

Content Delivery Networks (Enero 2012)

Iván López Espejo

Este texto trata de la presentación y descripción de los sistemas conocidos con el nombre de *content delivery networks* (CDNs) cuya misión es la de replicar, usualmente, datos multimedia a lo largo de diferentes nodos dentro de una red con la finalidad de reducir la latencia que el usuario experimenta en la recuperación de contenidos.

I. INTRODUCCIÓN

UNA RED de distribución de contenidos o, del inglés, *content delivery network* o *content distribution network* (CDN), la cual comienza a surgir a partir de 1998, puede ser descrita como un sistema de computadores tal que contiene réplicas de datos (redundancias), usualmente multimedia, a lo largo de diversos servidores en una red con el fin de mejorar el acceso a dichos datos a través del aumento del ancho de banda total de acceso y del rendimiento mientras se reduce la latencia. Como se ha dicho, el tipo de datos usualmente accesible a través de este tipo de sistemas es multimedia, como son los objetos descargables: vídeo, audio, imágenes, software o documentos. Sin embargo cubre un amplio espectro que además incluye objetos web como texto, gráficos, URLs y scripts, así como *streaming* y otros. Normalmente las fuentes de contenido son las grandes empresas como, por ejemplo, pueden ser Google o Amazon, pero también puede hacer uso cualquier usuario de un servicio de CDN para sus datos o sitios web tal y como ejemplificaremos en el punto tercero. Otras fuentes de contenido son los proveedores de servicios, las medianas empresas o los broadcasters de noticias. La figura 1 muestra un ejemplo de cómo los clientes interactúan con la CDN especificando la petición de contenido o servicio a través del dispositivo correspondiente.

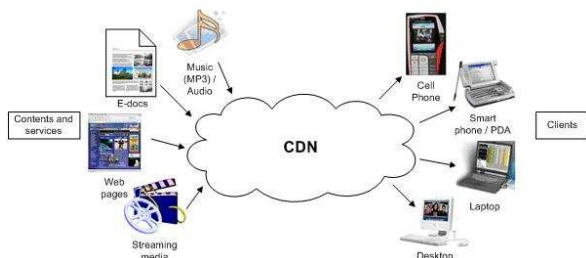


Fig. 1. Servicios y contenidos usualmente provistos a través de una CDN.

Con frecuencia, existen determinadas horas punta de acceso a sitios web y aplicaciones produciéndose una gran afluencia de tráfico por este u otros motivos y, a menos que exista un sistema CDN desplegado, usualmente dichos sitios web y aplicaciones pueden quedar fuera de servicio. Tal y como se ha introducido, el hecho de tener replicada la información en diferentes servidores estratégicamente ubicados geográficamente puede incrementar la capacidad total de acceso respecto de la que proporcionan las principales conexiones troncales, permitiendo un incremento en el número concurrente de usuarios que quieren recuperar un mismo

contenido al acceder estos al servidor adecuado en función, normalmente, de su proximidad geográfica. Por ejemplo, cuando hay una red troncal a 10Gbps y el servidor central tiene una capacidad de 200Gbps, sólo pueden enviarse datos a una tasa máxima de 10Gbps. Sin embargo, si 10 servidores se resitúan en sendas localizaciones frontera apropiadamente, la capacidad total del sistema frente a peticiones puede llegar a ser de $10 \times 10\text{Gbps}$ (100Gbps), pudiendo mejorar el rendimiento y atender a una mayor cantidad de usuarios a lo largo del mundo. Así, en esencia, cuando un contenido es solicitado, la red de distribución de contenido automáticamente direcciona al usuario final al servidor, típicamente, más cercano en términos geográficos mejorando la tasa de transferencia y protegiendo al sistema de picos de tráfico sobre cualquier servidor del mismo.

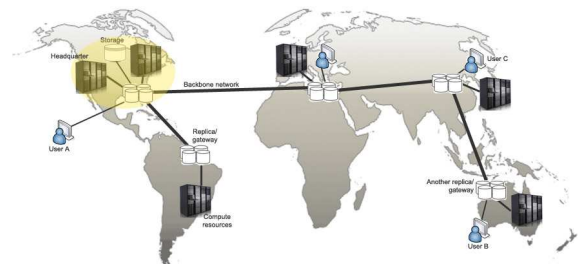


Fig. 2. Ejemplo de CDN.

Generalmente, el contenido de una CDN es distribuido sobre conexiones TCP o UDP. El throughput TCP en una red queda afectado por la pérdida de paquetes y la latencia. Con el fin de reducir la magnitud de estos dos parámetros, las CDNs tradicionalmente sitúan los servidores cerca de las redes frontera donde los usuarios de los contenidos muy posiblemente se sitúen. Al estar cercanos los datos, estos son recibidos con una menor latencia por los usuarios finales a la par que experimentan un menor jitter (variabilidad en la latencia). El incremento de la fiabilidad permite al operador de CDN distribuir contenido en HD (High Definition) con QoS (Quality of Service), bajos costes y baja carga en la red. También mejora esta tecnología la disponibilidad del servicio, pues al disponerse de diversas réplicas de la información es posible aplicar balanceo de carga en el caso de que uno o varios servidores estén caídos. El balanceo de carga también es responsable directo del incremento total de la capacidad y de la mejora de la escalabilidad.

II. TECNOLOGÍA Y PRINCIPIOS DE FUNCIONAMIENTO

Conozcamos a continuación el flujo básico de funcionamiento de una CDN. Consideremos para ello el diagrama de flujo de la figura 3, el cual provee una visión de alto nivel de la interacción básica entre los distintos componentes en un entorno de CDN. Supongamos que deseamos visitar la web *discovery.com* y que Akamai es la empresa de hosting que almacena el contenido de dicha web y además proporciona servicio de distribución de contenidos. El flujo de interacción se describe a continuación.

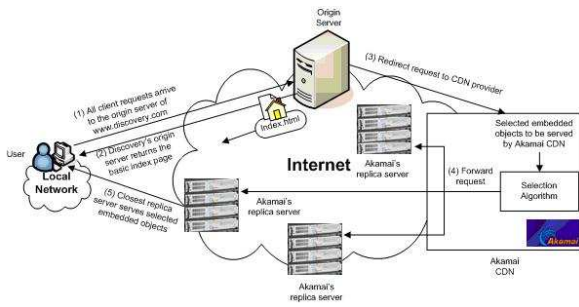


Fig. 3. Flujo básico de interacción en una CDN.

- 1) El cliente envía una petición HTTP a *discovery.com* mediante la especificación de su URL en el navegador web. Dicha petición se dirige al servidor origen de *discovery.com*.
- 2) Cuando *discovery.com* recibe la petición, su servidor web origen toma la decisión de tan sólo proveer los contenidos básicos (por ejemplo, el *index.html*).
- 3) El servidor origen redirecciona ahora la petición del cliente al proveedor del servicio de CDN (Akamai en el presente ejemplo).
- 4) Usando Akamai un algoritmo propietario de selección (usualmente basado en alguna métrica de tipo estática, estadística o dinámica), dicho proveedor de CDN selecciona el servidor réplica más cercano al cliente geográficamente hablando o más apropiado con el fin de que este último proporcione el resto de contenidos de la web (por ejemplo, objetos, barra de navegación, vídeo, imágenes, etc).
- 5) El servidor réplica seleccionado consigue el resto de contenidos a través del servidor origen, sirve la petición del cliente y cachea dicho contenido para futuras peticiones de servicio.

Como se ha esbozado, las peticiones de contenido típicamente se dirigen, mediante el uso de algún algoritmo, a los nodos réplica que son óptimos en base a algún criterio. Métricas típicas optimizables son el número de saltos, el tiempo de transmisión o la disponibilidad del servidor (tanto actual como histórica). En un escenario óptimo, usualmente este se traduce en la selección del servidor más cercano al usuario final en términos geográficos.

III. EJEMPLO PRÁCTICO: SYSLIPE

A continuación vamos a observar un ejemplo práctico de aplicación de la tecnología CDN a la mejora del rendimiento

de un sitio web. El tiempo de carga de un sitio web puede determinar en buena parte el éxito del mismo. Si el visitante del website tiene que esperar un tiempo excesivo para que se termine de completar la carga, muy probablemente se marche del sitio antes incluso de ver el contenido.

De otra parte, es cierto que los motores de búsqueda difícilmente liberan partes de su algoritmo o de cómo trabajan. Sin embargo, Google ha revelado que tiene en cuenta el tiempo de carga de un sitio web a la hora de clasificar este en los resultados de búsqueda. Esto significa que los sitios web que más rápido carguen tendrán más posibilidades de aparecer en el motor de búsqueda y en primeras posiciones.

Por estos motivos, resulta importante aprender cómo mejorar la velocidad de acceso a su web, siendo un método para ello el empleo de una CDN. Si mi compañía de hosting web se localiza en USA, sin el empleo de una CDN todo el contenido de mi sitio provendrá de servidores geográficamente localizados en USA. Si las conexiones de los visitantes a mi sitio se llevan a cabo desde otras partes del mundo, el contenido de mi web será distribuido con una mayor latencia respecto de si sus ISPs (Internet Service Providers) y ellos mismos se localizasen en USA.



Fig. 4. Ejemplo de sistema CDN.

En función de nuestro conocimiento del concepto de CDN, resulta intuitivo que podemos aprovechar esta tecnología para mejorar el rendimiento de nuestra web a través de, por ejemplo, la reducción del tiempo de carga de los contenidos a lo largo del mundo mediante la replicación de los datos estáticos del website como, por ejemplo, imágenes, hojas de estilo o ficheros javascript.



Fig. 5. Estadística de rendimiento de una web antes de emplear un servicio de CDN.

Syslipe (<http://www.syslipe.com>) es un ejemplo de empresa online que permite contratar servicios de CDN. Los resultados siguientes se refieren a la mejora del rendimiento de un sitio web (<http://www.paulund.co.uk>) experimentado por un usuario real. Antes de emplear el servicio de CDN de Syslipe, hizo uso del servicio web que ofrece Pingdom (<http://fpf.pingdom.com>) con el fin de conocer, en esencia, el número de peticiones de

acceso y el tiempo medio de carga de su website, mostrándose los resultados a priori en la figura 5. Una vez contratados los servicios de distribución de contenidos de Syslipe, repitió la experiencia obteniendo los resultados de la figura 6.



Fig. 6. Estadística de rendimiento de una web después de emplear un servicio de CDN.

Después de ello, el tiempo medio de carga del website pasó de ser de 5.8 segundos a 1.51 segundos, lo que manifiesta una mejora considerable. La diferencia fundamental radica en la localización del servidor del que se recuperaron los datos. Tanto el propietario de la web como su ISP se sitúan en UK, y ya que las pruebas las hizo desde dicha localización, es lógico que el contenido debiese de ser recuperado a través de un servidor cercano en términos geográficos. Los servidores de hosting de su web se sitúan en USA, por lo que únicamente se podía recuperar en un principio el contenido desde allí antes del empleo del servicio de CDN. Ahora, sin embargo, lo puede hacer también desde un servidor situado geográficamente en Ámsterdam con la consiguiente reducción de la latencia experimentada.

En conclusión, se ha comprobado cómo de beneficioso puede ser para un sitio web el empleo de una CDN. Teniendo un website más rápido podemos reducir la tasa de salidas de visitantes antes de la carga del mismo y mejorar la posición de nuestro sitio en los resultados ofrecidos por los motores de búsqueda.

REFERENCIAS

- [1] Content Delivery Network, CDN Price Quotes, <http://contentdeliverynetworkcdn.com/>.
- [2] Syslipe Content Delivery Network Review, <http://www.paulund.co.uk/content-delivery-network-review>.
- [3] Using Content Delivery Networks (CDN) to Speed Up Content Load on the Web, <http://tech.amikelive.com/node-285/using-content-delivery-networks-cdn-to-speed-up-content-load-on-the-web/>.
- [4] Content Delivery Network (CDN) Research Directory, <http://ww2.cs.mu.oz.au/~apathan/CDNs.html>.
- [5] Content Delivery Network, http://en.wikipedia.org/wiki/Content_delivery_network.