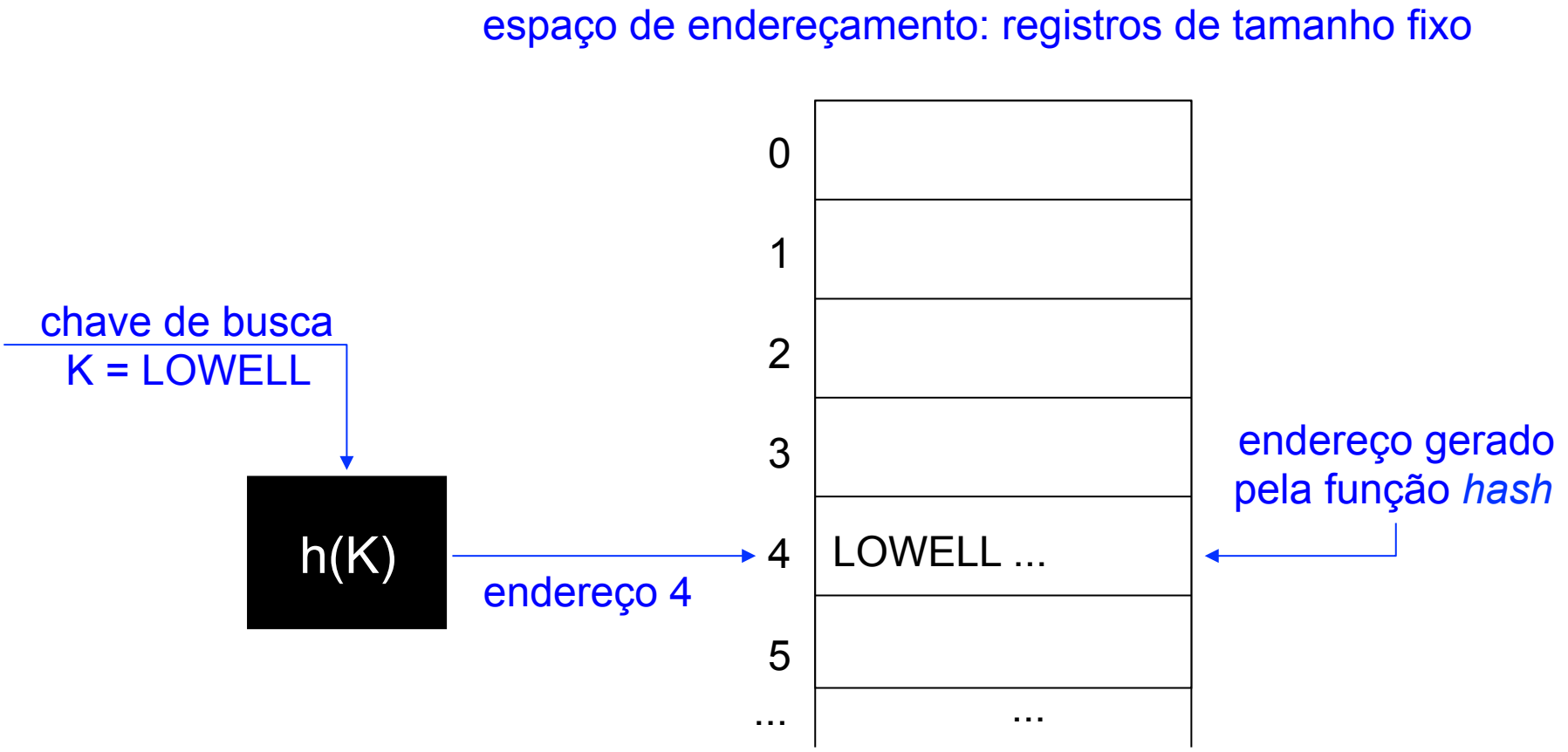
Hashing

Profa. Dra. Cristina Dutra de Aguiar Ciferri

Hashing

- Função *hash*
  - caixa preta que produz um endereço toda vez que uma chave de busca é passada como parâmetro
- Endereço resultante
  - usado para o armazenamento e a recuperação de registros no arquivo de dados
- Nomenclatura
  - h(K) → endereço
    - K: chave de busca

Hashing



Duas chaves diferentes podem ser transformadas para o mesmo endereço: COLISÃO

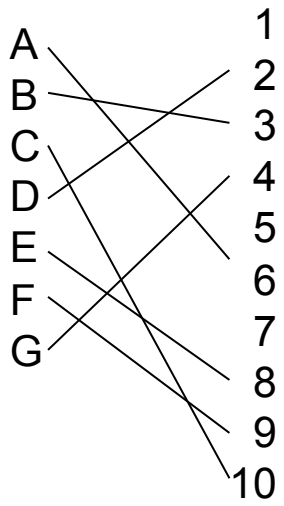
Exemplos de Métodos de *Hash*

- Pegar o resto da divisão da chave pelo tamanho do espaço disponível
- Examinar as chaves em busca de um padrão
- Segmentar a chave em diversos pedaços e depois fundir os pedaços
- Dividir a chave por um número
- Elevar a chave ao quadrado e pegar o meio
- Transformar a base

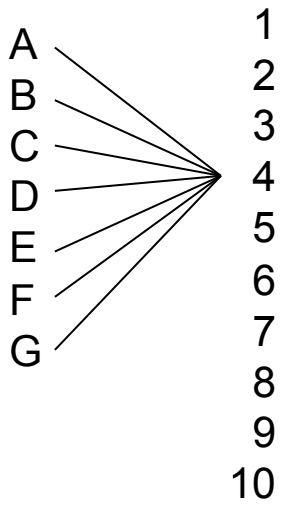
Distribuição de Registros

- Como uma função *hash* distribui (espalha) os registros no espaço de endereços?

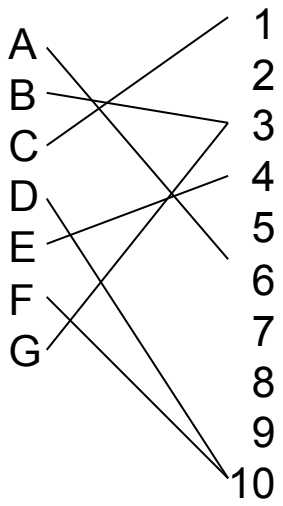
(a) melhor caso



(b) pior caso



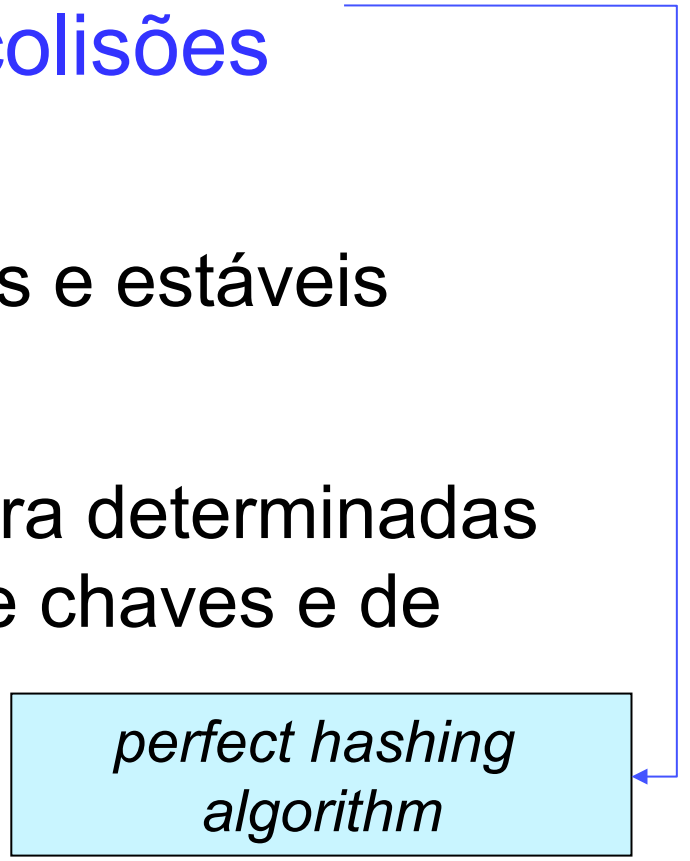
(c) caso aceitável



7 registros x 10 endereços

Colisão: Solução 1

- Encontrar um algoritmo de *hashing* perfeito que não produza colisões
- Cenário de uso
  - conjunto de dados pequenos e estáveis
- Limitação
  - abordagem não indicada para determinadas configurações de número de chaves e de dinâmica dos dados



Colisão: Solução 2

- Encontrar um algoritmo de *hashing* que produza poucas colisões
- Objetivo
  - evitar o agrupamento de registros em certos endereços
- Funcionalidade
  - espalhar os registros aleatoriamente no espaço disponível para armazenamento
  - distribuir o mais uniformemente possível

Colisão: Solução 3

- Ajustar a forma de armazenamento dos registros
  - Possibilidade 1: aumentar o espaço de endereçamento para um mesmo conjunto de registros
  - Possibilidade 2: armazenar mais de um registro em um único endereço
    - cada endereço é suficientemente grande para armazenar diversos registros

Colisão: Solução 3

- Para as possibilidades 1 e 2
  - soluções convencionais para o tratamento da colisão
    - overflow progressivo
    - *hashing* duplo
    - encadeamento