

Redes entre pares (II)

(peer-to-peer)

Repaso Conceptos

- P2P es un nuevo paradigma, todos son clientes y todos son servidores.
- Todos ofrecen y utilizan recursos.
- No existe autoridad central.
- Sistemas flexibles y robustos.

Repaso Conceptos, II

- Tienen cierta problemática asociada:
 - ¿Cómo acceder a los recursos?
 - ¿Como encontrar recursos?

Repaso Localización

- Se crea una red superpuesta para enrutar búsquedas.
- Sistemas no estructurados.
 - Centralizados (Napster).
 - Super-Nodos (eDonkey).
 - Inundación (Gnutella).
 - Caminos aleatorios (Gia, Dante).
- Sistemas estructurados.

Redes Estructuradas / DHT

- Cada recurso es 'asignado' a un nodo. Este debe de saber dónde está localizado el recurso (aunque no lo gestione/posea).
- Hay un control estricto de las conexiones de la red superpuesta.
- Se basan en el uso de Tablas Hash Distribuidas (DHTs).

Tablas Hash Distribuidas

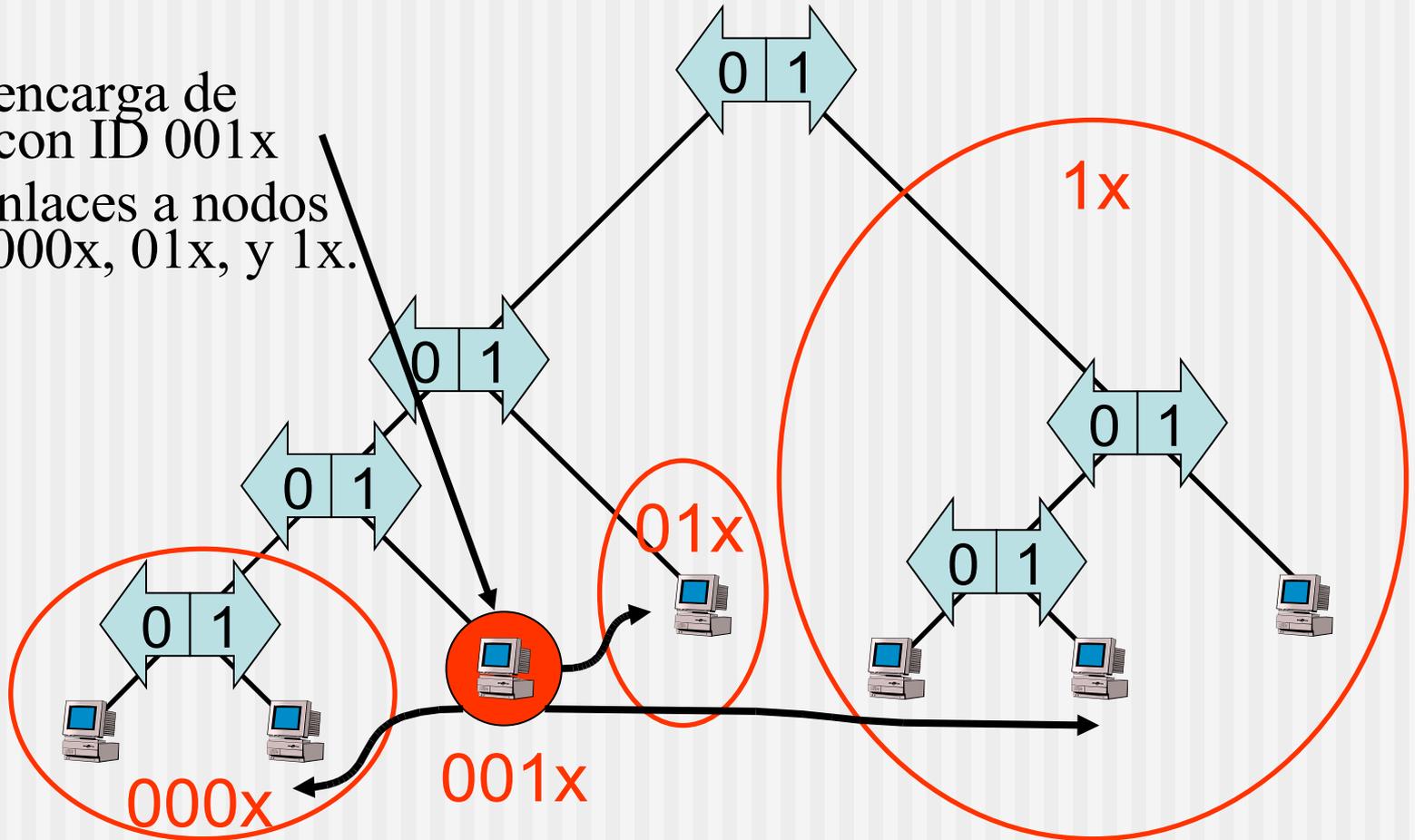
- Se crea un espacio de identificadores.
- Mediante una función hash, a cada recurso y nodo se le asigna un id.
- Cada nodo se encarga de un segmento del espacio, asignado según su id.
- Cada nodo debe saber dónde están los recursos de su segmento.

Enrutamiento en DHT

- Cada nodo conoce sólo un subconjunto de otros nodos.
- Cada uno de esos nodos hace de 'representante' de un área del espacio de ids.
- Se enrutan búsquedas a los vecinos, según el área al que pertenezca la clave buscada.

Ejemplo DHT

Nodo se encarga de recursos con ID 001x
Guarda enlaces a nodos en áreas 000x, 01x, y 1x.

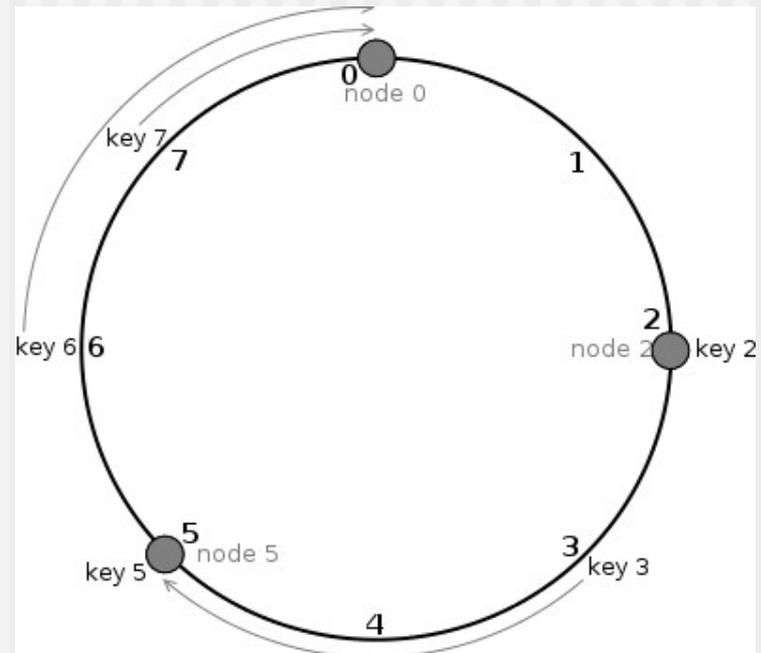


Chord

- Propuesto en 2001 por Stoica *et al.*
- Representa el espacio de identificadores en un círculo.

Ejemplo:

Círculo Chord para longitud de identificadores $m=3$.



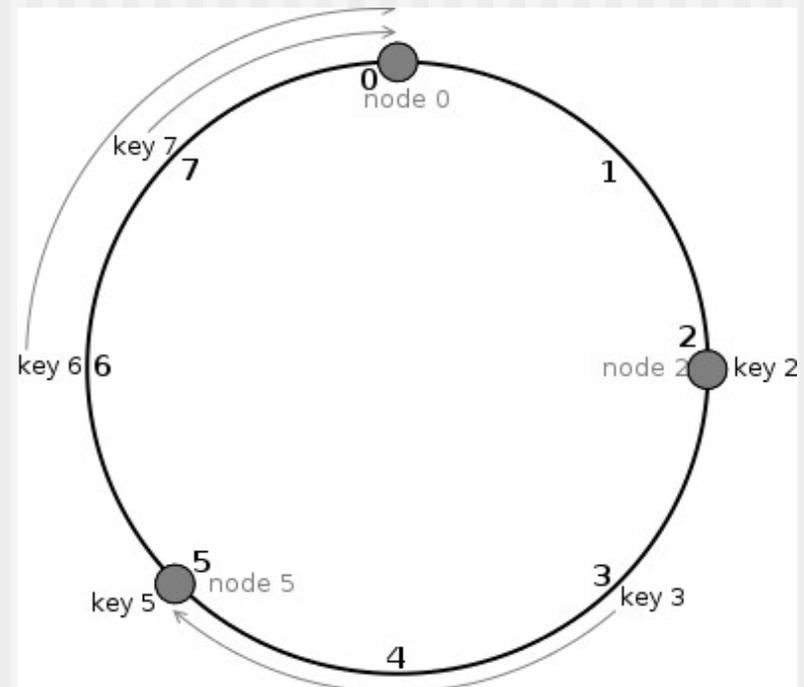
Chord

- Cada recurso es asignado al nodo con el mismo id o sucesor en el círculo de ids.

Ejemplos:

Recurso 2 es asignado a nodo 2.

Recurso 3 es asignado a nodo 5.



Chord

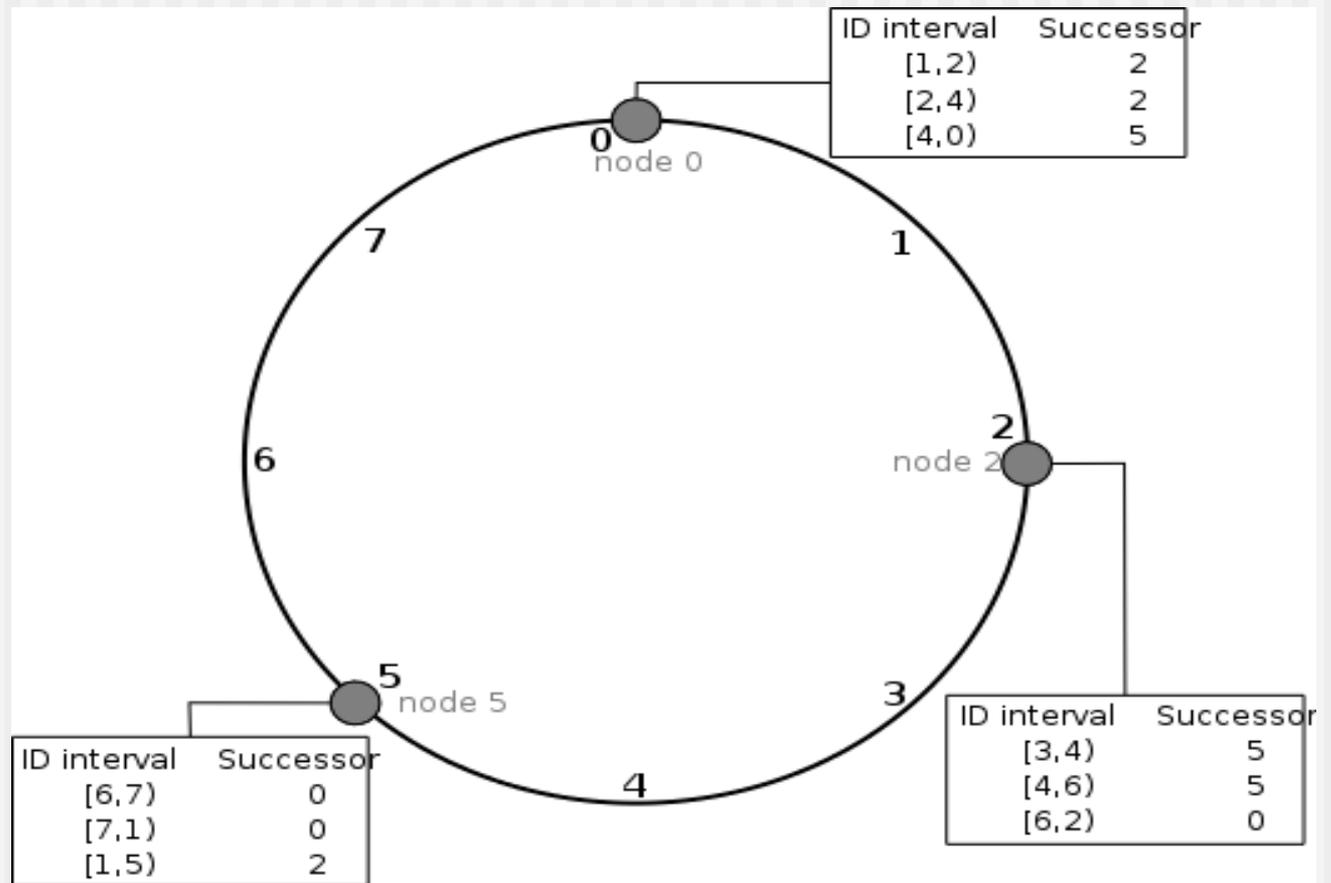
- Cada nodo mantiene una tabla de enrutamiento.
- La entrada i -ésima de la tabla:
 - Representa el intervalo de ids a distancia 2^{i-1} y 2^i-1 del nodo.
 - Contiene el nodo sucesor más cercano dentro del intervalo.

Ejemplo, tabla de enrutamiento nodo 2:

Entrada	Intervalo	Nodo Sucesor
1	[3,4)	5
2	[4,6)	5
3	[6,2)	0

Chord

Ejemplo completo:



Chord

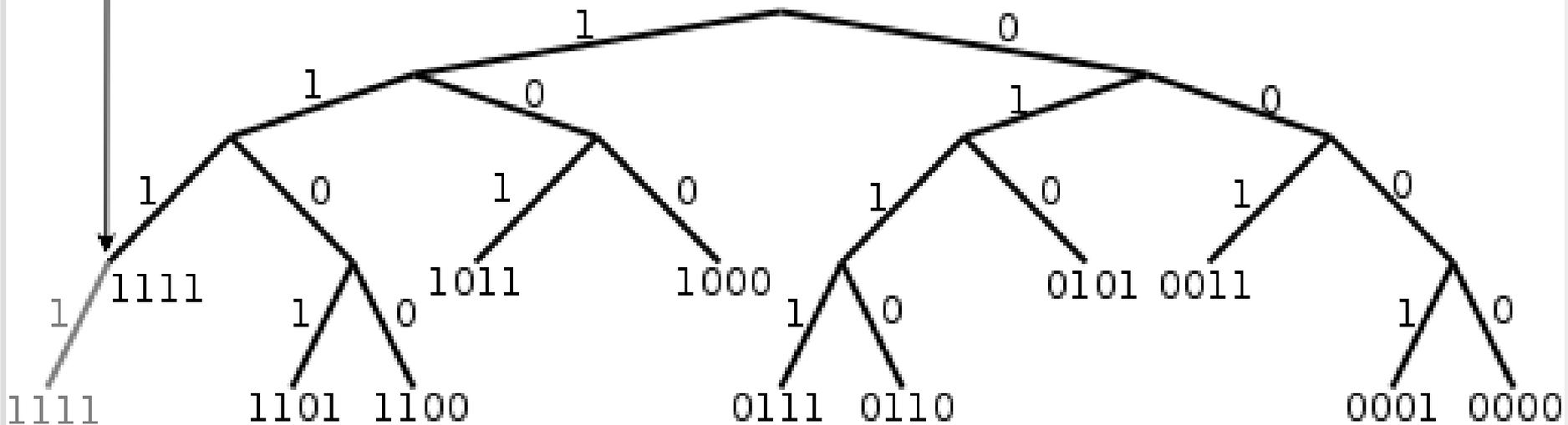
- Al enrutar búsquedas:
 - Si se conoce al nodo sucesor (encargado del id), la búsqueda está terminada.
 - Si no, se reenvía al nodo en la tabla cuyo id es menor que el id del recurso.
 - Las búsquedas acaban en $O(\log_2 N)$ pasos.

Kademlia

- Usa un árbol binario para organizar los ids.

Ejemplo árbol para ids de 4 bits:

Position of node with id=1111, determined by 111 (shortest unique prefix)

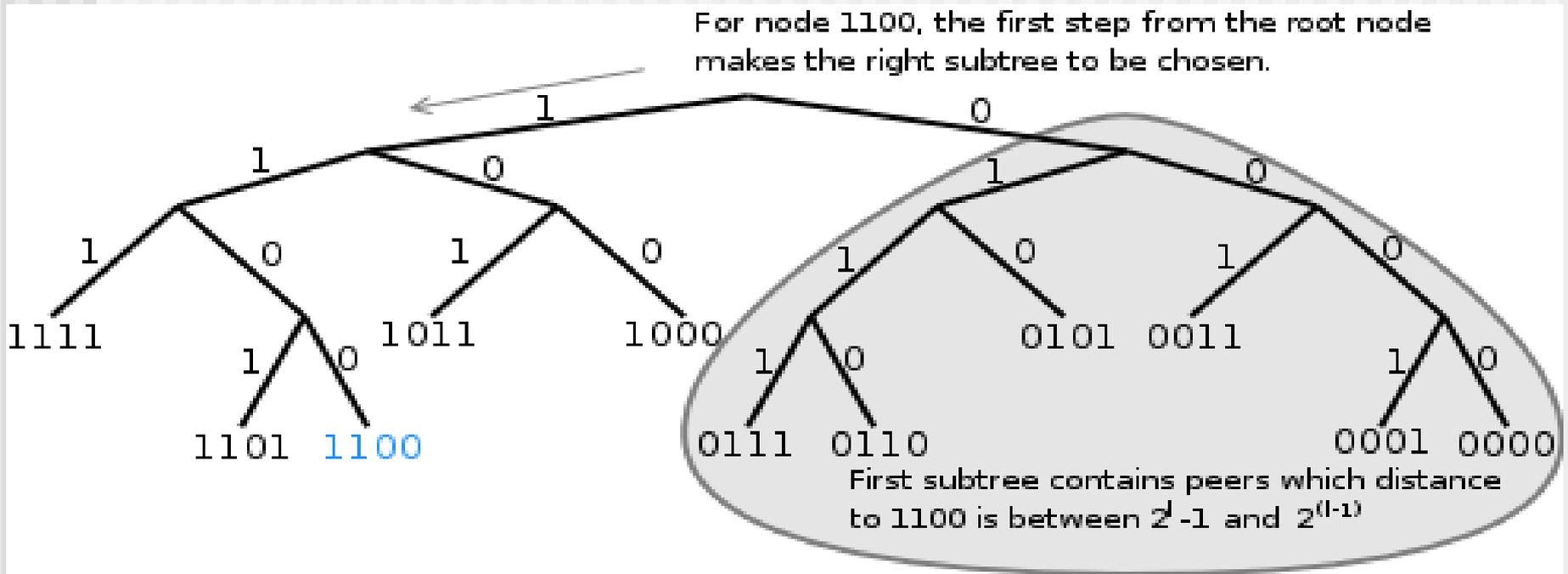


Kademlia

- Un recurso es responsabilidad del primer nodo por encima de él en el árbol.
- Cada nodo organiza el árbol en sub-árboles.
- Empezando desde la raíz, se sigue el camino que lleva al nodo.
- En cada paso, se escoge el subárbol que NO contiene al nodo.

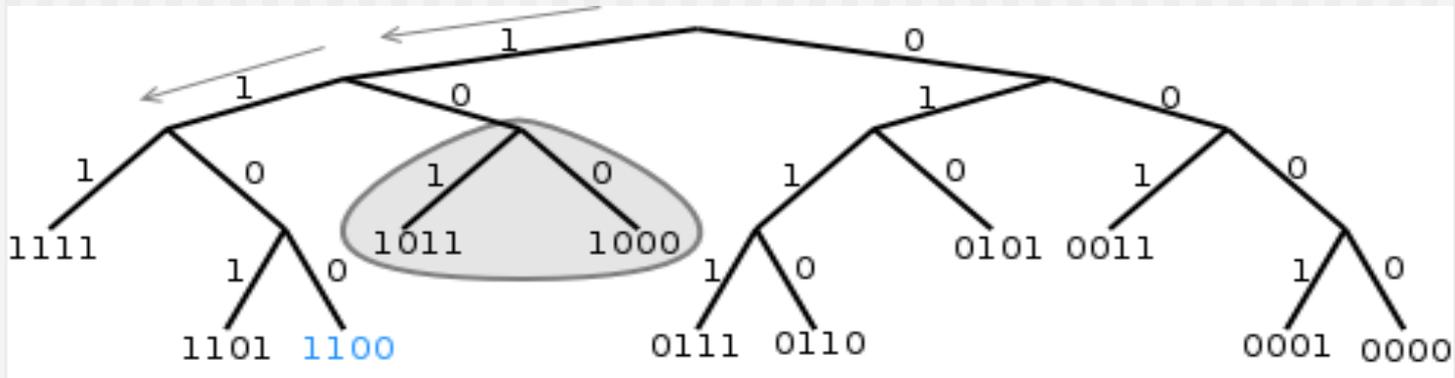
Kademlia

Primer subárbol nodo 1100:

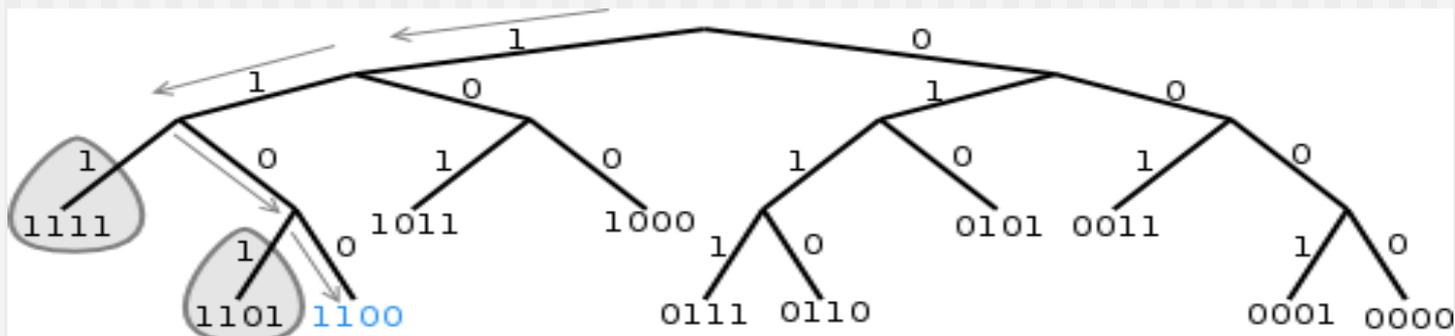


Kademlia

Segundo subárbol nodo 1100:

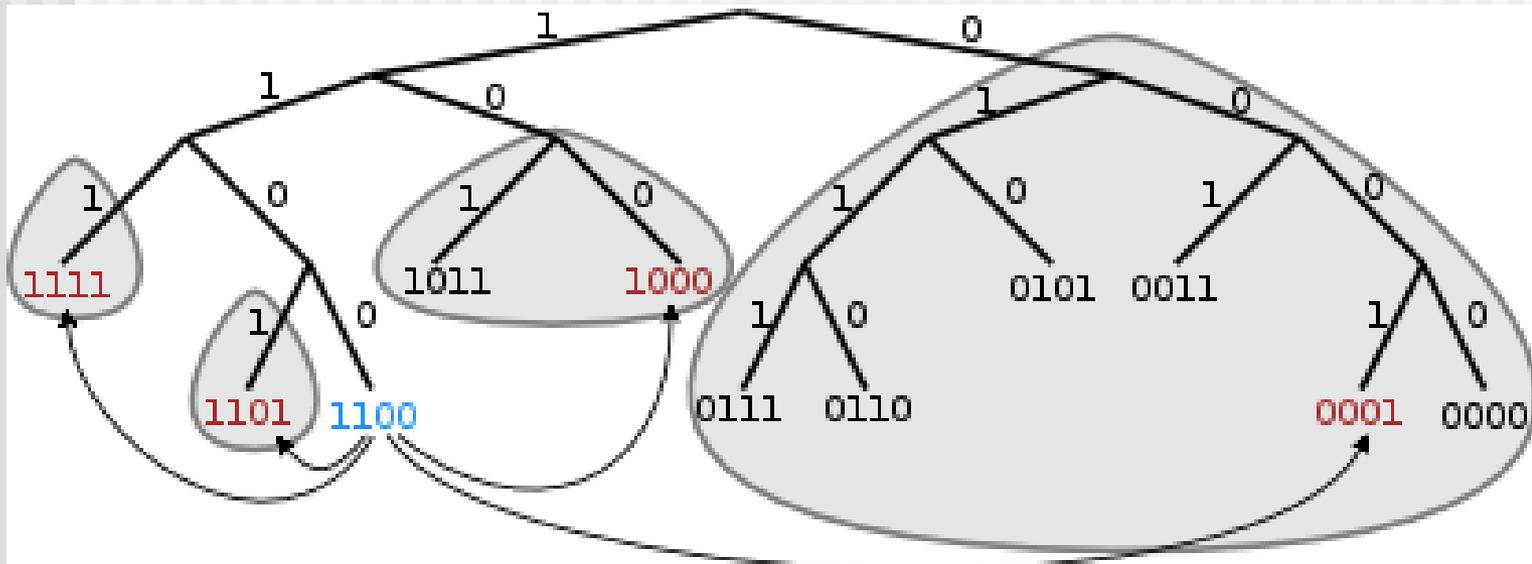


Tercer y cuartos subárboles nodo 1100:



Kademlia

- Cada nodo conoce al menos un nodo dentro de cada subárbol.



Kademlia

- En Kademlia, la distancia se mide usando la operación XOR.

El primer árbol contiene los nodos a distancia $[2^4-1, 2^4-1)$.

El segundo árbol contiene los nodos a distancia $[2^4-1, 2^3-1)$.

- Observación:

Dados x e y dentro de un árbol, para cualquier z de fuera del mismo se cumple que:

$\text{distancia}(x,y) < \text{distancia}(z,x)$, $\text{distancia}(x,y) < \text{distancia}(z,y)$

Kademlia

- Al dirigir búsquedas, se busca al nodo con distancia mínima al recurso buscado.

Ejemplo:

El nodo 1100 quiere localizar el recurso 1010. Comprueba las distancias a los nodos que conoce:

$$1111 \text{ XOR } 1010 = 0101 = 5$$

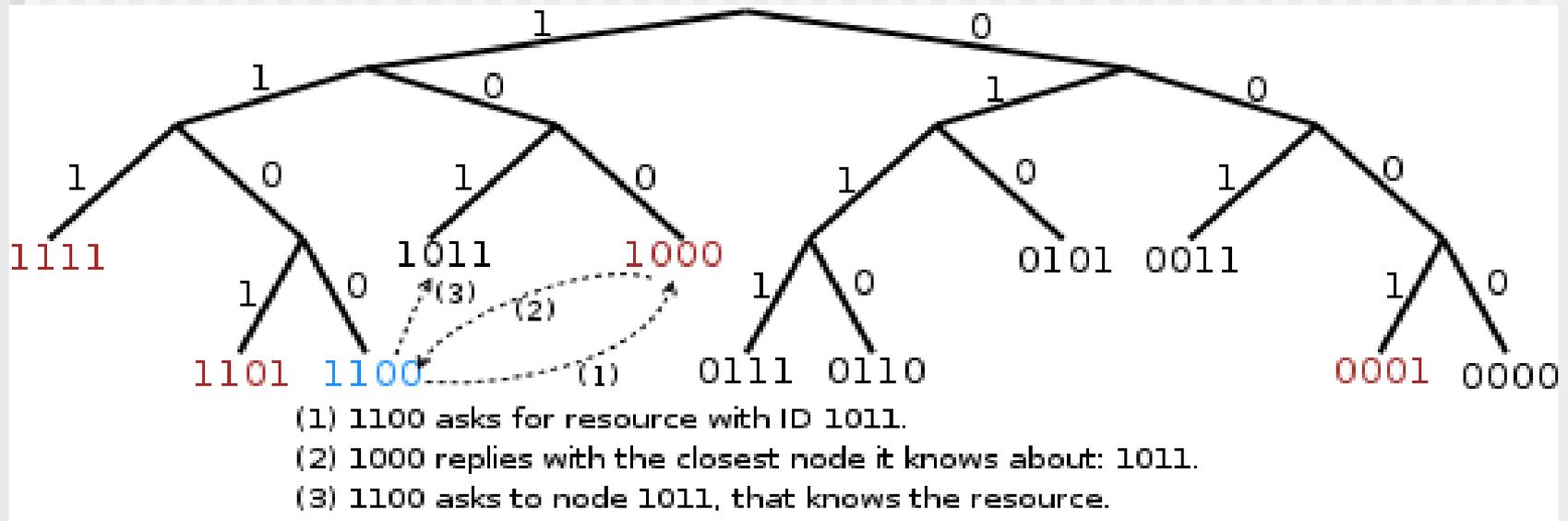
$$1101 \text{ XOR } 1010 = 0111 = 7$$

$$1000 \text{ XOR } 1010 = 0010 = 2 \quad \leq \text{ Nodo 1000 es el más cercano.}$$

$$0001 \text{ XOR } 1010 = 1011 = 11$$

Kademlia

Se pregunta al nodo 1000, que conoce al 1011. El 1011 es el encargado de conocer el recurso 1010.



Las búsquedas no se reenvían. El nodo que buscador ejecuta el mismo proceso hasta que encuentra al par responsable del recurso.

Limitaciones P2P estructurada

- La gestión de entradas y salidas de nodos es costosa.
- Difícil realizar búsquedas por palabras clave.
- No aprovecha replicación 'natural'.