

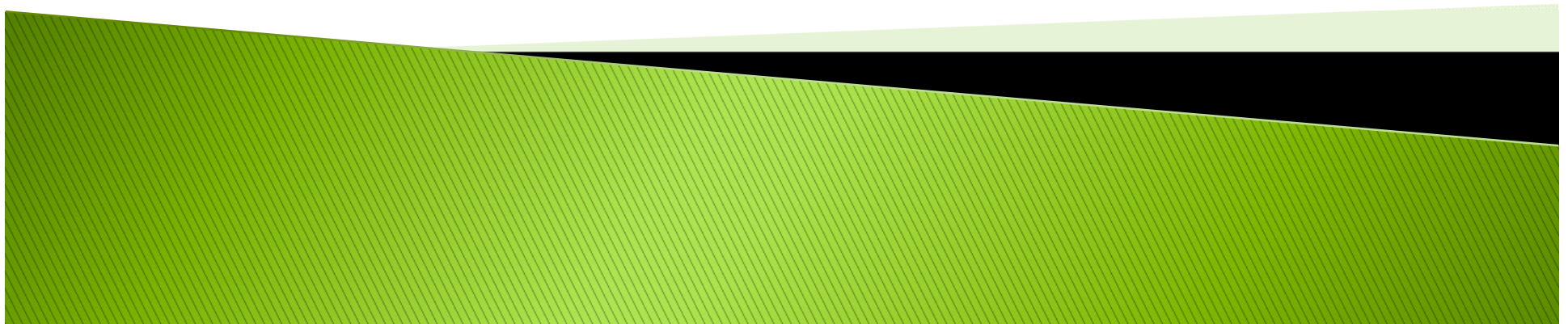


# **5. Redes de distribución de contenidos**

Servicios Telemáticos Avanzados

4º Grado en Ingeniería en Tecnologías de Telecomunicación

Especialidad de Telemática



## Indice

### *Hora 1*

1. Problemática de sistemas de información un servidor-un sitio
2. Mirrors
3. Proxy caches
4. Redes de distribución de contenidos (CDN)
  - 4.1 CDN localización
  - 4.2 CDN redirección

### *Hora 2*

- 4.3 CDN replicación
- 4.4 Estructura CDN
- 4.5 ESI
- 4.6 Proveedores CDN
  - 4.6.1 Akamai
- 4.7 Efecto de la versión HTTP en CDNs

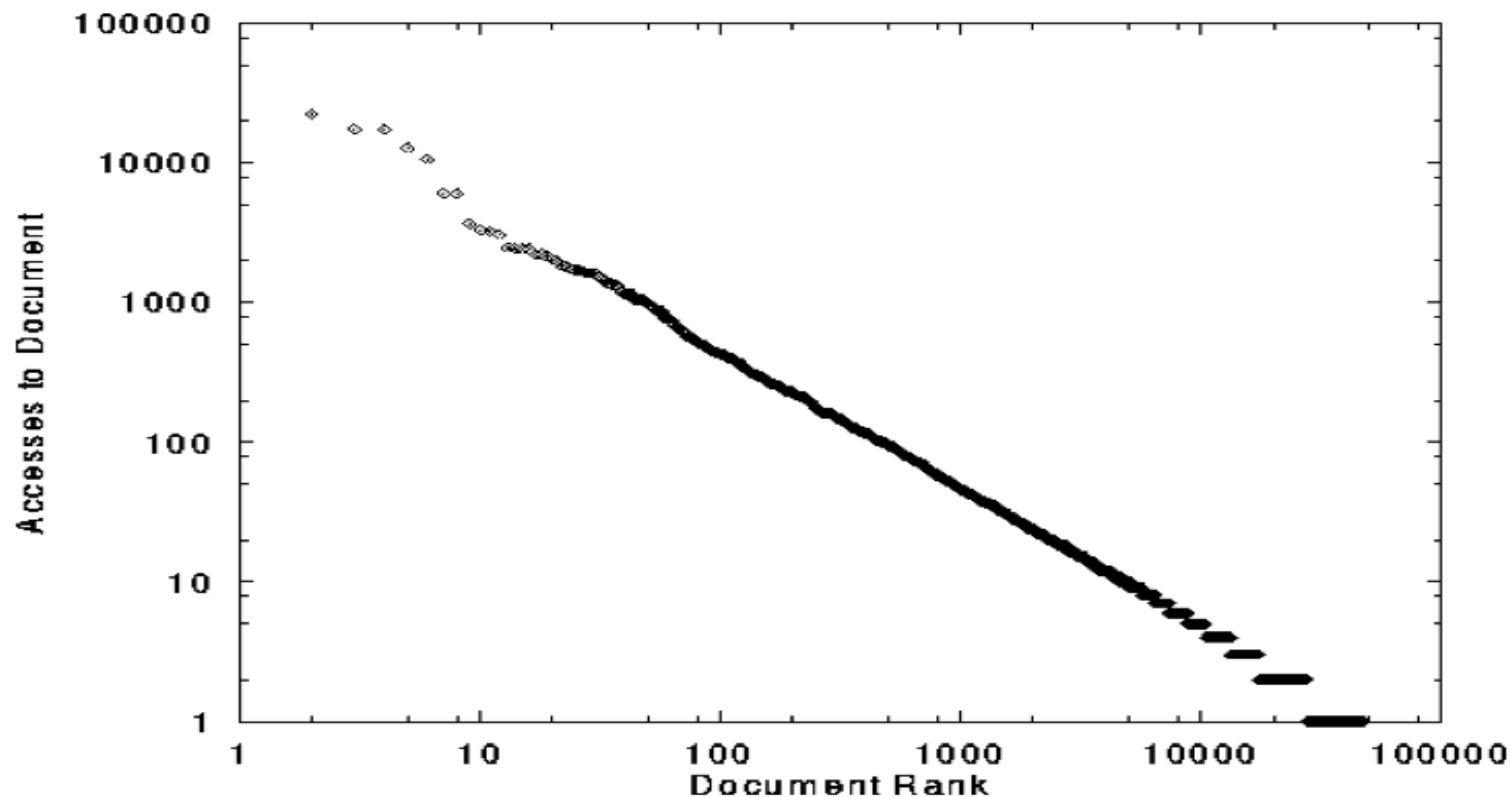
### Referencias

# 1. Problemática de sistemas de información un servidor-un sitio

- ▶ Único servidor
  - Un único punto de fallo
    - Hardware/software
    - Ataques de denegación de servicio
  - Fácilmente sobrecargado
    - Contenido popular, accesos masivos: “fash crowd” o “slashdot effect”
- ▶ Único sitio
  - Lejos de la mayoría de clientes
  - Conectividad limitada a uno o pocos proveedores o ISPs
    - Que incluso pueden compartir cableado

# Problemática de sistemas de información un servidor-un sitio

- Variabilidad en la popularidad de los sitios web
  - Distribución Zipf o de ley de potencias
  - Ejemplo: Characteristics of WWW Client-based Traces. Carlos R. Cunha, Azer Bestavros, Mark E. Crovella, BU-CS-95-01

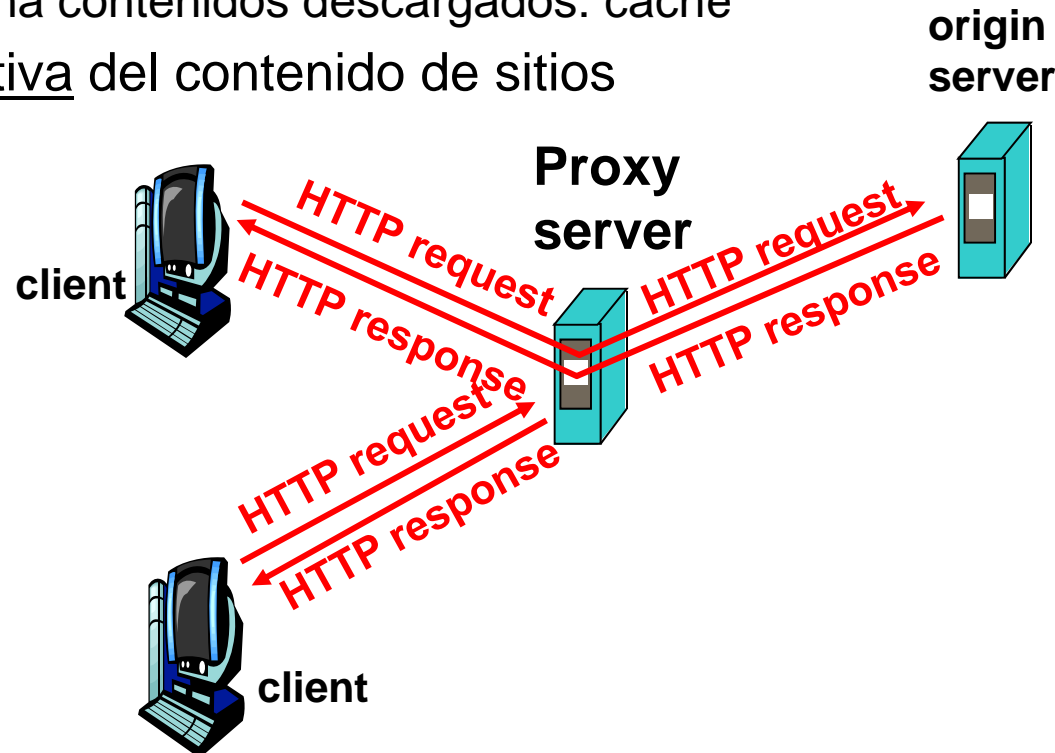


## 2. Mirrors

- ▶ Copias de servidores (*mirrors*) instaladas en áreas de la red geográficamente alejadas.
  - Sincronización por FTP o *rsync*
- ▶ Redirección a veces automática generada por el propio servidor del contenido (modificación o listado de diferentes URLs destino para el mismo contenido):
  - *cookies*
  - *preferred language* del cliente
  - algunos productos comerciales hacen balanceo de carga (según métricas de la red)
- ▶ Problemática
  - Soluciones muy rígidas
  - Únicamente contenido estático
  - Actualización normalmente diaria

### 3. Proxy caches

- ▶ Un servidor proxy web es una entidad de red a la que los clientes delegan la realización de solicitudes HTTP
  - Una conexión TCP entre cliente y proxy
  - Otra conexión TCP entre el proxy y el servidor original
  - El proxy almacena contenidos descargados: caché
- ▶ Replicación reactiva del contenido de sitios



## Forward proxy

- ▶ **Caché cercana a los clientes**
  - Normalmente localizada en la red del ISP o de la empresa de los clientes
  - Mejora eficiencia de los clientes
  - Reduce los costes de cursar tráfico de red
- ▶ **Tipos de funcionamiento**
  - Con configuración explícita
    - Requiere su configuración en el navegador del cliente
  - Con configuración implícita o transparente
    - Localizado en el camino del cliente a Internet, intercepta las conexiones de éste
    - Funcionamiento “transparente” para el usuario

## Forward proxy

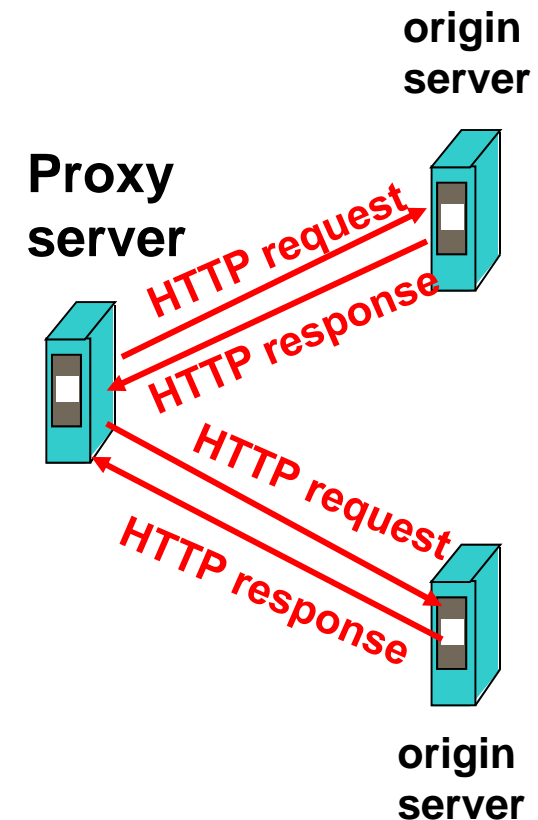
Secuencia de uso de un proxy:

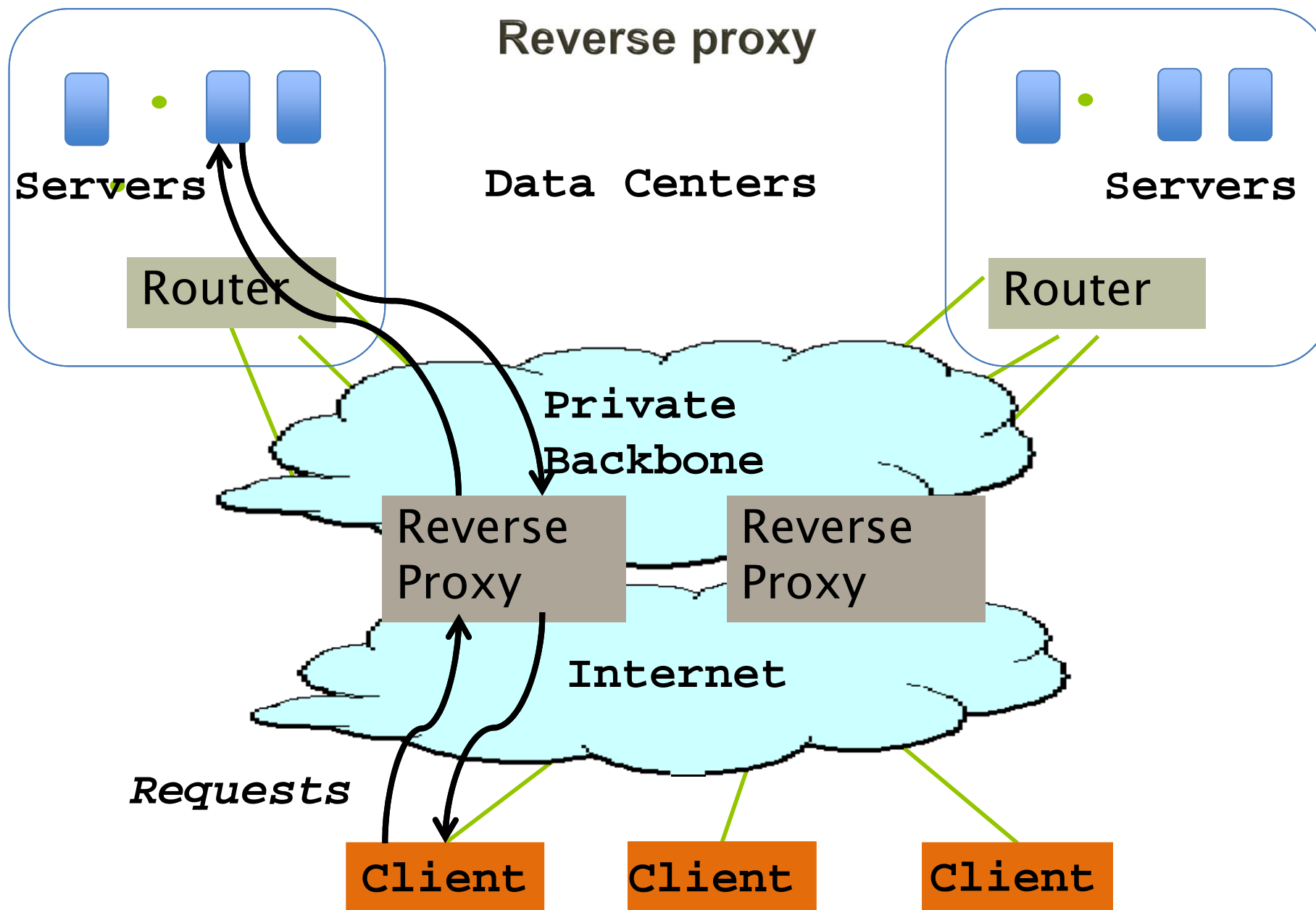
- ▶ 1. El navegador establece una conexión con el proxy y le envía la solicitud HTTP. La cabecera Host indica el servidor original.
- ▶ 2. El proxy comprueba si tiene una copia del objeto almacenada localmente. En tal caso, le envía el objeto al navegador en un mensaje HTTP de respuesta.
- ▶ 3. Si el proxy no tiene el objeto, abre una conexión con el servidor original y le envía una petición HTTP del objeto requerido por el navegador. El servidor original envía la respuesta al servidor proxy.
- ▶ 4. El proxy almacena una copia en su caché y redirige otra copia en una respuesta HTTP al navegador, por la conexión establecida en el primer paso.



## Reverse proxy

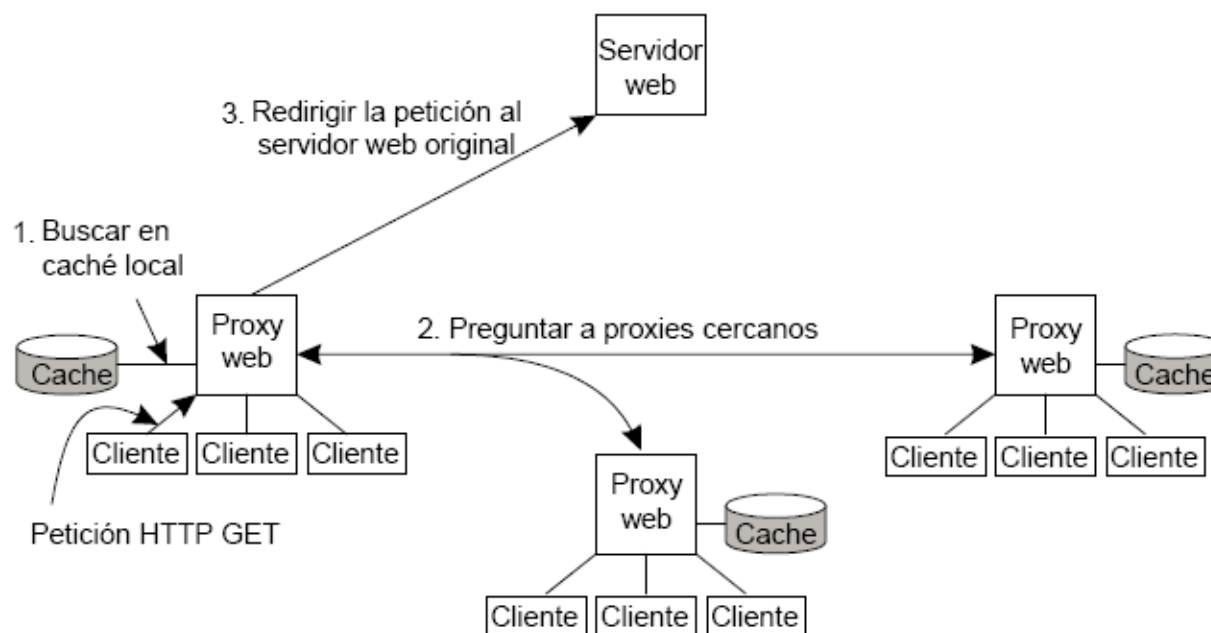
- ▶ Caché cercana a los servidores
  - Mejora la eficiencia de los servidores
  - Funcionalidades de balanceo de carga, ensamblado de contenido, transcodificación, etc.
- ▶ Desde el punto de vista del cliente, el dominio de la web original se mapea en la dirección IP del proxy
- ▶ Reciben también el nombre de publicadores





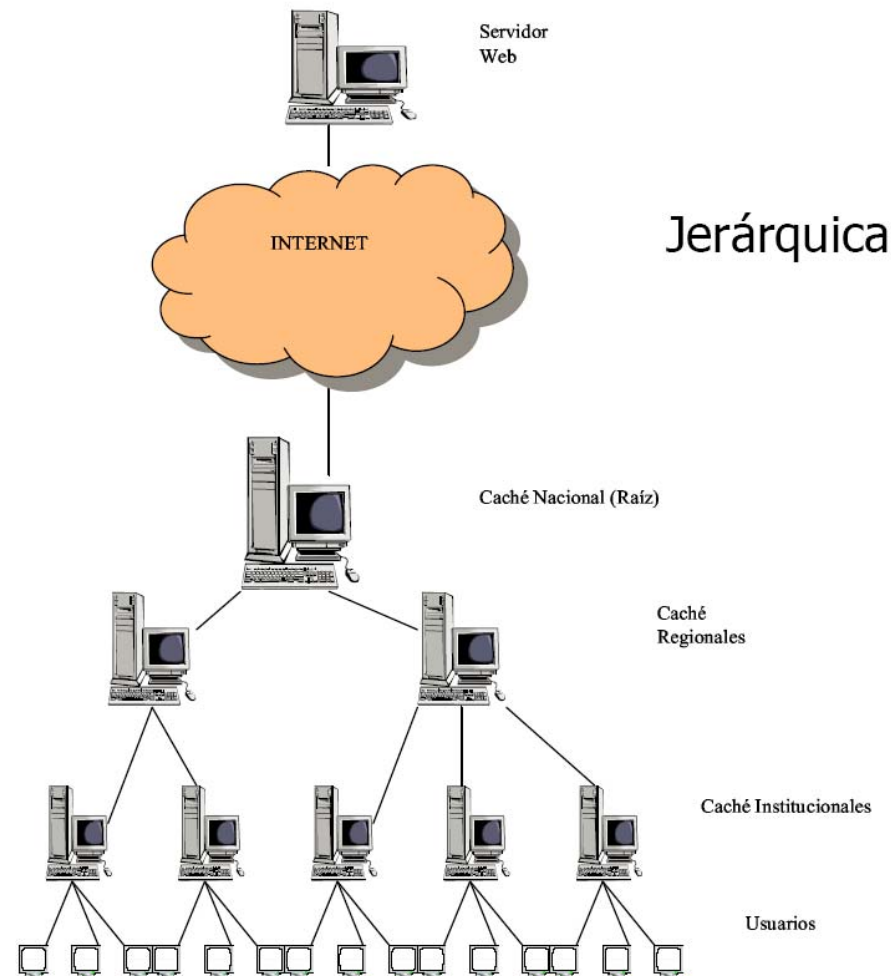
## Proxy caches cooperativos

- Se pueden establecer servidores proxy cooperativos que aumentan la probabilidad de disponer de una copia cercana



## Proxy caches, jerarquía

- ▶ **Múltiples cachés web**
  - caché cooperativa
  - topología jerárquica
  - protocolo *squid*
- ▶ **Árbol de distribución de contenidos**

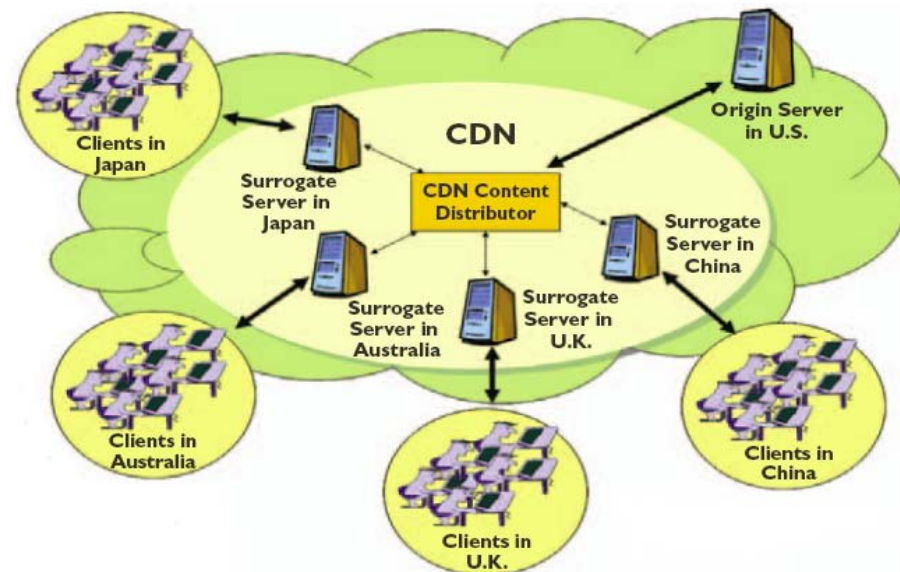


## Proxy caches

- ▶ Ventajas del uso de proxys:
  - Reducción del tiempo de respuesta
  - Reducción del tráfico web a través de la red de acceso a Internet, y por tanto reducción del coste de conexión
  - Reducción de carga en el servidor original
  - Funcionalidades extra de seguridad: antivirus
  - Funcionalidades extra de restricción de contenidos
- ▶ Limitaciones
  - Rompe el paradigma TCP extremo a extremo
  - Útil sólo para contenidos estáticos, nada de aplicaciones, cookies, etc.
  - La comunicación SSL no es cacheable
    - Reverse proxy puede encargarse de la encriptación SSL con los clientes, dejándola en claro para la comunicación con los servidores
  - Analítica: el servidor original pierde información de parte de sus accesos
  - Datos obsoletos: compromiso entre los tiempos de refresco de las caches

## 4. Redes de distribución de contenidos (CDN)

- ▶ Content Distribution Networks (CDN)
- ▶ Replicación proactiva de contenidos
- ▶ Servidores que poseen la misma información replicada que el original y atienden peticiones de usuarios por algún criterio que:
  - Equilibre la carga entre servidores.
  - Minimice la latencia en el volcado de contenidos al cliente.
- ▶ Facilitan las transferencias de ficheros grandes o aumentan la rapidez en las consulta.
  - ❑ Los contenidos se descargan en los servidores del CDN.
  - ❑ Colocar los servidores cerca de los usuarios minimiza retardo y pérdidas.
  - ❑ El servidor CDN se coloca normalmente en el borde de la red de acceso.



## Redes de distribución de contenidos

- ▶ El proveedor del servicio CDN puede ser diferente del proveedor de los contenidos (Ej: Akamai <-> CNN).
- ▶ La actualización de los contenidos es automática y simultánea en todos los servidores CDN.
- ▶ CDN crea un mapa sobre cercanía entre sus servidores e ISPs (usuarios finales).
- ▶ Sirven contenidos estáticos y dinámicos
  - Habitualmente se encargan de contenidos estáticos de forma sencilla imágenes, vídeos, publicidad, etc.
  - La dinamicidad del servidor original es difícil de replicar
    - Necesidad de adaptaciones



## 4.1 CDN localización

- ▶ Un CDN **localiza** su servidor sustituto (surrogate server) más “cercano” al cliente según criterios diversos:
  - Zona geográfica
    - Normalmente colocados en POPs (Points of Presence) de diferentes ISPs (Internet Service Providers)
  - Topología
    - Número de saltos (routers) entre cliente y servidor
  - Retardo
    - Medido de manera activa o pasiva
  - Carga de los servidores
    - En CPU o ancho de banda consumido en su acceso
- ▶ Esto supone llevar control de esta información de la forma más actualizada posible en centros de control que puedan tomar las mejores decisiones en cada momento
- ▶ Los proveedores de CDN suelen elegir como localización para sus servidores aquellos puntos con ancho de banda o electricidad más barata, o si es posible con mejores condiciones ambientales (evitar el calor que eleve los costes de refrigeración)



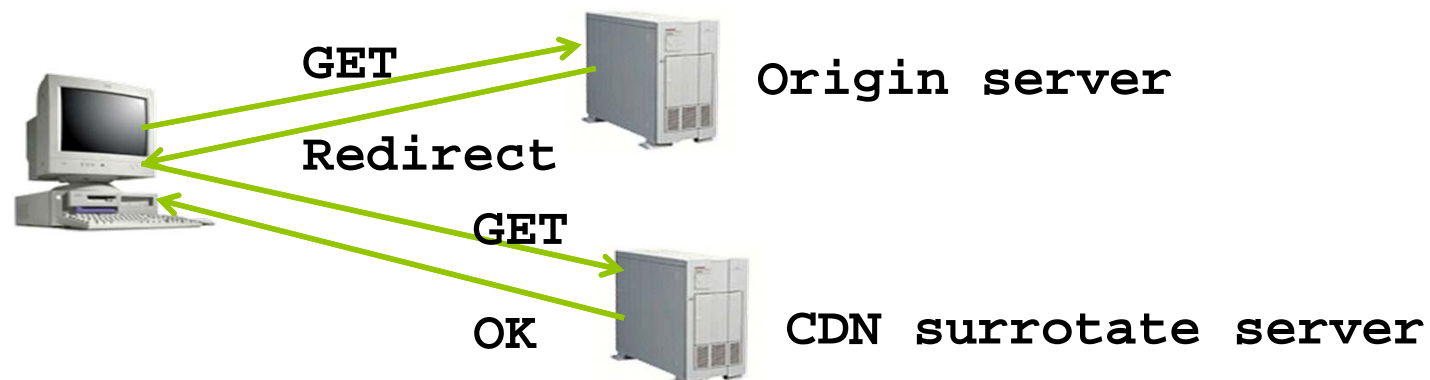
## 4.2 CDN redirección

- ▶ Un CDN **redirecciona** la petición del cliente al servidor CDN más “cercano” de manera transparente al usuario.
- ▶ Técnicas:
  - Redirección HTTP
  - Routing anycast
  - Redirección DNS
  - Reescritura de URL

## CDN redirección

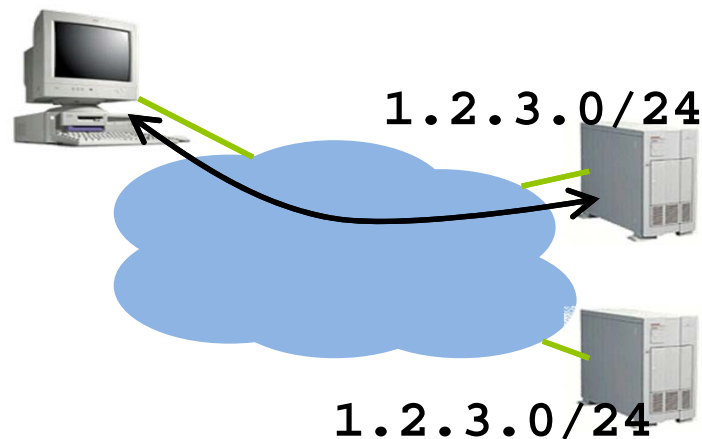
### ► 1) Redirección HTTP

- El cliente accede al servidor original del que recibe una redirección HTTP rediriéndolo a la localización real del contenido
- Ventajas
  - Control absoluto por el servidor original
- Desventajas
  - Extra RTT en el establecimiento de las conexiones
  - Sobrecarga del servidor original
  - El cliente ve que cambia la URL que carga el navegador



## CDN redirección

- ▶ 2) Anycast routing
  - Servidores con la misma IP en diferentes localizaciones, accesibles según el camino tomado desde la red origen
  - Ventaja
    - Ruta al servidor más cercano
  - Desventajas
    - No considera carga de la red o de los servidores
    - Diferentes paquetes pueden llegar a alcanzar diferentes servidores
    - Utilizado en esquemas muy simples petición-respuesta



## CDN redirección

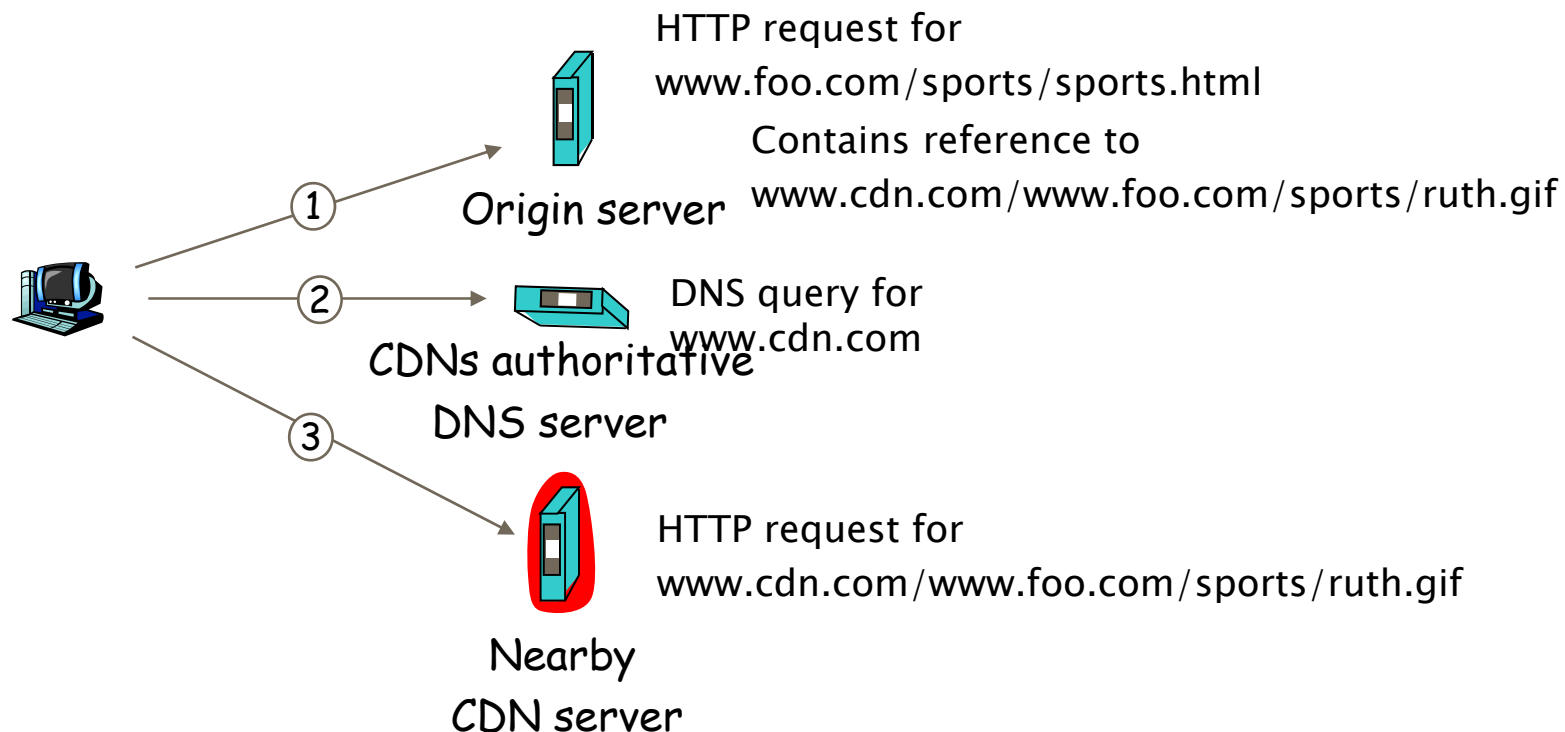
- ▶ 3) Redirección DNS (Domain Name Service)
  - Los servidores DNS autoritativos son controlados por el CDN
  - El nombre DNS del recurso se resuelve en una IP u otra del servidor CDN según interese (disponibilidad de recursos y condiciones de red)
  - Habitualmente se establecen tiempos de validez del mapeo DNS (TTL) cortos, para que pueda hacer un balanceo mejor según las condiciones cambiantes de la red y servidores (10-30sg)
  - Tipos:
    - Sitio completo: se delega por completo la resolución del nombre del servidor original en los servidores DNS del CDN. El servidor original queda oculto a los clientes.
    - Sitio parcial: el servidor original modifica las URL embebidas de objetos (principalmente imágenes) para que hagan referencia a nombres propios resolubles por los servidores DNS del CDN.

## CDN redirección

- ▶ Redirección DNS (Domain Name Service)
  - Ventajas
    - Control relativamente fino del destino de las conexiones
  - Desventajas
    - Compromiso entre cache de DNS y el control fino de la redirección

## Ejemplo CDN redirección DNS parcial

- ▶ Proveedor de contenidos ([www.foo.com](http://www.foo.com))
  - Distribuye las páginas HTML.
- ▶ Proveedor CDN ([www.cdn.com](http://www.cdn.com))
  - Distribuye las imágenes gif.
  - Transparente al usuario al utilizar redirecciones de DNS.



## CDN redirección

- ▶ 4) Reescritura de URL (Uniform Resource Locator)
  - El servidor original reescribe dinámicamente las URLs embebidas en sus páginas según criterios (disponibilidad de recursos y condiciones de red) para redireccionar el cliente a diferentes servidores CDN.
  - A la hora de servir la página, el servidor original reescribe la URL colocando la dirección IP concreta del servidor CDN al que se quiere direccionar, evitando la necesidad de búsqueda de DNS.

