

# Hashing Universal

Guillermo Palma

Universidad Simón Bolívar  
Departamento de Computación y T.I.

CI-2612: Algoritmos y Estructuras de Datos II



(USB)

Hashing Universal

CI-2612 enero-marzo 2020

1 / 12

## Plan

### 1 Hashing Universal



(USB)

Hashing Universal

CI-2612 enero-marzo 2020

2 / 12

# Sobre el Hashing Universal

- En el mundo real las claves a ser insertadas en una tabla no son aleatoriamente distribuidas
- Un usuario mal intencionado, sabiendo la función a usar, puede escoger las claves a ser insertadas en una tabla, de manera tal que la función de hash las asigne todas a una misma casilla y entonces se obtiene el peor caso posible  $O(n)$ , para las operaciones sobre la tabla de hash
- Se quiere tener funciones de hash que produzcan índices de casillas de tablas de hash, que sean aleatorios, sin importar las claves que sean usadas



# Características del Hashing Universal

## Idea Básica

Se quiere seleccionar una buena función de hash al azar e independiente de las claves a insertar, de un conjunto de funciones de hash, al comienzo de la ejecución de las operaciones sobre una tabla de hash

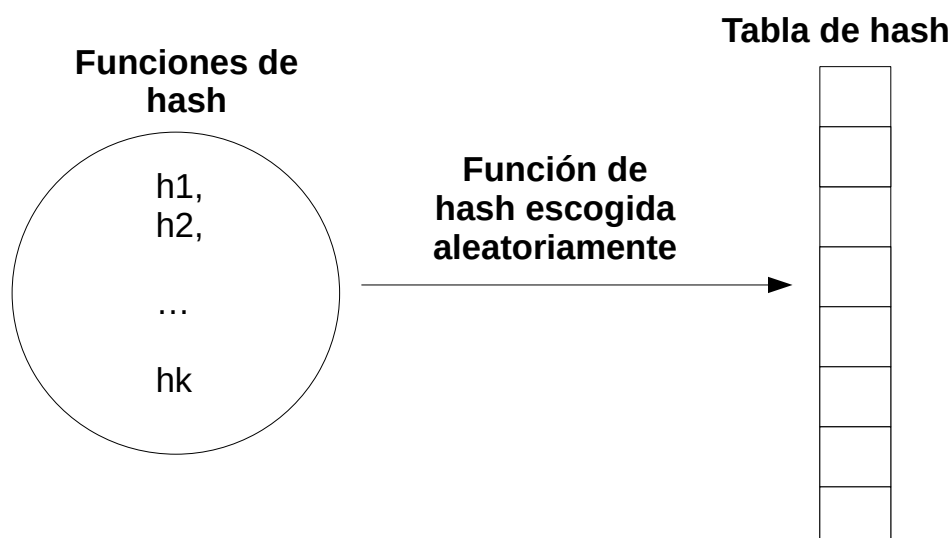


## Características del Hashing Universal

- El seleccionar de manera aleatoria funciones de hash garantiza con gran probabilidad que una secuencia de operaciones sobre una tabla de hash, no pueda obtener el peor caso
- La idea es que una misma secuencia de operaciones tenga un resultado diferente cada vez que se ejecutan, como en el caso de RandomizedQuicksort.



## Ejemplo de Hashing Universal



## Definición de Funciones de Hash Universales

- Sea  $m$  el número de casillas de una tabla de hash

$$H = \{h(k) : U \rightarrow \{0, 1, \dots, m-1\}\}$$

- Un colección de funciones de hash  $H$  es universal si cumple que:

$$\forall k \neq l \quad \text{se tiene que} \quad |h \in H : h(k) = h(l)| = \frac{|H|}{m}$$



## Análisis del Hashing universal

- Se quiere saber cuál es la probabilidad de una colisión usando Hashing universal
- Se tiene que escogiendo una función cualquiera  $h \in H$ , se cumple que  $k \neq l \rightarrow h(k) = h(l)$ , en consecuencia

$$\text{Probabilidad}(h(k) = h(l)) = \frac{|H|/m}{|H|} = \frac{1}{m}$$



## Análisis del Hashing universal

Usando Hashing universal la probabilidad de una colisión entre dos claves diferentes  $k$  y  $l$  es a lo sumo  $\frac{1}{m}$ , si  $h(k)$  y  $h(l)$  fueron escogidas aleatoriamente



## Análisis del Hashing universal

- Considere un Hashing universal con las colisiones resueltas por encadenamiento
- Suponga un secuencia de  $n$  operaciones, insertar, buscar y eliminar, en una tabla con  $m$  casillas y donde se hacen  $O(m)$  inserciones
- Como hay  $O(m)$  inserciones, el factor de carga es  $a = O(1)$ . Las eliminaciones y las inserciones son tiempo constante. Las búsquedas requieren tiempo  $O(1 + a) = O(1)$
- Se tiene que las  $n$  operaciones toman un tiempo  $\Theta(n)$
- En consecuencia no hay una secuencia de operaciones que pueda generar el peor caso.



## Diseñando clases universales de funciones de hash

- Seleccione un primo  $p$  suficientemente grande para todas las claves  $k$  a usar.
- $Z_p = \{0, 1, \dots, p-1\}$
- $Z_p^* = \{1, \dots, p-1\}$
- Se define la función de hash:

$$h_{a,b}(k) = ((ak + b) \bmod p) \bmod m, \quad \forall a \in Z_p^* \text{ y } b \in Z_p$$

- La familia de todas funciones de hash universales es:

$$H_{p,m} = \{h_{a,b} : a \in Z_p^* \text{ y } b \in Z_p\}$$

- Se tiene que  $a, b$  se escogen al comienzo de la ejecución de las operaciones sobre la tabla de hash.

