

¿Qué información recoge un servidor web?

Programación Web - Curso 2018-2019

David Gil Bautista

45925324M

Índice

- Introducción
- HTTP
- Cookies
- Facebook y las empresas de terceros
- Conclusión
- Bibliografía

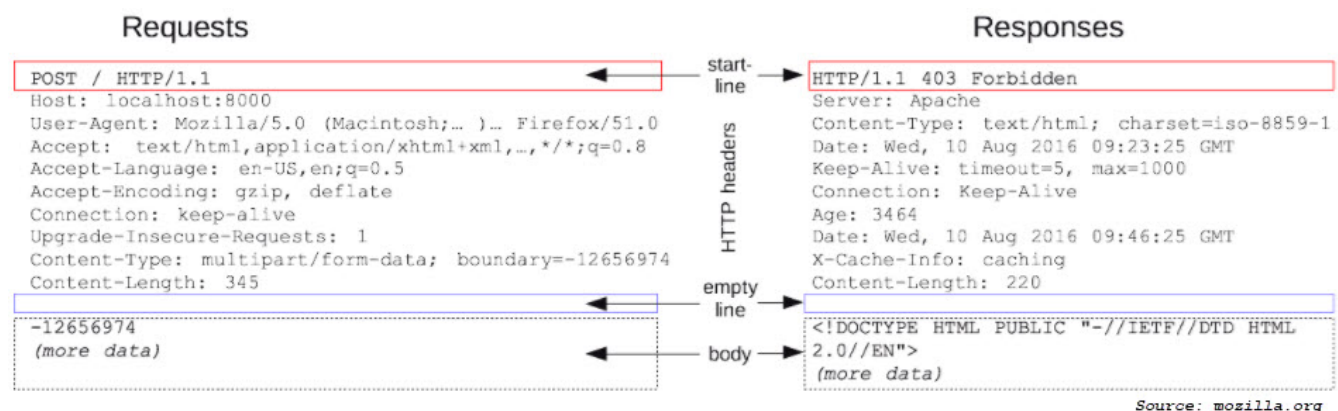
Introducción

Para responder esta pregunta lo primero que debemos preguntarnos es lo que es un servidor web. Un servidor es una máquina a la que nos conectamos y nos ofrece un servicio, en este caso un servicio web.

La web surge a finales de los 80 desarrollada por Tim Berners-Lee y es una herramienta que nos permite compartir información a través de una "telaraña", "red", "malla". Para almacenar esta información usamos páginas web. Dichas páginas las creamos usando un lenguaje de marcado, **HTML** (*Hypertext Markup Language*) generalmente. Para poder acceder a estas páginas necesitamos estar conectados a la web y un traductor capaz de transformar estas páginas en algo que podamos leer e interpretar, es aquí donde entra en juego el navegador. Este va a ser el encargado de conectarnos a través de la red y comunicarse con el servidor web para que éste nos preste un servicio.

Las comunicaciones en la web se regulan principalmente mediante el protocolo **HTTP** (Hypertext Transfer Protocol), aunque también hay más protocolos que nos permiten compartir ficheros, uso de correo electrónico...etc, por lo que la primera información que recogerá un servidor web es una petición HTTP que realiza un cliente cuando demanda un servicio. Vamos a ver qué podemos obtener de una petición HTTP.

HTTP



A priori esta imagen no aporta mucho, pero es lo que permite la conexión con el servidor. Cuando buscamos una página web nuestro navegador realiza un *"three way handshake"*, que es un proceso de tres pasos con los cuales consigue sincronizarse y conectarse con el servidor antes de comenzar el traspaso de datos.

Desde el navegador se manda una petición en la que aparece la máquina que manda el mensaje, el usuario, su idioma, la petición que realiza, el navegador que está usando... etc. Una vez establecida la conexión ya sabiendo el servidor toda esta información lo único distinto en las peticiones HTTP son las peticiones en sí mismas, es decir, a qué intenta acceder el usuario.

¿Podría ser relevante esta información para una empresa? Por supuesto, pongamos un ejemplo.

Entras en amazon y buscas robots de cocina. Amazon ya sabe que te estás informando sobre robots de cocina. Pero una vez cerrada la conexión debería olvidarse, pero un día distinto entras en amazon y ves que amazon comienza a recomendarte robots de cocina. **¿Qué está pasando?**

Cookies



Guardar toda la información de un usuario en el servidor es absurdo, tendríamos muchísimos más datos y la mayoría serían inútiles. Por ello se hace uso de las cookies, las cuales nos permiten almacenar información del usuario en su propia máquina y una vez establece la conexión con el servidor tan sólo tiene que leer las cookies, cosa que es mucho más rápida que buscar en su almacenamiento la información que concuerda con este usuario.

En páginas web con anuncios puede que estos anuncios se guarden en las cookies y al entrar en otra página cuyos anuncios pertenezcan a la misma compañía que la primera web, esta página lea las cookies y sepa qué páginas has estado visitando. Para empresas de venta de productos o anuncios esta información es muy valiosa ya que el crear anuncios dirigidos a los intereses de un usuario aumenta las posibilidades de que el usuario sea un potencial comprador.

Aunque las cookies son un recurso que genera muchísimos datos que las empresas pueden llegar a usar hay una empresa experta en sacar un beneficio a partir de los datos de los usuarios.

Facebook y las empresas de terceros

Facebook nace como una red social creada por un estudiante sin ánimo de lucro, pero conforme empieza a aumentar la cantidad de personas que la usan se da cuenta de que todos esos datos son una mina por explotar. Ya de por sí damos nuestros datos personales a Facebook, ya sea el nombre, cumpleaños, trabajo, escuela... etc. Aunque la mayoría de la gente que usa Facebook también usa aplicaciones de terceros en las que se puede vincular la cuenta de Facebook, por lo que te ahorras registrarte y tener que memorizar otra contraseña más. Cada vez que te conectas a una aplicación nueva usando Facebook estás generando más datos que Facebook, o cualquier empresa de terceros puede utilizar. Ya hemos visto cómo las empresas usan estos datos para generar anuncios, aunque puede que algunos de estos datos sean vendidos en la *deep web* y que posteriormente alguien pueda llegar a robarnos la identidad.

Conclusión

En estos últimos años el uso de la web ha ido creciendo exponencialmente hasta el punto de poder encontrar cualquier cosa en internet. Debido a esto los usuarios han comenzado a usar la web como una herramienta para absolutamente todo, desde hablar con tu madre hasta gestionar los movimientos de tu cuenta bancaria, es por esto que cualquier filtración de estos datos puede hacer que pierdas la vida en internet, metafóricamente hablando.

Para intentar evitar esto siempre podemos configurar la privacidad de nuestras aplicaciones, o algo que no mucha gente hace, leerse los términos y aplicaciones de un servicio web y en caso de que no lo consideremos fiable no usarlo. La web es una herramienta muy poderosa, pero también un arma, y las empresas que recolectan nuestros datos son las que tienen la munición.

Bibliografía

<http://www.whatarecookies.com/>

<https://stackoverflow.com/questions/1968722/how-cookies-work>

<https://docs.trafficserver.apache.org/en/8.0.x/developer-guide/plugins/http-headers/index.en.html>

<https://en.wikipedia.org/wiki/Handshaking>

<https://en.wikipedia.org/wiki/HTML>

<http://fortune.com/2018/03/21/facebook-personal-data-cambridge-analytica/>