

Universidad de Granada

Doble Grado en Ingeniería Informática y Matemáticas

Fundamentos de Redes

La web distribuida: el protocolo IPFS

Autores:

José María Martín Luque

Adolfo Soto Werner

Profesor:

Antonio Ruiz Moya

17 de octubre de 2017

Índice

1 Problemas de HTTP	3
1.1 Fragilidad	3
1.2 Hipercentralización	5
1.3 Ineficiencia	5
1.4 Dependencia	7
2 La web distribuida	7
2.1 Tecnologías de web distribuida	7
3 El protocolo IPFS.	7
3.1 IPFS en detalle	7
3.2 Cómo soluciona IPFS los problemas de HTTP	7
4 La web distribuida en la actualidad	7
4.1 La Wikipedia descentralizada	7
4.2 Neocities	7
4.3 La web del referéndum catalán de 2017	7

Introducción

El Protocolo de transferencia de hipertexto (HTTP por sus siglas en inglés) es uno de los protocolos fundamentales de Internet. El desarrollo de HTTP comenzó en 1989 en el CERN por parte de Tim Berners-Lee. El desarrollo de un estándar fue un trabajo colaborativo entre el Grupo de Trabajo de Ingeniería de Internet (Internet Engineering Task Force, IETF) y el Consorcio WWW (World Wide Web Consortium, W3C), proceso que culminó con la publicación de una serie de *Request for Comments* (RFCs)¹. La primera definición de HTTP/1.1, la versión de HTTP más utilizada, apareció en el RFC 2068 de 1997.

Estamos hablando por tanto de un protocolo cuyo diseño comenzó hace más de 25 años. En aquel momento era inimaginable pensar que la tecnología que se estaba desarrollando fuese a ser usada por miles de millones de personas (3.885.567.619 a nivel mundial según las últimas estadísticas[1] de junio de 2017) ni se esperaba que tuviese tal repercusión sobre nuestras vidas. HTTP ha simplificado y facilitado la transmisión de información a nivel mundial. Gracias a ello hemos avanzado hacia una sociedad conectada donde la información y la cultura fluye libremente.

Pero como es de esperar, un protocolo que no fue diseñado con la visión del mundo actual presenta una serie de problemas a resolver. La intención de este trabajo es mostrar las deficiencias de HTTP y explorar una alternativa a la web actual, la web distribuida y el protocolo IPFS.

¹Los RFCs son un tipo de documentos técnicos del IETF que datallan técnicamente diversos aspectos del funcionamiento de Internet y otros protocolos.

1 Problemas de HTTP

1.1 Fragilidad



Figura 1: Primer servidor de HTTP del mundo. Se trata del ordenador personal de Tim Berners-Lee durante su estancia en el CERN.

Para entender por qué decimos que HTTP es *frágil* solo hay que observar la pegatina del primer servidor de HTTP: “*Esta máquina es un servidor. ¡¡No apagar!!*”. Está ahí para recordarnos que si se apagaba el servidor no se podía acceder al contenido. Otros sitios web en distintos servidores enlazaban a su contenido, de forma que si dejaba de estar en la red, todos esos enlaces no servían para nada. Por otro lado, si ese servidor se movía a otra ubicación, con otra dirección, todos esos enlaces habían muerto.

Hablamos en pasado pero efectivamente este sigue siendo un problema en la actualidad. No es raro entrar en algún enlace y encontrarnos con el error que podemos ver en la figura ???. Incluso si no conoces la especificación del protocolo

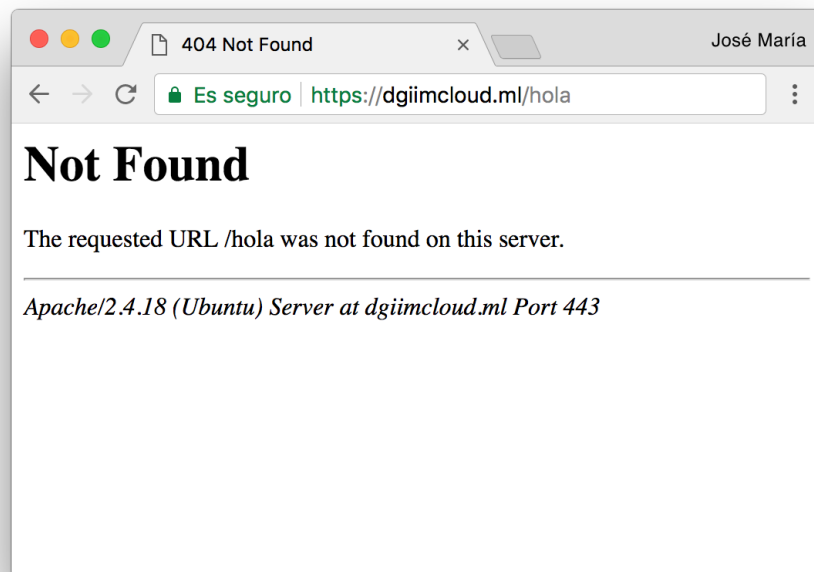


Figura 2: Error 404.

HTTP es probable que ya sepas que 404 es el código de error que nos indica que no hay nada que ver en una dirección.

La desaparición de enlaces (o *link rot* en inglés) es mucho más habitual de lo que se pueda pensar. El creador de [Pinboard](#), Maciej Cegłowski, estima que alrededor del 5 % de los enlaces que almacenan los usuarios en este servicio dejan de funcionar cada año[2]. Añade además que uno de sus clientes ha visto cómo dejaban de funcionar el 90 % de los enlaces que llevaba almacenados desde 1997. En 2014, un estudio de Jonathan Zittrain, Kendra Albert y Lawrence Lessig de la Harvard Law School concluyó que aproximadamente el 50 % de los enlaces que aparecen en resoluciones del Tribunal Supremo de los Estados Unidos ya no redireccionan a la información original[3]. Por estos motivos existen servicios como [The Internet Archive](#), que se dedica a guardar copias de las páginas web para preservarlas de cara al futuro, o el propio Pinboard, que ofrece a sus usuarios la posibilidad de almacenar copias de los artículos que añadan al servicio.

1.2 Hipercentralización

En la *Declaración de Independencia del Ciberespacio*[4] John Perry Barlow describía una utopía digital en la que los ciudadanos de la red se autogobiernan y las antiguas instituciones no tienen nada que hacer. “De parte del futuro, os pido a vosotros del pasado que nos dejéis en paz. No sois bienvenidos entre nosotros. No tenéis la soberanía del lugar donde nos reunimos.”

Por desgracia, esta no es la realidad en el año 2017. En la actualidad la web está altamente centralizada. La práctica totalidad de los usuarios de Internet dependen de una serie de servicios concretos. Por poner algunos ejemplos, en 2015 Facebook anunció que más de mil millones de usuarios utilizaron el servicio en un mismo día[5], mientras que una caída del servicio de Google en 2013 provocó una reducción del tráfico de internet del 40 %[6].

La *hipercentralización* de la red trae consecuencias negativas. Organizaciones como la NSA sólo tienen que intervenir el tráfico de unas pocas empresas para espiarnos, tal y como revelan las filtraciones de Edward Snowden[7]. La censura es mucho más fácil de establecer ya que sólo hay que bloquear el acceso a una serie de sitios concretos.

1.3 Ineficiencia

Para comprobar la ineficiencia de transmitir información por HTTP vamos a poner un ejemplo. El vídeo más visto de YouTube según Wikipedia a 17 de octubre de 2017 es “Luis Fonsi - Despacito ft. Daddy Yankee”[8], con 4.055.733.709 visualizaciones a las 11:47.

Supongamos que el vídeo siempre se reproduciese en 720p, de tal forma que el archivo pesa exactamente 67,1MB. 4.055.733.709 visualizaciones de un archivo de 67,1MB son 272.139.731.874MB descargados. Suponiendo que a Google le costase 1 céntimo transmitir 1GB de información (incluyendo todos los gastos del servidor), ya se habría gastado más de 272.139.731,874€ en transmitir un único vídeo.

Este precio de 1 céntimo por GB quizás sea posible para Google pero no lo es ni mucho menos para el ciudadano medio. La tabla de precios del CDN de Amazon, CloudFront es la siguiente[9]:

	Estados Unidos	Europa	Japón	India
Primeros 10TB/mes	0,085USD	0,085USD	0,140USD	0,170USD
Siguientes 40TB/mes	0,080USD	0,080USD	0,135USD	0,130USD
Siguientes 100TB/mes	0,060USD	0,060USD	0,120USD	0,110USD
Siguientes 350TB/mes	0,040USD	0,040USD	0,100USD	0,100USD

Figura 3: Tabla de precios de Amazon CloudFront en algunas regiones.

Podemos observar fácilmente que los precios son mucho mayores, ya que además se cobran las peticiones HTTP y HTTPS. La conclusión a la que queremos llegar es que a pesar de que HTTP ha abaratado muchísimo los costes de distribución de información, siguen siendo altos. Si el contenido a distribuir se encuentra en un servidor concreto, es necesario pagar los gastos generados tanto por distribuir el contenido como por el propio mantenimiento del servidor.

1.4 Dependencia

2 La web distribuida

2.1 Tecnologías de web distribuida

3 El protocolo IPFS

3.1 IPFS en detalle

3.2 Cómo soluciona IPFS los problemas de HTTP

4 La web distribuida en la actualidad

4.1 La Wikipedia descentralizada

4.2 Neocities

4.3 La web del referéndum catalán de 2017

Referencias

- [1] Miniwatts Marketing Group. *Internet World Stats*. 2017. URL: <http://www.internetworldstats.com/stats.htm> (visitado 14-10-2017).
- [2] Maciej Cegłowski. *Web Design: the first 100 years*. 2014. URL: http://idlewords.com/talks/web_design_first_100_years.htm (visitado 14-10-2017).
- [3] Jonathan Zittrain, Kendra Albert y Lawrence Lessig. “Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations”. En: *Legal Information Management* 14.2 (2014), págs. 88-99. DOI: [10.1017/S1472669614000255](https://doi.org/10.1017/S1472669614000255).
- [4] John Perry Barlow. *A Declaration of the Independence of Cyberspace*. 1996. URL: <https://www.eff.org/cyberspace-independence> (visitado 14-10-2017).
- [5] BBC. *Facebook has a billion users in a single day, says Mark Zuckerberg*. 2015. URL: <http://www.bbc.com/news/world-us-canada-34082393> (visitado 16-10-2017).
- [6] Sky News. *Google Outage Internet Traffic Plunges 40 percent*. 2013. URL: <http://news.sky.com/story/google-outage-internet-traffic-plunges-40-10437065> (visitado 16-10-2017).
- [7] Washington Post Barton Gellman y Ashkan Soltani. *NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say*. 2013. URL: https://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html (visitado 16-10-2017).
- [8] Wikipedia. *Videos más vistos en YouTube*. 2017. URL: https://es.wikipedia.org/wiki/Anexo:Videos_m%C3%A1s_vistos_en_Youtube (visitado 17-10-2017).
- [9] Amazon. *Precios de Amazon CloudFront*. 2017. URL: <https://aws.amazon.com/es/cloudfront/pricing/> (visitado 17-10-2017).