



Sistemas Distribuidos

Grado en Ingeniería Informática

Universidad de Cádiz

Sistemas Peer-to-Peer

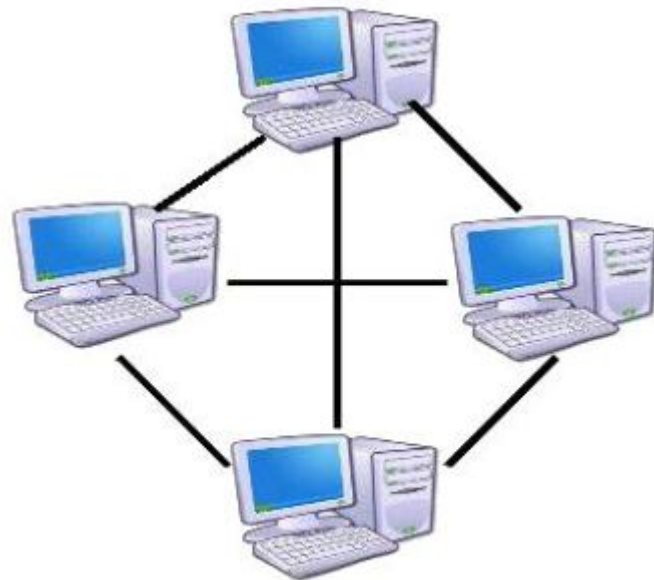
Introducción

¿Qué es P2P?

⊕ ¿Que es P2P?

«Peer-to-peer es un Sistema auto-organizado de entidades iguales, autónomas (peers), cuyo objetivo es el **uso compartido de recursos distribuidos** en un ambiente de red evitando servicios centralizados»

Oram y otros, 2001



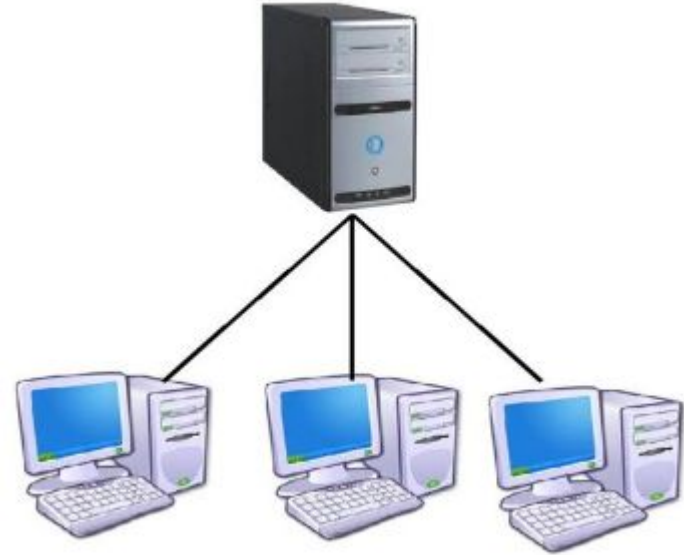
Introducción

¿Qué no es?

➔ Red cliente-servidor

Frente a este tipo de redes, se encuentran las que se basan en una arquitectura **cliente-servidor**, en la que los clientes solicitan recursos a uno o varios servidores. En este tipo de arquitectura, según aumenta el número de usuarios en la red, **la tasa de transferencia disminuye**.

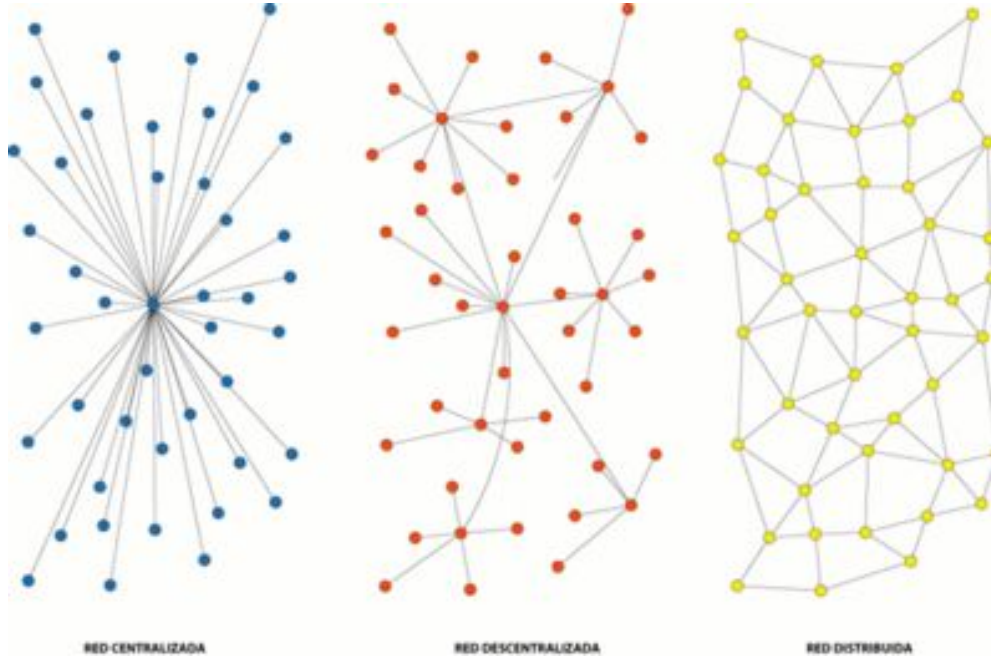
Esto ocurre porque los recursos de los que dispone el servidor se consumen debido al intenso tráfico que se genera.



Definición

Características

⊕ 1/9 Descentralización.



Definición

Características

⊕ 1/9 Descentralización.

Los diferentes miembros (iguales) del sistema son los **propietarios** y tienen el **control de los datos y recursos de su ordenador**. Esto complica la implementación de los sistemas de igual a igual, porque **no hay una visión global del sistema**. Por este motivo, algunos sistemas de igual a igual adoptan aproximaciones híbridas, como Napster o eDonkey, que disponen de directorios centralizados de ficheros, pero los nodos descargan los ficheros directamente de los iguales que aportan los ficheros.

Definición

Características

⊕ 2/9 Escalabilidad.

Una consecuencia directa de la descentralización es la mejora de la escalabilidad. El hecho de que **las operaciones se hagan entre los iguales (peers) hace que el sistema pueda soportar muchas más operaciones** que si se tuviese que recurrir a un nodo que centraliza las operaciones.

Definición

Características

⊕ 3/9 Autoorganización.

En sistemas de igual a igual, esta **autoorganización** es necesaria ya que acostumbran a ser sistemas de gran escala (por ejemplo, es difícil saber cuántos iguales, usuarios y carga hay);

Definición

Características

⊕ 4/9 Coste de la propiedad

En los sistemas de igual a igual, la propiedad es compartida (los recursos que forman parte del sistema los aportan los propios participantes en el sistema). La propiedad compartida **reduce el coste** de poseer el sistema y los contenidos, y el coste de mantenerlo

Definición

Características

⊕ 5/9 Conectividad puntual (ad-hoc connectivity)

Muchos de los iguales que forman un sistema **no están conectados todo el rato**. Se conectan para hacer unas actividades concretas y se vuelven a desconectar.

Definición

Características

⊕ 6/9 Rendimiento

El rendimiento acostumbra a ser una preocupación en los sistemas de igual a igual. Estos sistemas mejoran el rendimiento agregando capacidad de almacenamiento distribuido o capacidad de procesamiento en dispositivos repartidos por toda la red.

- ⊕ REPLICACIÓN
- ⊕ USO DE MEMORIAS CACHÉS
- ⊕ ENCAMINAMIENTO INTELIGENTE

Definición

Características

➦ 7/9 Seguridad

La mayoría de los problemas de los sistemas de igual a igual relacionados con la seguridad son los mismos que tendría cualquier sistema distribuido:

- Cifrado multiclave
- Seguridad del código (que no ejecute o modifique cosas en el ordenador)
- Gestión de derechos digitales (propiedad intelectual)
- Reputación y contabilización (Como de bueno es un igual, se realiza contabilización de los recursos compartidos)

Definición

Características

⊕ 8/9 Transparencia y usabilidad

Es deseable que los sistemas de igual a igual puedan funcionar **independientemente del dispositivo** de los iguales (ordenador, agenda electrónica o PDA, teléfono móvil).

Otro tipo de transparencia que tienen que proporcionar es la relacionada con la **movilidad y la seguridad**. Los usuarios móviles deben poder trabajar independientemente del lugar desde donde estén conectados.

Definición

Características

⊕ 9/9 Resistencia a fallos

Uno de los objetivos de diseño de los sistemas de igual a igual **es evitar que si falla un componente del sistema, todo el sistema deje de funcionar**. Sería deseable que, a pesar de que algún igual dejará de funcionar o que haya particiones en la red, los iguales que queden continuarán colaborando.

Ataques en P2P

Ataques

Las redes P2P promueven la libre compartición de archivos y servicios, es por eso que es más vulnerable a determinados tipos de ataques

- **"envenenamiento" (poisoning):** Consiste en distribuir información y/o archivos que no corresponden a la descripción dada (archivos falsos).
- **"contaminación" (polluting):** Insertar piezas o paquetes malignos dentro de paquetes legítimos, a veces solo para afectar el rendimiento y otras conteniendo malware.
- **"saturación" (DoS, choking):** El tráfico provocado por las redes P2P puede ser excesivo para los sistemas, provocando que la red se ahogue, pero también puede haber ataques orquestados para producir Negación de Servicios.
- **"identificación":** Algunas redes P2P permiten identificar a los usuarios, algo que han aprovechado bastante los organismos de propiedad intelectual para localizar usuarios.
- **"estafas" (scamming):** muchas "empresas" se aprovechan de la ignorancia del usuario y le venden servicios especiales, que no son otra cosa mas que acceso a lo que es gratis, sólo quitándoles su dinero.

Áreas de aplicación

Almacenamiento

⊕ Manejo de Documentos :

Problemática:

- Usualmente organizados en forma centralizada
- Pero grandes porciones de documentos creados en una compañía **son distribuidos entre** PCs sin un repositorio central que tenga algún conocimiento de su existencia.

Solución:

- Las redes P2P crean un **repositorio conectado** de los datos locales de los peers individuales.
- **Indexación y categorización** de datos por cada peer sobre la base de un criterio seleccionado individualmente.
- Agregado de información **autoorganizada desde áreas del conocimiento**.

Áreas de aplicación

Almacenamiento

⊕ Archivos compartidos en P2P

- Almacena contenidos en nodos individuales en vez de un lugar central.
- Peers que bajan archivos los ponen disponibles para otros peers.
- El 70% del tráfico de Internet puede ser atribuido a intercambio de archivos.
- Es la aplicación más extendida de los sistemas P2P.

Áreas de aplicación

Almacenamiento

⊕ Problemas de Búsqueda

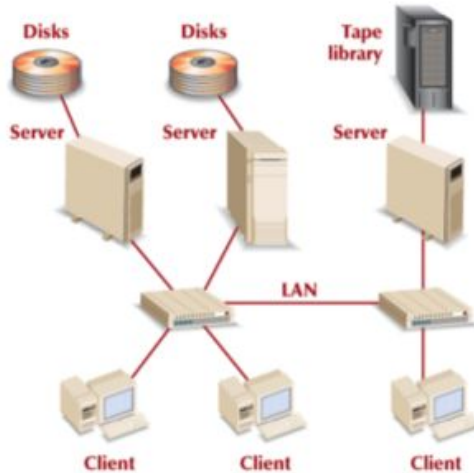
- Localizar recursos es, en general, un problema central de las redes P2P y compartir archivos en particular
- Las redes P2P proveen de diferentes métodos para el almacenaje, búsqueda y recuperación de los archivos:
 - Modelo de directorio centralizado
 - Modelo de requerimiento por “inundación”
 - Modelo de ruteo de documentos

Áreas de aplicación

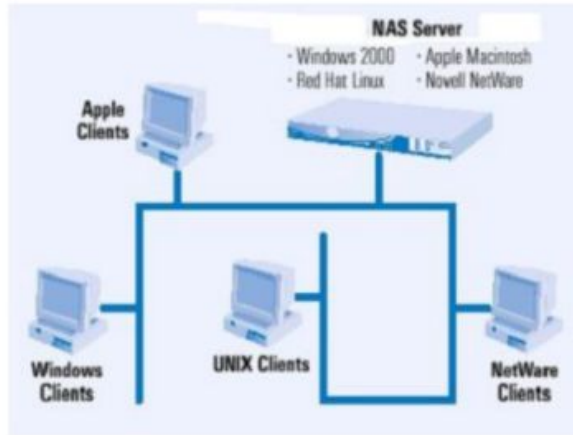
Soluciones centralizadas posibles

⊕ DAS, NAS, SAN

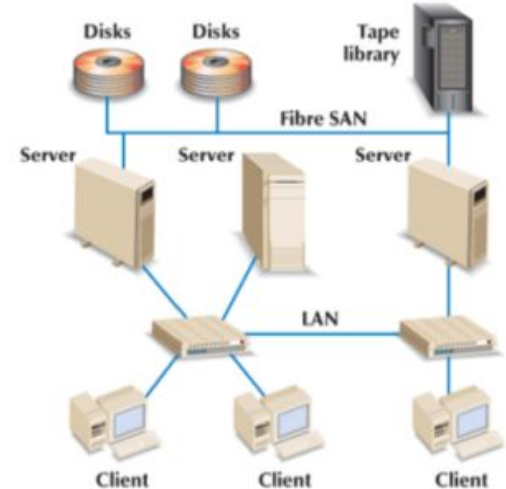
► Direct Attached Storage (DAS)



► Network Attached Storage (NAS)



► Storage Area Networks (SAN)



Áreas de aplicación

Soluciones centralizadas posibles

⊕ **DAS, NAS, SAN**

<https://youtu.be/CczKEbEIR9U>

HBA : Host Bus Adapter



Áreas de aplicación

Soluciones centralizadas posibles

⊕ Problemas DAS, NAS, SAN

- Uso ineficiente del almacenaje disponible.
- Carga adicional en la red de la compañía.
- Necesidad de personal entrenado.
- Soluciones adicionales de respaldo

Áreas de aplicación

Procesamiento

⊕ Aumento de ciclos de procesador

- Incremento de Requerimientos para Computación de Alto Rendimiento (bio-informática, logística, sector financiero)
- Poder de computación de entidades en red con poco uso
- Uso de aplicaciones P2P para uso de ciclos de procesador:
 - Forma un cluster de computadoras independientes conectadas en red que para una simple computadora es transparente y todos los nodos en red son combinados en una computadora lógica simple.
 - Lograr poder de computación que aún las más caras supercomputadoras escasamente ofrecen.
 - “Grid Computing”

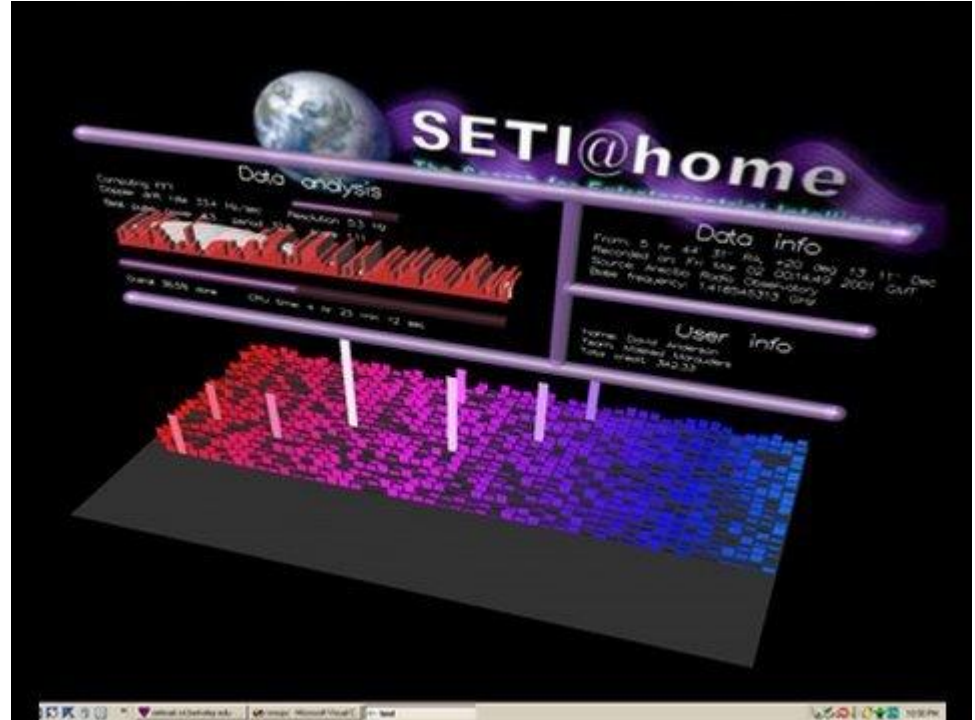
→ *Globus Toolkit: <http://toolkit.globus.org/toolkit/>*

Áreas de aplicación

Procesamiento




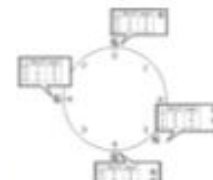
➡ Aumento de ciclos de procesador

https://youtu.be/_aIJV5aQR68



Arquitecturas de las redes p2p

Resumen general

P2P sin Estructura			P2P Estructurado
P2P Centralizado	P2P Puro	P2P Híbrido	Basado en DHT
<ol style="list-style-type: none">1. Todas las características de P2P incluidas2. Una entidad central es necesaria para proveer el servicio3. La entidad central es una especie de Db índice/grupo <p>Ejemplo: Napster</p>	<ol style="list-style-type: none">1. Todas las características de P2P incluidas2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad3. → Entidades no centrales <p>Ejemplos: Gnutella 0.4, Freenet</p>	<ol style="list-style-type: none">1. Todas las características de P2P incluidas2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad3. → Entidades centrales dinámicas <p>Ejemplo: Gnutella 0.6, JXTA</p>	<ol style="list-style-type: none">1. Todas las características de P2P incluidas2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad3. → Entidades no centrales4. Las conexiones en overlay son "fijas" <p>Ejemplos: Chord, CAN</p>
			
1 st Gen.			2 nd Gen.

Arquitecturas de las redes p2p

Modelo de directorio centralizado

⌕ Modelo de directorio centralizado

- El servicio de indexado es provisto centralmente **por una entidad de coordinación**.
- Un requerimiento de **búsqueda es atendido por la entidad coordinadora** que presenta una lista de peers que tienen los archivos requeridos .
- Luego el peer obtiene los archivos respectivos **directamente de los otros peers** que los ofrecen.

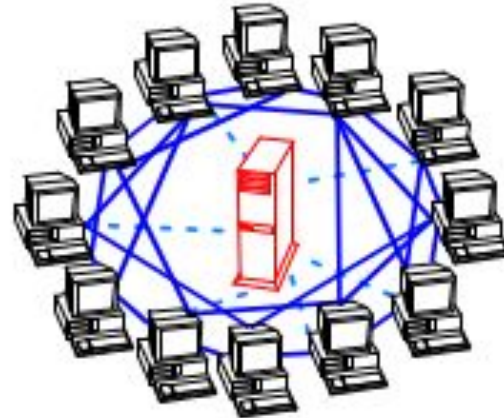


RED CENTRALIZADA

Arquitecturas de las redes p2p

Modelo de directorio centralizado

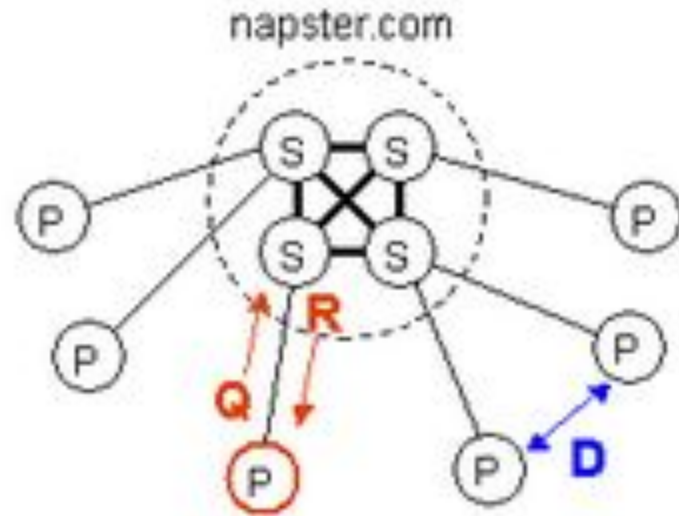
➡ Modelo de directorio centralizado



Arquitecturas de las redes p2p

Modelo de directorio centralizado

➦ Modelo de directorio centralizado

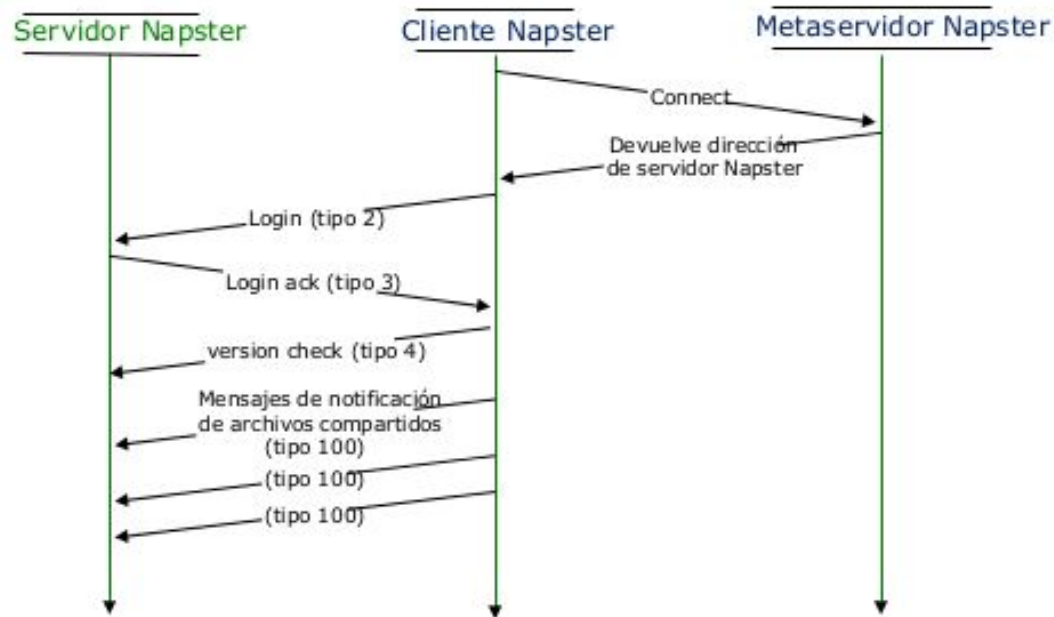


Napster

Arquitecturas de las redes p2p

Modelo de directorio centralizado

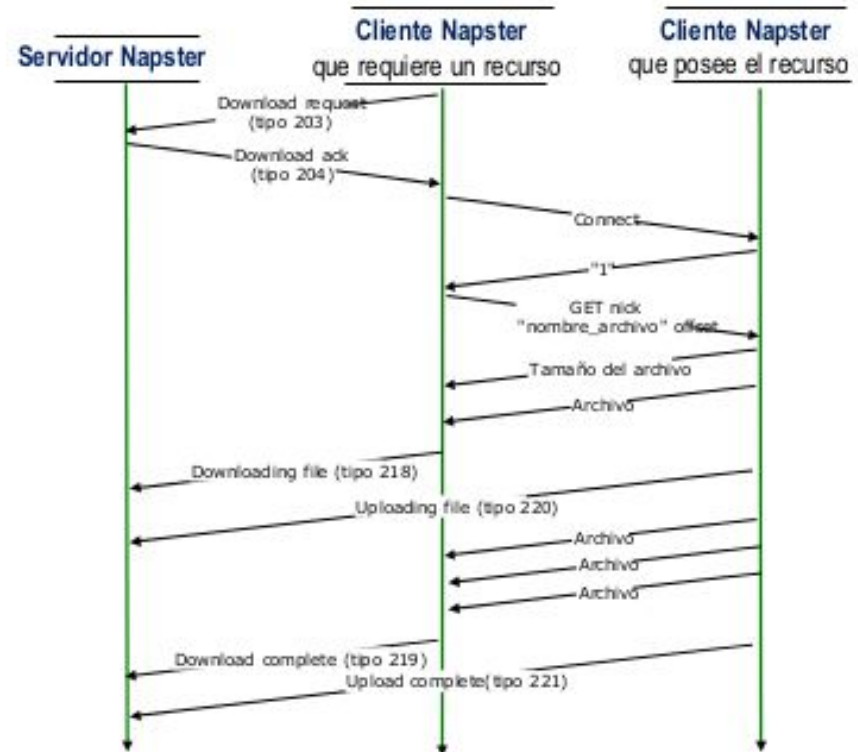
➦ Modelo de directorio centralizado (conexión)



Arquitecturas de las redes p2p

Modelo de directorio centralizado

➤ Modelo de directorio centralizado (descarga)



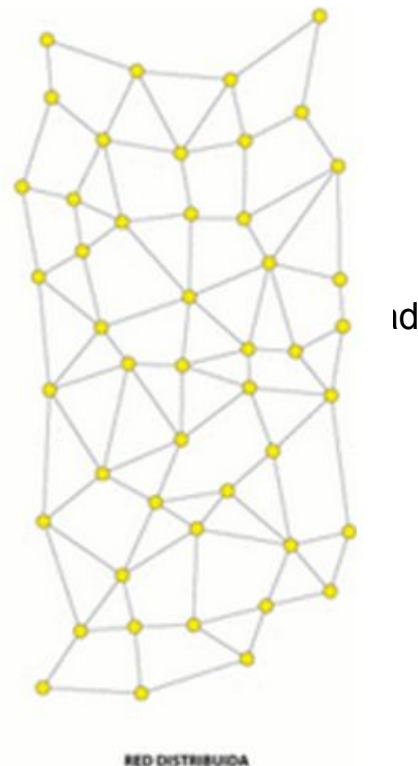
Arquitecturas de las redes p2p

Modelo de descentralizado o puro

⊕ Modelo de directorio centralizado

- **Sin autoridad central** de coordinación (todos los peers son iguales).
- El requerimiento de búsqueda pasa por un determinado número de peers.
- Si no pueden responder el requerimiento, lo pasan sobre otros nodos hasta c
(ttl=time-to-live).
- Cuando el archivo requerido ha sido localizado, los resultados positivos de la búsqueda son enviados a la entidad que los requirió.
- El peer puede ahora bajar el archivo deseado directamente de las entidades que lo ofrecen

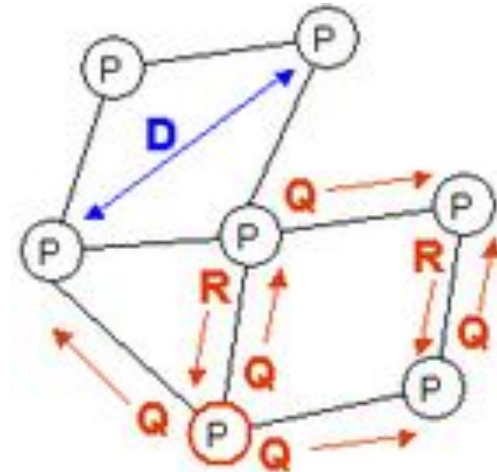
Problemas: Escalabilidad y búsqueda de documentos



Arquitecturas de las redes p2p

Modelo descentralizado o puro

⌕ Modelo descentralizado o puro



Gnutella

Arquitecturas de las redes p2p

Modelo de descentralizado o puro

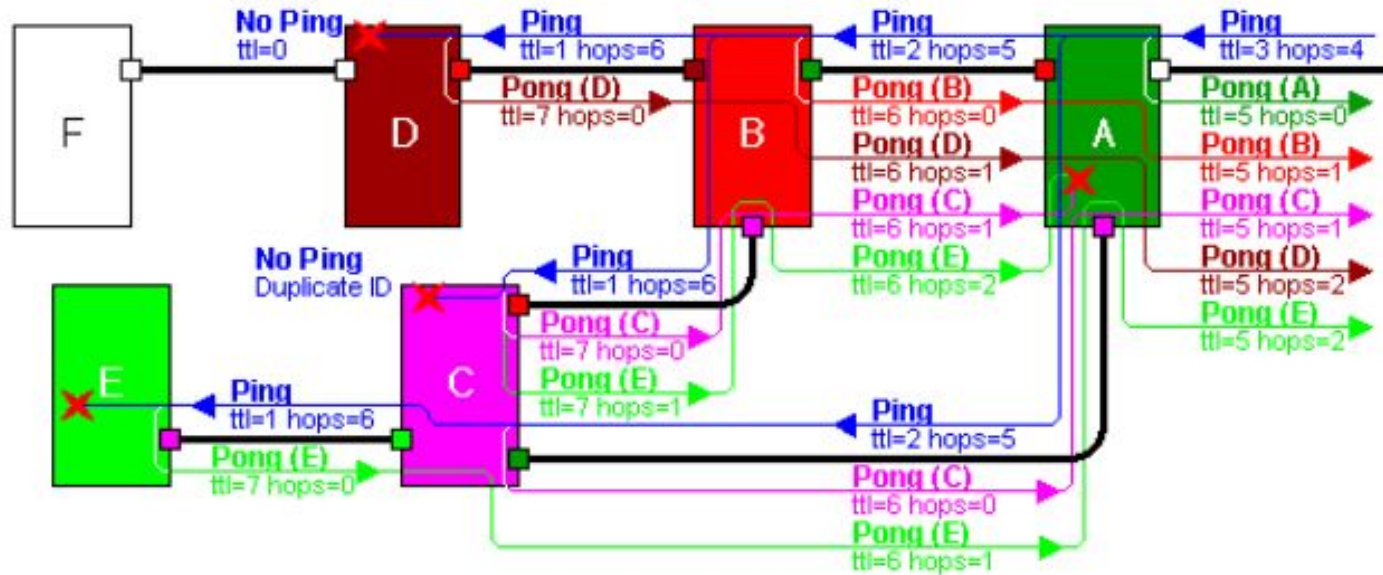
⌕ Primitivas para la comunicación

- **Ping:** Requerimiento de un host para autoanunciarse
- **Pong:** Respuesta al Ping. Contiene el IP y el p rtico del host que responde, el n mero y tama o de archivos compartidos.
- **Query:** Un requerimiento de b squeda. Contiene un string de b squeda y los m nimos requerimientos de velocidad del host que responde.
- **QueryHits:** respuesta a Query. Contiene IP, p rtico y velocidad, el n mero de archivos encontrados y el conjunto de  ndices resultado

Arquitecturas de las redes p2p

Modelo descentralizado o puro

⊕ Modelo descentralizado o puro

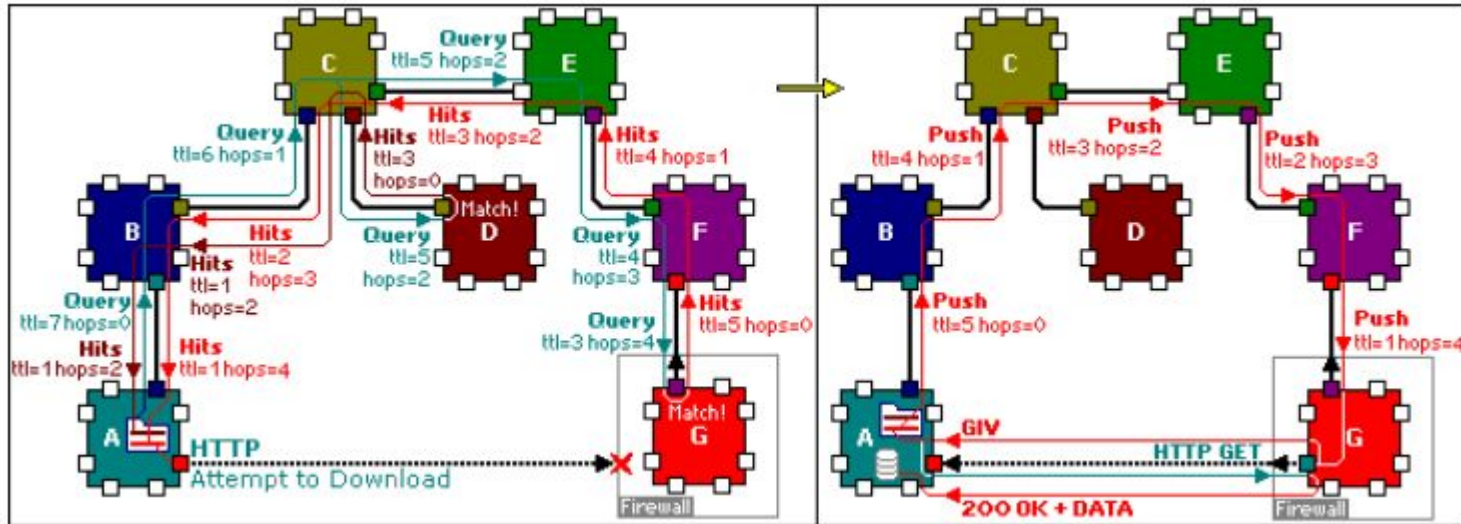


Arquitecturas de las redes p2p

Modelo descentralizado o puro

⊕ Modelo descentralizado o puro

<http://rfc-gnutella.sourceforge.net/developer/stable/>



Arquitecturas de las redes p2p

Modelo híbrido

⊕ Modelo híbrido

- Principal característica comparada con P2P puro: Introducción de otra capa dinámica jerárquica
- Un servidor central especial, que ayuda a encaminar el tráfico y a administrar los recursos de la red sin conocer la identidad de cada nodo y sin almacenar información, por lo que no comparte ningún recurso
- Reduce la carga de señalamiento sin reducir la confiabilidad
- Proceso de elección para seleccionar la asignación de un **Superpeer**
- **Leafnodes**: conectado a uno o más Superpeers (<7)



Arquitecturas de las redes p2p

Modelo híbrido

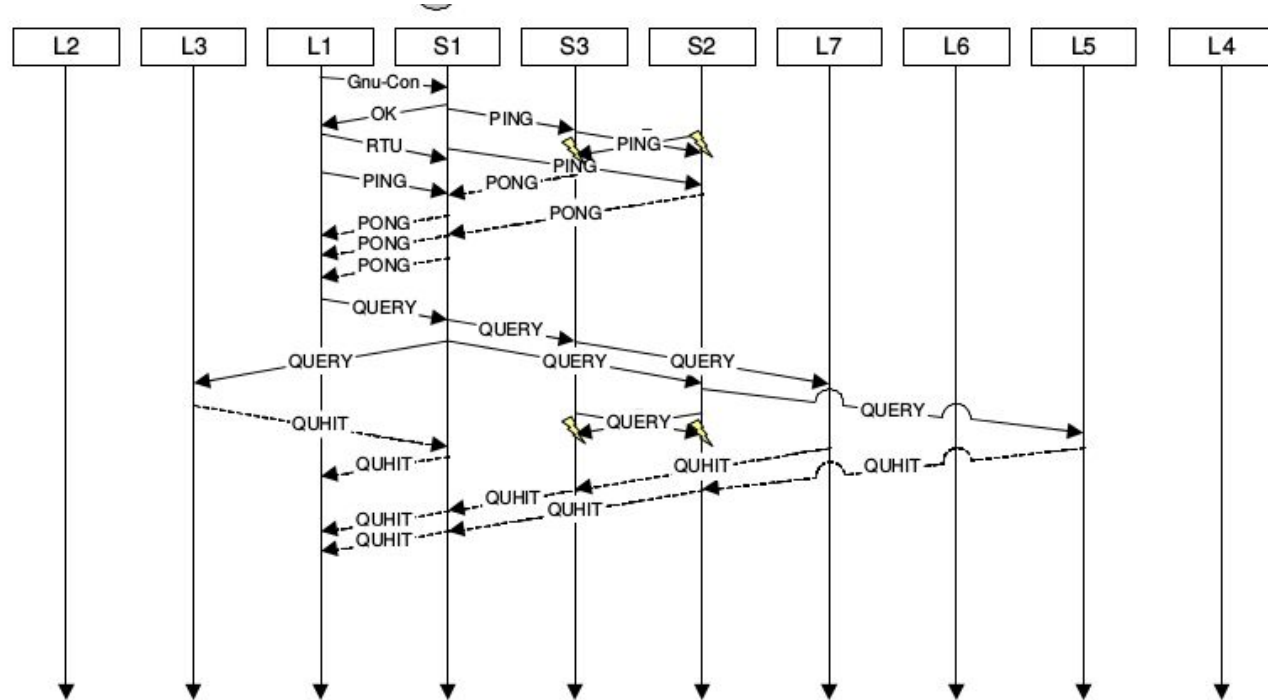
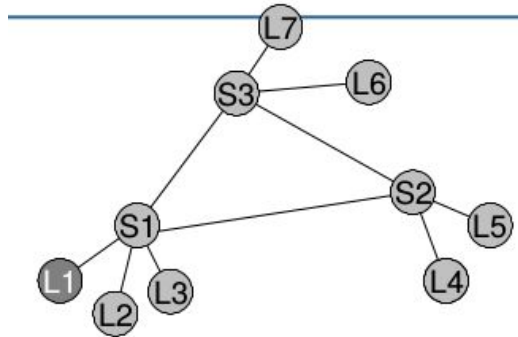
⊕ **Modelo híbrido**



Arquitecturas de las redes p2p

Modelo híbrido

⚡ Modelo híbrido



Estructuras de las redes p2p

Concepto

Estructura de las redes P2P

Las redes P2P se pueden clasificar por tener o no una estructura determinada en base a cómo se enlazan unos nodos a otros.

Estructuras de las redes p2p

No estructuradas

⊕ Estructura de las redes P2P

En las **redes P2P no estructuradas**, la localización de recursos no está determinada en nodos concretos, a cada nodo se le asignan enlaces arbitrariamente, que irá actualizando. Utilizan un mecanismo de búsqueda no determinista que no garantiza que se vaya a encontrar el recurso aunque esté en la red.

Las redes Gnutella son no estructuradas.

Estructuras de las redes p2p

Estructuradas

⊕ Estructura de las redes P2P

Las redes P2P estructuradas son aquellas en las que los recursos están situados en nodos precisos. Cada nodo cuenta con su propia **tabla de hash** y permiten que cada usuario sea responsable de una parte específica del contenido de la red.

Estas redes utilizan una **función de hash** para asignar valores a cada contenido y a cada usuario.

Ejemplos: Pastry, Chord, CAN, Tapestry, Freenet.

Estructuras de las redes p2p

Modelo de ruta en las estructuradas

⊕ Modelo de ruteo en las redes estructuradas

- Sin autoridad central de coordinación (todos los peers son iguales).
- Los archivos no están almacenados en el HW de los peers que lo proveen.
- Los archivos están almacenados en otras locaciones de la red P2P.
 - Asigna responsabilidad por un conjunto de archivos a cada peer (de acuerdo a una función definida).
 - Rutea archivos a peers asociados que los almacenan.
 - Se usa una función definida para determinar el peer asociado cuando hay un requerimiento.
 - Se baja el archivo del peer asociado.
- Usa una función hash: “Distributed Hash Table (DHT)”

Estructuras de las redes p2p

Problemas

⊕ Problemas de las redes estructuradas

- El problema reside en que el usuario debe **conocer exactamente el nombre** del archivo para poder aplicar la función de hash y obtener un resultado adecuado.
- Otro punto crítico se da debido al hecho de que cada vez que un usuario se conecta o desconecta de la red, se inicia un proceso de **sincronización** para actualizar las tablas de hash de los nodos vecinos. Esto crea una gran cantidad de mensajes de mantenimiento que en un caso extremo podrían dar lugar al colapso de la red.

Estructuras de las redes p2p

Chord

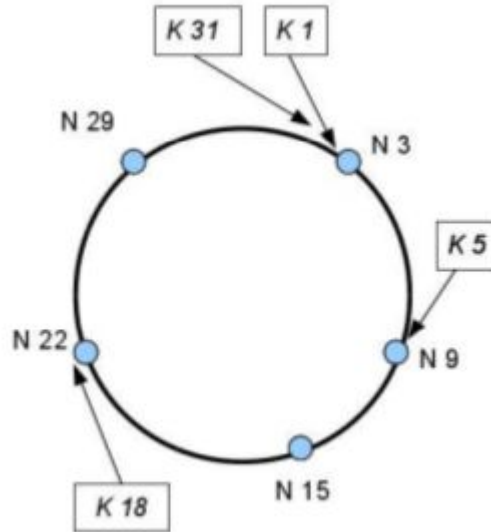
Chord

La topología de una red Chord es básicamente un anillo de 2^m elementos en el cual los nodos están ordenados según sus identificadores, en orden creciente siguiendo el sentido de las agujas del reloj. Cada *key* K se asigna **al primer** nodo cuyo identificador sea igual o superior a K en el anillo.

Estructuras de las redes p2p

Chord

⊕ Chord



Estructuras de las redes p2p

Operaciones

⊕ Mapeo de nodos y datos en el mismo espacio de direcciones

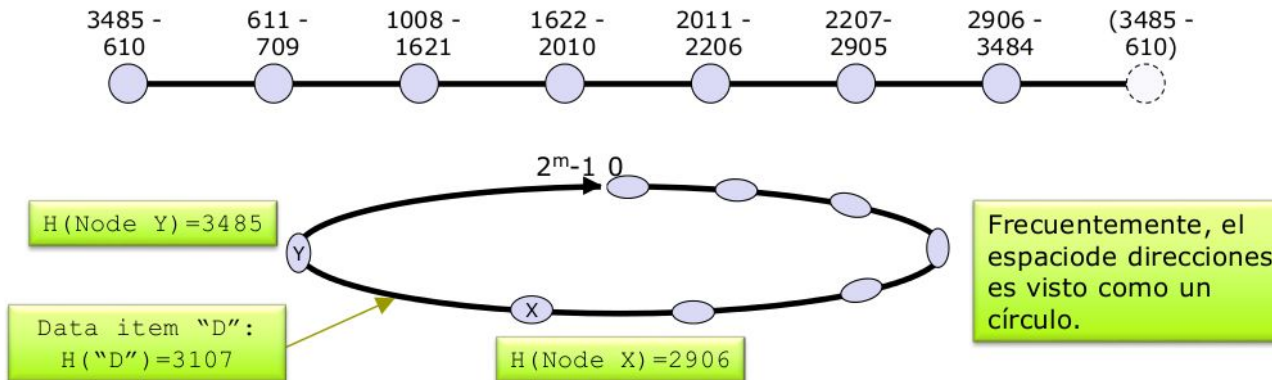
- Peers y contenidos son direccionados usando identificadores flat (IDs)
- Espacio común de direcciones para datos y nodos
- Los nodos son responsables de los datos **en ciertas partes** del espacio de direcciones
- La asociación de datos a nodos puede cambiar dado que los nodos pueden desaparecer

Estructuras de las redes p2p

Operaciones



- Paso 1: Mapeo de contenido/nodo en espacio lineal
 - Usualmente: $0, \dots, 2^m - 1 \gg$ número de objetos a ser almacenados
 - Mapeo de datos y nodos en el espacio de direcciones (con hash)
 - ▶ Ej., $\text{Hash}(\text{String}) \bmod 2^m$: $H(\text{"my data"}) \rightarrow 2313$
 - Asociación de partes de espacio de direcciones a nodos HT



Estructuras de las redes p2p

Operaciones

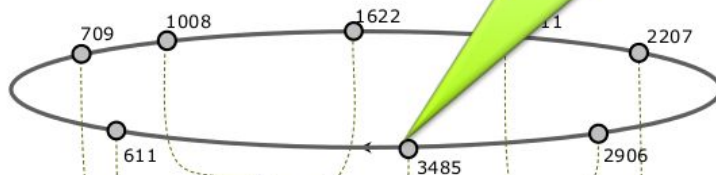


Cada nodo es responsable por una parte del rango de valores

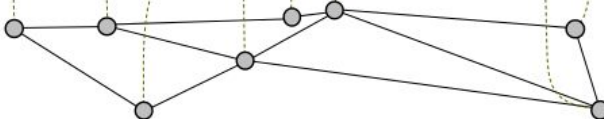
- Frecuentemente con redundancia (solapado de partes)
- Adaptación continua
- Las topologías real (underlay) y lógica (overlay) no correlacionan

Nodo 3485 es responsable por items de datos en rango 2907 to 3485

Vista lógica de DHT



Mapeo a la topología real



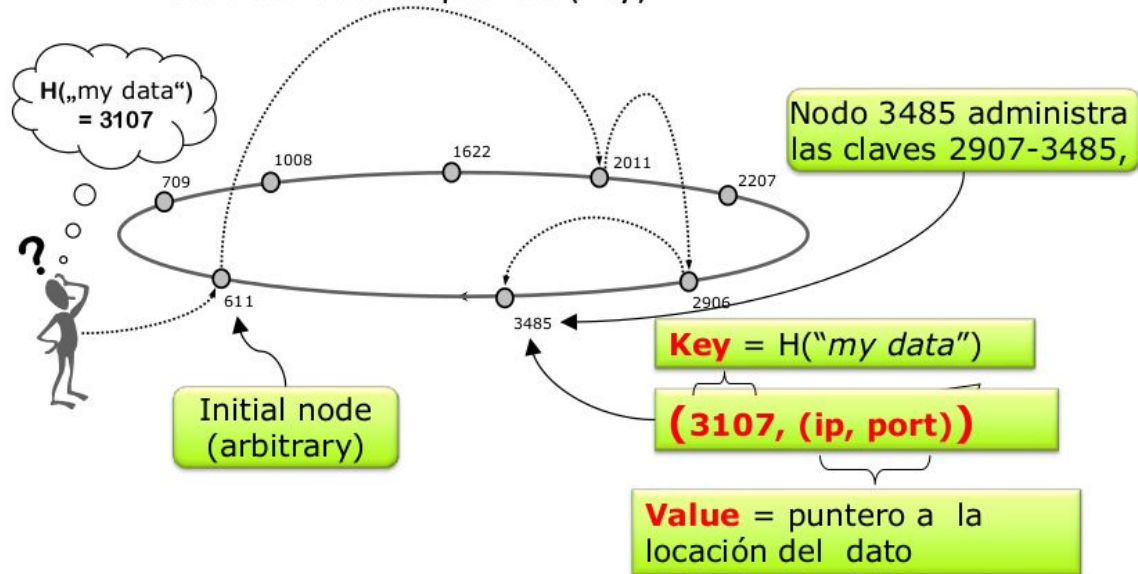
Estructuras de las redes p2p

Operaciones



■ Ruteo de un par K/V

- Comienza la búsqueda en un nodo arbitrario de DHT
- Ruteo del dato requerido (key)



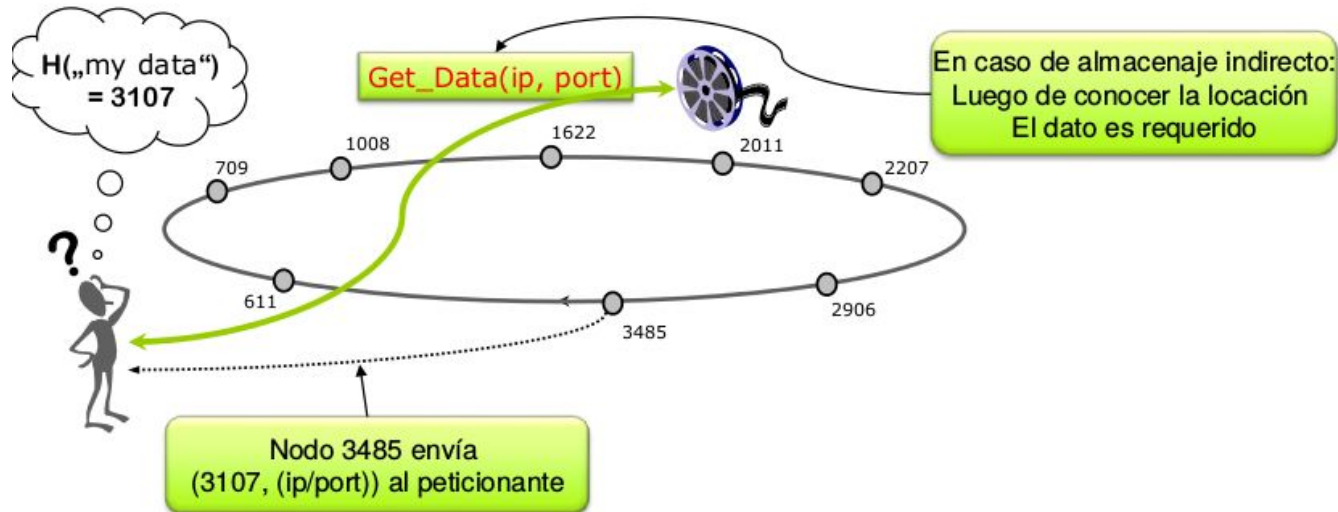
Estructuras de las redes p2p

Operaciones

■ Obteniendo el contenido



- El par K/V es atendido por el peticionante
- El peticionante analiza la K/V-tupla (y baja el dato de la actual locación – en caso de almacenaje indirecto)



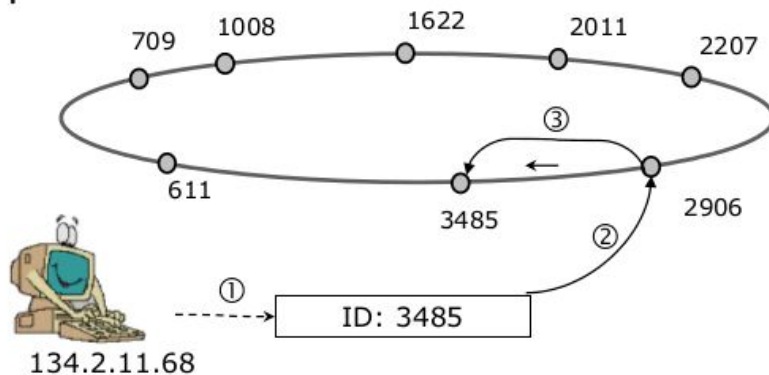
Estructuras de las redes p2p

Operaciones



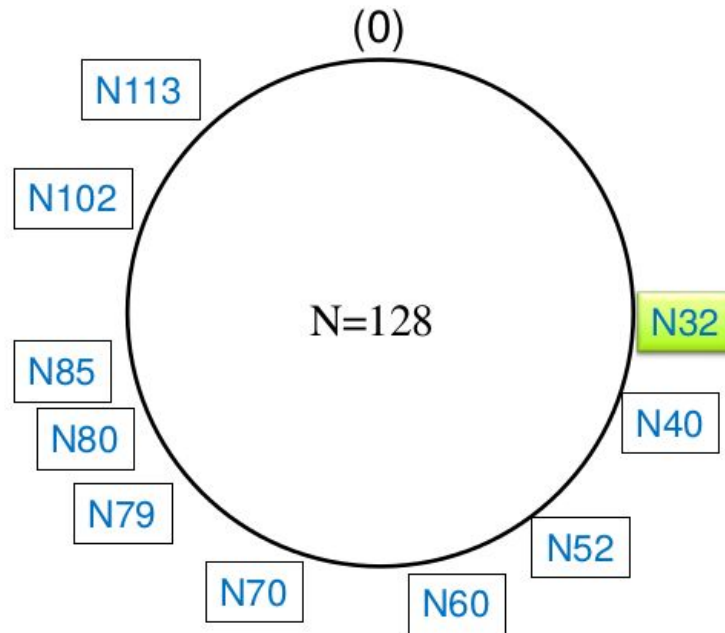
Agregando un nuevo nodo

1. Cálculo del ID del nodo
2. El nuevo nodo contacta DHT vía un nodo arbitrario
3. Le asigna un particular rango hash
4. Copia de los pares K/V del rango hash (usualmente con redundancia)
5. Mapeo en el ambiente de ruteo



Estructuras de las redes p2p

Operaciones



N32's Tabla Finger

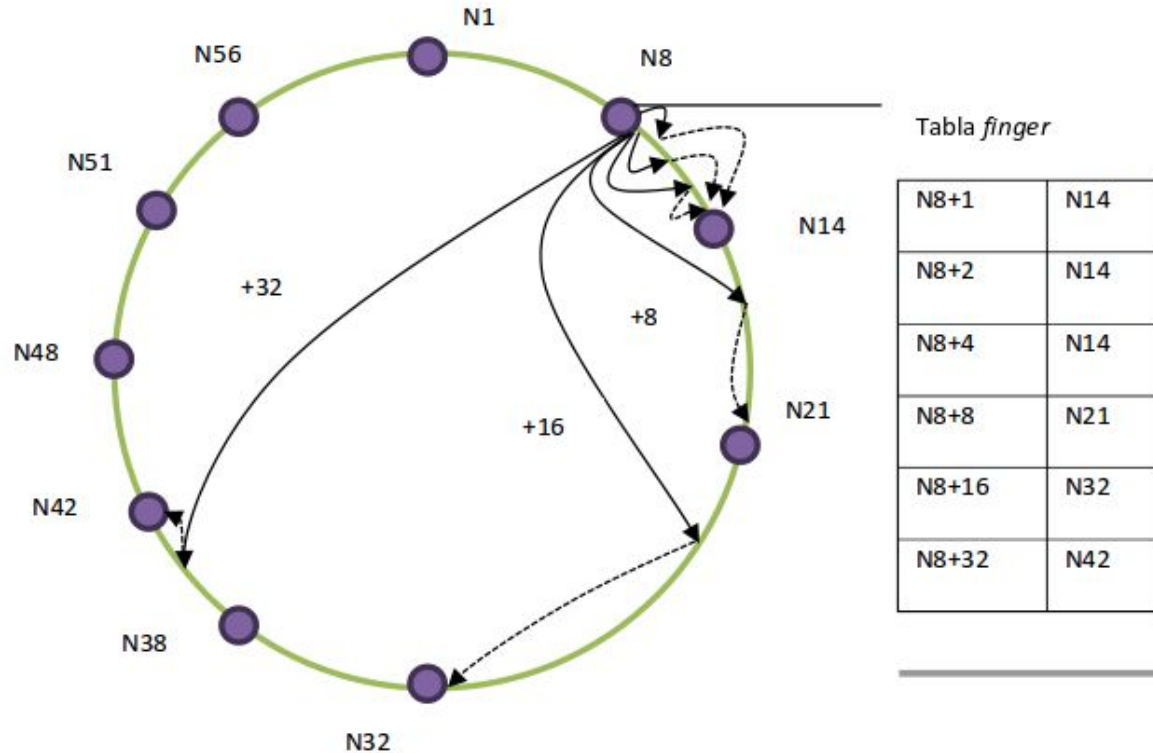
33..33	N40
34..35	N40
36..39	N40
40..47	N40
48..63	N52
64..95	N70
96..31	N102

Tabla Finger contiene actualmente
ID y dirección IP

Nodo n 's i -ésima entrada: **primer** nodo $\geq n + 2^{i-1}$

Estructuras de las redes p2p

Operaciones



Estructuras de las redes p2p

Operaciones



$\text{sucesor}(3+2^{1-1}) = \text{sucesor}(4) = \text{N5}$

$\text{sucesor}(3+2^{2-1}) = \text{sucesor}(5) = \text{N6}$

$\text{sucesor}(3+2^{3-1}) = \text{sucesor}(7) = \text{N9}$

$\text{sucesor}(3+2^{4-1}) = \text{sucesor}(11) = \text{N15}$

$\text{sucesor}(3+2^{5-1}) = \text{sucesor}(19) = \text{N22}$

Tabla 1: Cálculo de finger table de N3

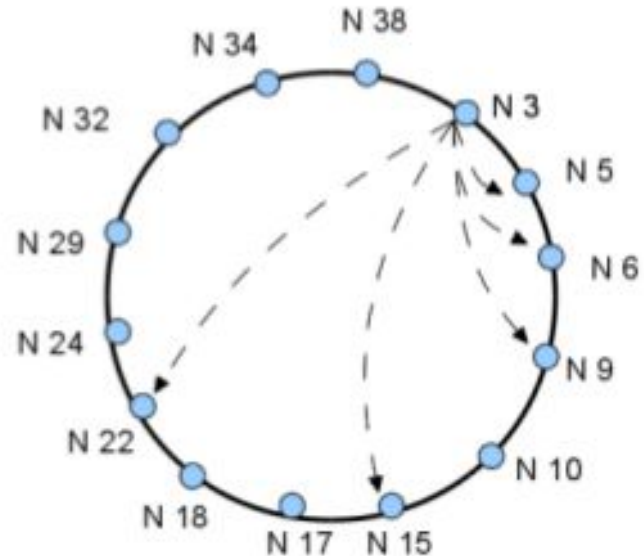
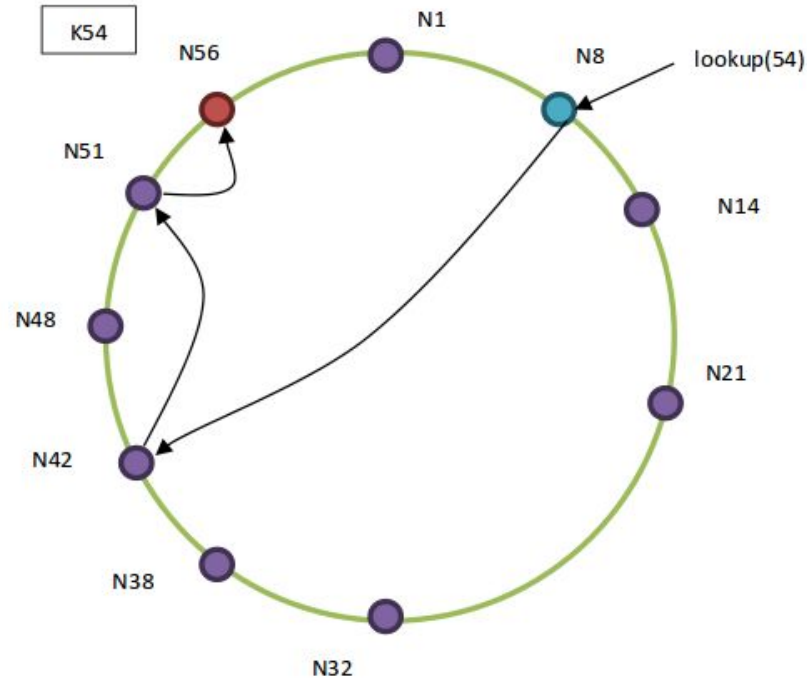


Fig. 7: Finger table de N3

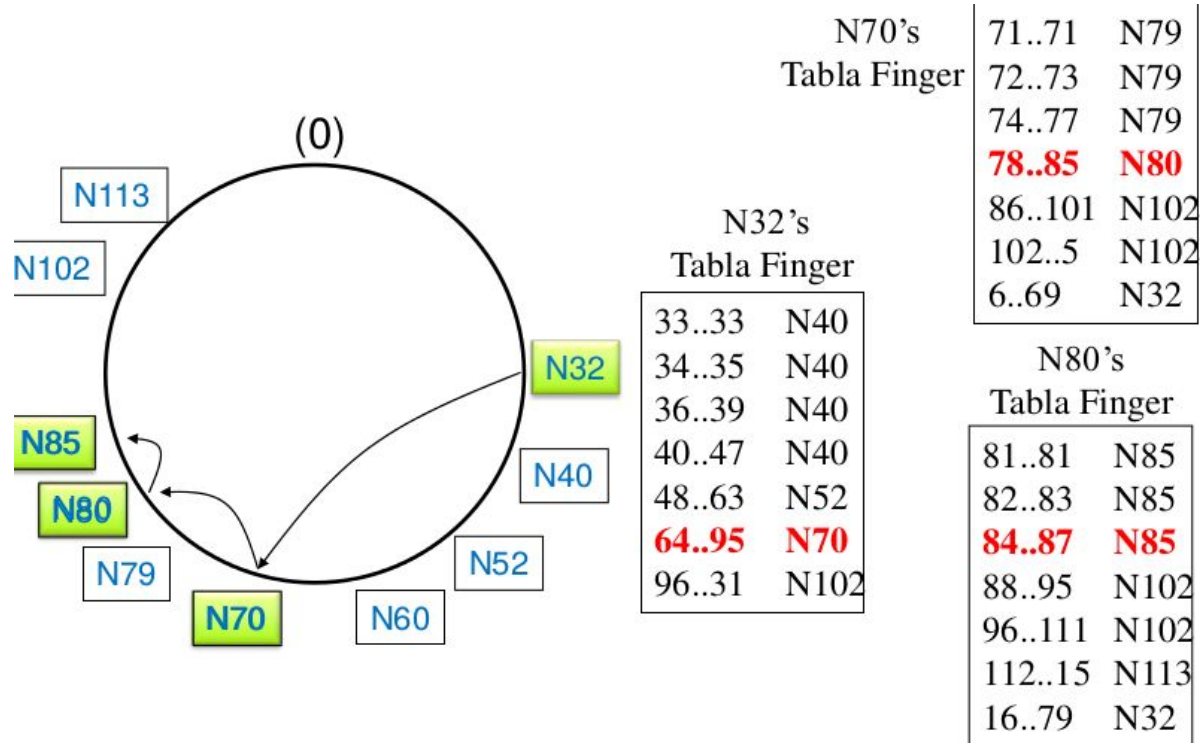
Estructuras de las redes p2p

Operaciones



Estructuras de las redes p2p

Operaciones



Resumen

Operaciones

Cliente-Servidor	Peer-to-Peer			
<ol style="list-style-type: none"> 1. El Servidor es la entidad central y el único proveedor del servicio y contenido. → La red es manejada por el servidor 2. El Servidor es el sistema de más alto rendimiento. 3. El Cliente es el sistema d más bajo rendimiento <p>Ejemplo: WWW</p>	<ol style="list-style-type: none"> 1. Los recursos son compartidos entre peers 2. Los recursos pueden ser accedidos directamente desde otros peers 3. Concepto de Servent 			
	P2P sin Estructura			
	P2P Centralizado	P2P Puro	P2P Híbrido	P2P Estructurado
	<ol style="list-style-type: none"> 1. Todas las características de P2P incluidas 2. Una entidad central es necesaria para proveer el servicio 3. La entidad central es una especie de Db índice/grupo <p>Ejemplo: Napster</p>	<ol style="list-style-type: none"> 1. Todas las características de P2P incluidas 2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad 3. → Entidades no central es <p>Ejemplos: Gnutella 0.4, Freenet</p>	<ol style="list-style-type: none"> 1. Todas las características de P2P incluidas 2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad 3. → Enidades centrales dinámicas <p>Ejemplo: Gnutella 0.6, JXTA</p>	<ol style="list-style-type: none"> 1. Todas las características de P2P incluidas 2. Alguna entidad terminal puede ser removida sin pérdida de funcionalidad 3. → Entidades no central es 4. Las conexiones en overlay son "fijas" <p>Ejemplos: Chord, CAN</p>
