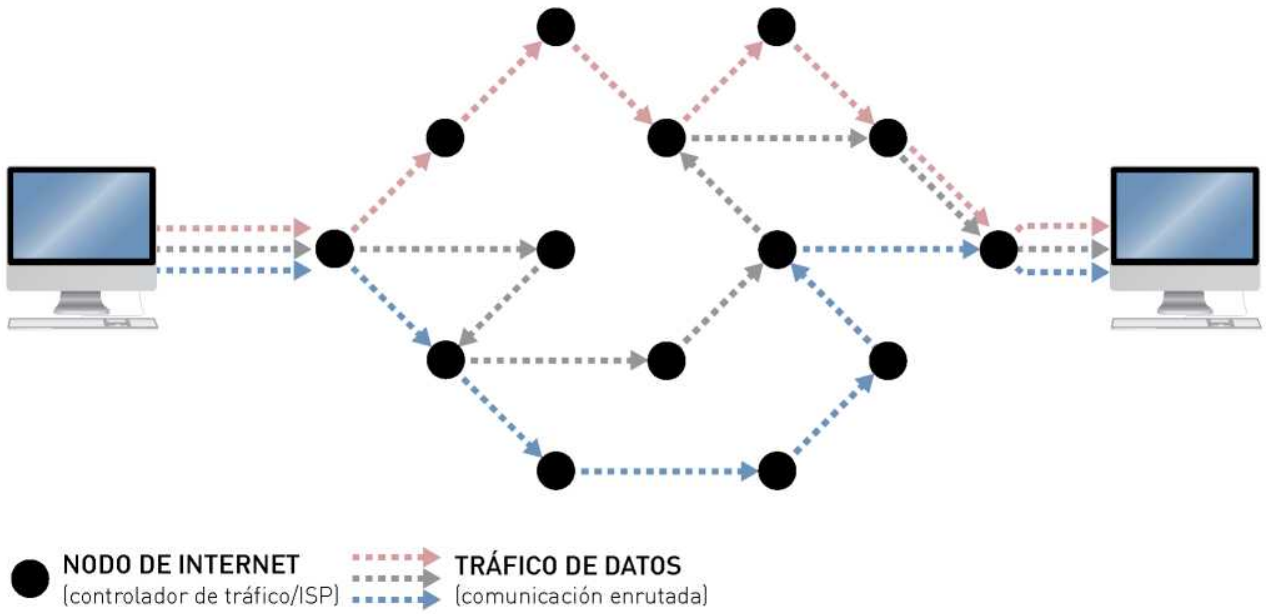
INTERNET

UNA RED DE REDES DE ORDENADORES



Internet es un sistema mundial de redes de ordenador interconectadas.

Cuando dos o más dispositivos electrónicos (ordenadores, por ejemplo) se conectan entre sí para poder comunicarse, pasan a formar parte de una red. Internet es la interconexión a escala mundial de esas redes, pertenecientes a empresas, gobiernos y particulares, lo que permite que todos los dispositivos conectados a ellas puedan comunicarse entre sí.

Para poder comunicarse, los ordenadores necesitan entenderse entre ellos. En Internet, esto es posible porque todos los dispositivos emplean el mismo "lenguaje" o protocolo, concretamente el Protocolo IP (Internet Protocol), un "mercado único" sin barreras físicas, técnicas o nacionales. Es la base de todos los sistemas de comunicación a través de Internet.

Enviar información por Internet usando el protocolo IP es como mandar por correo postal las páginas de un libro en muchos sobres distintos. Todos los sobres tienen la misma dirección del remitente y la misma dirección del destinatario. Aunque algunos sobres viajen por mar y otros por aire, tarde o temprano todos llegarán a su destino previsto y el libro podrá ser encuadernado de nuevo. Da lo mismo que la página 1 llegue después que la 47.

En Internet, el contenido del sobre también sigue protocolos (convenciones, formatos consensuados), uno para cada tipo de comunicación. Algunas de estas convenciones sobre IP son:

- SMTP para enviar correos electrónicos
- HTTP para acceder a sitios web y

- para compartir archivos por P2P (una forma de intercambiar archivos de datos con gran número de personas).

Cualquiera puede inventarse un protocolo y usarlo en Internet, siempre y cuando funcione sobre IP. Dicho de otro modo: el único límite es la imaginación y la única regla es que la dirección del sobre esté en un formato estándar. Su carácter abierto es lo que ha hecho de Internet un fenómeno global y cualquier restricción de su transparencia reduciría su potencial de evolución.

El uso universal de un protocolo único para todas las comunicaciones tiene varias ventajas importantes. Los routers (aparatos encargados de transportar datos a través de Internet) no necesitan ser programados en función del tipo de datos. Ni siquiera necesitan tener información sobre los datos que transportan siempre y cuando se use el Protocolo IP. Solo tienen que leer lo que pone en el sobre para poder entregar el mensaje, igual que el cartero que reparte las cartas.

Da igual que en el sobre haya una factura o una carta de amor (salvo para quien lo reciba, claro).

Esto lleva a:

- Posibilidades ilimitadas de innovación en materia de nuevos protocolos y aplicaciones;
- "Privacy by design" (privacidad desde el diseño): no es necesario conocer el contenido de las comunicaciones;
- Flujo de datos rápido y flexible;

En el fondo, Internet solo ofrece un servicio flexible: llevar datos de un dispositivo a otro independientemente de la naturaleza de estos, de dónde y cómo se conecten a Internet o del tipo o contenido de los datos.

Este carácter abierto y flexible es la causa principal de la innovación en Internet y de su éxito democrático y económico.

“Este carácter abierto y flexible es la causa principal de la innovación en internet y de su éxito democrático y económico."

LA DIRECCIÓN IP

UNA DIRECCIÓN DIGITAL

La dirección IP es una dirección numérica asignada a cada uno de los dispositivos conectados a Internet.[1]

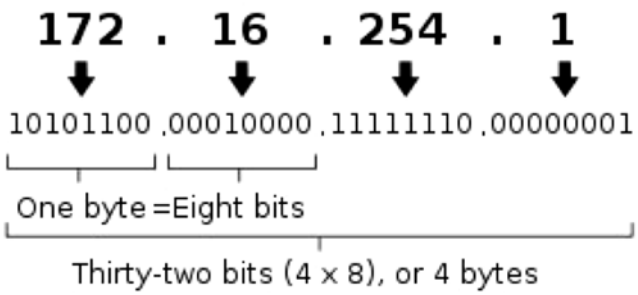
En muchos casos, las direcciones IP pueden servir para identificar a una organización o a un particular que ha contratado Internet a través de un Proveedor de Servicios de Internet para conectar uno o más dispositivos.

En otros casos, especialmente en redes de empresa, conexiones inalámbricas públicas o desprotegidas y conexiones móviles, la dirección IP no siempre permite identificar a la persona que ha llevado a cabo una acción rastreable electrónicamente.

Como la mayoría de routers domésticos y de empresa suelen mostrar una única dirección IP para toda la gente conectada a esa red, la IP identificaría a un grupo de personas y no a un individuo concreto. Por este motivo, a menudo es complicado, si no imposible, saber con seguridad quién hizo exactamente qué basándose solo en la IP.

Por otro lado, dado que muy a menudo la dirección IP puede identificarte personalmente, como medida de precaución hemos de darlo por supuesto, salvo que estemos seguros de lo contrario.

Dirección IPv4 (notación decimal con puntos)



“la dirección IP no siempre permite identificar a la persona que ha llevado a cabo una acción rastreable electrónicamente"

[1]Debido a que la versión actual del protocolo tiene un límite de direcciones de red admisibles, es cada vez más frecuente, sobre todo en las redes de empresa, compartir una dirección IP (entre todos los ordenadores de una oficina, por ejemplo). Esta restricción se solventará con la implantación de direcciones IPv6.

CIFRADO

PRIVACIDAD EN UNA RED PÚBLICA

Una carta puede ser interceptada, abierta, leída y luego cerrada sin dejar huella. Una llamada telefónica también puede ser intervenida. ¿Cómo hacemos para enviar un mensaje comprometedor sin que acabe en las manos equivocadas?

El desarrollo de la criptografía se disparó en el siglo XX con los avances de las tecnologías informáticas. Los ordenadores multiplicaban la velocidad con la que se podían cifrar los mensajes electrónicos y permitían descifrar mucho más rápido las claves criptográficas empleadas hasta entonces.

El cifrado no es infalible y no garantiza al cien por cien la confidencialidad. Una técnica habitual para esquivar el cifrado es capturar el mensaje antes de que sea cifrado -por ejemplo, por un troyano sigiloso instalado en el ordenador del remitente que captura todas las pulsaciones del teclado o incluso del teléfono móvil de la víctima.

Otro atributo que casi siempre deberás proteger al cifrar un mensaje es su integridad (es decir, que el archivo esté intacto). De lo contrario, el mensaje podría ser manipulado aún sin conocer la clave de cifrado. La mayoría de las herramientas de cifrado más respetadas lo hacen automáticamente por ti.

La imagen anterior muestra las fases de cifrado con clave pública, que utiliza dos claves, una pública y otra privada:

1. El remitente solicita una copia de la clave pública.
2. Usando el programa adecuado, el remitente cifra el mensaje utilizando la clave pública del destinatario.
3. El mensaje es enviado.
4. El destinatario descifra el mensaje utilizando la clave pública y la clave privada.

EL SISTEMA DE NOMBRES DE DOMINIO (DNS)

Cuando cuelgas un sitio web en Internet, este será accesible mediante la dirección IP numérica del servidor web donde está alojado (por ejemplo, en el momento de escribir esta guía, la dirección IP de edri.org es 217.72.179.7). El problema es que las direcciones IP no son fáciles de recordar para los humanos. Usarlas para identificar recursos online tampoco es práctico, porque en Internet en ocasiones los servicios cambian de IP (si cambian de proveedor de servicios, por ejemplo).

Como el uso de direcciones IP para páginas web no es ni práctico ni intuitivo, se crearon los "nombres de dominio" (como edri.org). El Sistema de Nombres de Dominio (DNS) es en parte el equivalente en Internet a la guía telefónica.

Si conoces el nombre de dominio del sitio web que quieres visitar, el DNS encontrará automáticamente la dirección IP del servidor web donde se aloja la página. Es decir, si escribes http://edri.org, tu ordenador identifica el dominio con la IP 217.72.179.7 y envía una petición a nuestra página web.

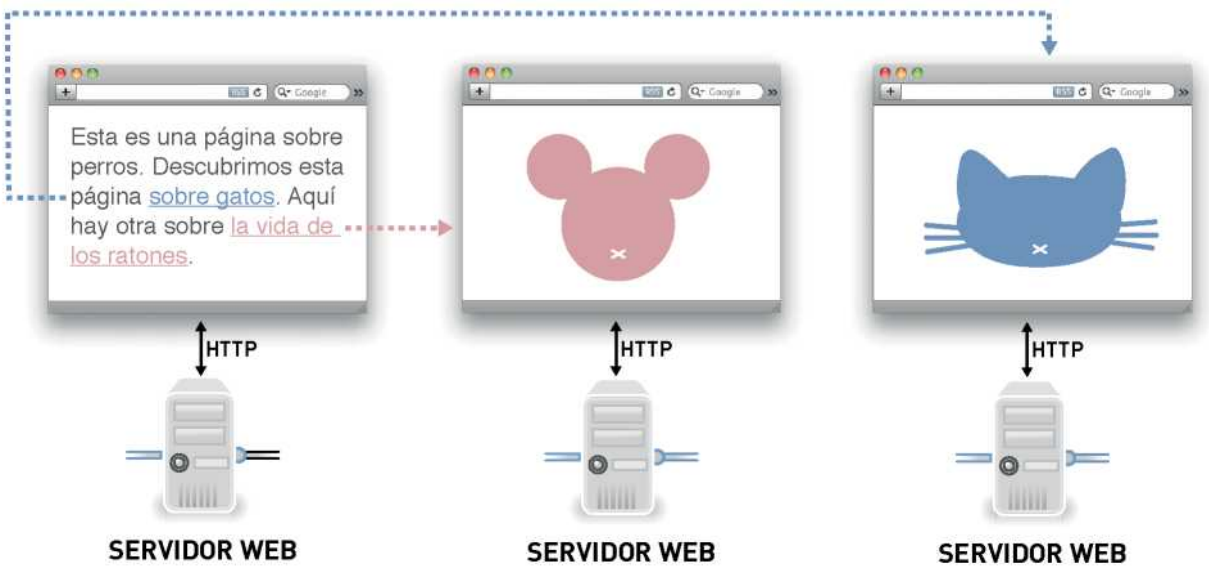
El sistema de búsqueda del nombre de dominio sigue una estructura jerárquica de árbol. Al teclear http://edri.org, tu ordenador primero se conecta a un servidor para pedir la dirección IP.[2] Por defecto, el servidor DNS está gestionado por tu operador de Internet, pero se puede utilizar otro.

Si alguien ha accedido recientemente a http://edri.org, el servidor DNS "recuerda" los datos y te facilita la dirección IP correcta. En caso contrario, deriva la consulta a un nivel superior de autoridad, donde se repetirá el proceso. El máximo nivel de autoridad son 13 "servidores raíz" que recopilan los servidores DNS. Estos 13 servidores son muy estables y tienen una enorme capacidad, tanta que han seguido funcionando sin problemas incluso durante ataques graves (los llamados "ataques de denegación de servicio").

[2] Si tu ordenador ha accedido recientemente a http://edri.org, ya conoce su dirección IP y no necesita comprobarla con el proveedor de servicios.

LA WEB

ENLAZANDO LA SOCIEDAD DE LA INFORMACIÓN



La Web se basa en HTTP, un protocolo (lenguaje) relativamente reciente de nivel superior al Protocolo IP. El HTTP (HyperText Transfer Protocol - protocolo de transferencia de hipertexto) permite descargar documentos de hipertexto (lo que ahora se conocen como "páginas web") y enviar información básica a su servidor web.

Las páginas web están escritas en un lenguaje de etiquetas llamado HTML, (HyperText Markup Language). El World Wide Web Consortium (W3C) define las reglas del lenguaje y especifica marcadores especiales para indicar la tipografía y el estilo del texto. Por ejemplo, el texto en negrita tendrá **antes y** después.

Si bien hay varias versiones de la especificación, (la más reciente es HTML5), el protocolo HTML está en desarrollo permanente y abierto a la participación. Una vez definidos los estándares, no hay ninguna licencia o coste por usar HTML. La ventaja es que todos los sistemas informáticos disponibles entienden de la misma manera las instrucciones en HTML, por lo que cualquiera puede usar este lenguaje (gratis) con la seguridad de que todos los dispositivos mostrarán la página web de la misma manera. La web (y el mundo) sería bastante peor si la gente tuviera que pagar por diseñar páginas en los lenguajes de distintos tipos de ordenador.

Este carácter abierto y gratuito del HTML es imprescindible para garantizar que las páginas webs sean compatibles con todo tipo de dispositivos: ordenadores de sobremesa, portátiles, teléfonos móviles, tabletas, etc.

Respetar la especificación HTML al programar páginas web también garantiza la accesibilidad de las personas con deficiencias visuales - de lo contrario, los sistemas de lectura de texto no entenderán las páginas visitadas.

Las páginas web se publican en máquinas conocidas como "servidores web". Un servidor web es un ordenador que puede ser localizado por su dirección IP única (como se explica en la página 5). Normalmente variante segura llamada HTTPS. Las conexiones HTTP (y, por tanto, la información enviada y recibida) no están cifradas,

y cualquier persona con acceso al hardware de red entre el ordenador del usuario final y el servidor web podría interceptar la información enviada y recibida.

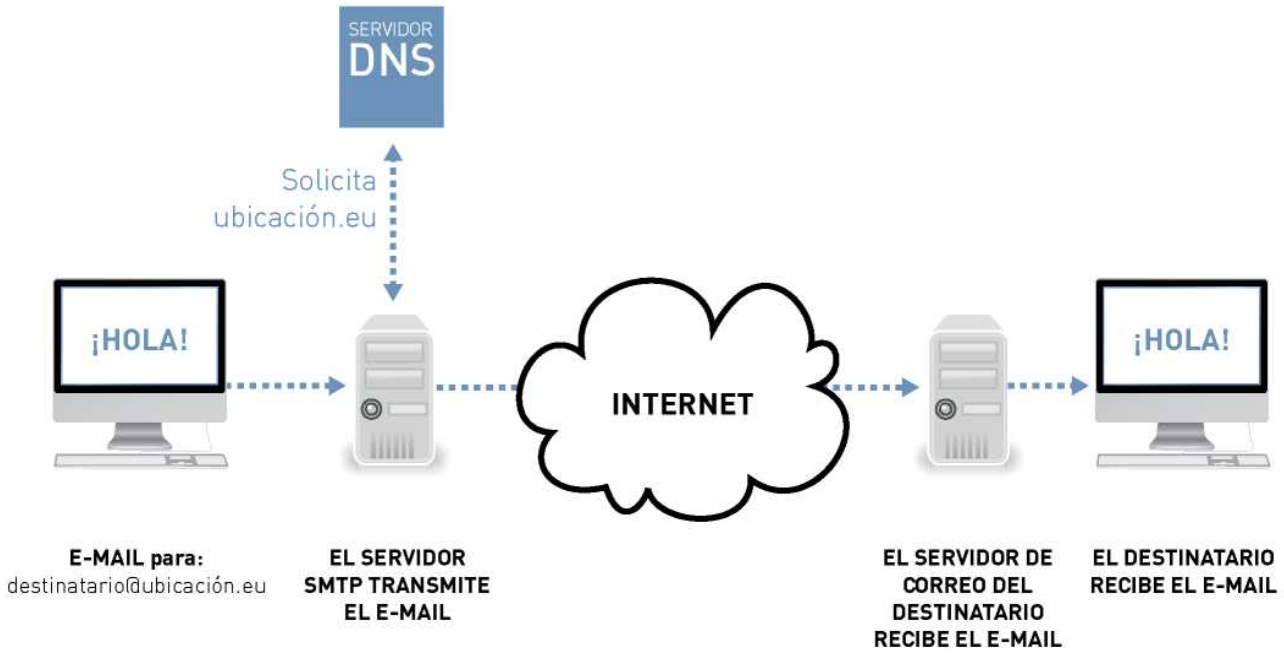
El HTTPS incorpora cifrado a esta conexión, de forma que (en teoría) solo el usuario final y el servidor web pueden descifrar la información enviada y recibida. Esto se basa en la confianza: el editor de una página web solicita a una entidad de confianza un certificado estrictamente personal, firmado una misma IP es compartida por muchos dominios (como www.edri.org y www.bitsoffreedom.nl), porque están almacenados ("alojados") en el mismo servidor. Es decir, un servidor, con una dirección IP única, puede alojar multitud de páginas web. En el caso de las empresas de alojamiento web, un único servidor puede contener cientos de páginas sin relación entre sí. Por tal motivo, los intentos de "bloquear" páginas web concretas basándose en su dirección IP siempre han tenido consecuencias nefastas para el resto de páginas alojadas en el mismo servidor.

Además del HTTP, existe también una electrónicamente para confirmar la identidad del editor, de forma parecida a los sellos de cera que se usaban hace siglos para sellar documentos.

Cuando te compras un ordenador o instalas un nuevo navegador, incluye por defecto un listado estándar de autoridades certificadoras de confianza que pueden ser utilizadas por los editores de páginas web. Esta lista estándar es el punto débil del sistema: incluye decenas de entidades. Con que una de ellas resulte no ser fiable, los usuarios estarán confiando en un servicio sin garantías.

CORREO, ELECTRÓNICO Y SEGURIDAD

MAIL IN THE DIGITAL WORLD



Los correos electrónicos, o eMails, son mensajes enviados por un remitente a uno o más destinatarios. La transferencia de estos mensajes se realiza por SMTP (Simple Mail Transfer Protocol - Protocolo simple de transferencia de correo) que, al igual que el HTTP, es un protocolo que funciona en un nivel sobre gatos superior sobre IP.

Cuando enviamos un eMail desde el correo web o a través de un cliente de correo, es transferido a un servidor de correo saliente, que a su vez lo transferirá a otro servidor, usando siempre SMTP, y así sucesivamente hasta llegar al servidor de destino.

Para saber adónde tienen que enviar los eMails, los servidores de correo realizan una consulta al DNS (como explicamos antes), que responde incluyendo información sobre qué servidores son responsables de procesar los eMails de ese dominio. El dominio puede extraerse de la parte de la dirección de correo del destinatario que va después de la arroba.

Una vez que el mensaje llegue al servidor de correo que gestiona todos los eMails del destinatario, permanecerá ahí hasta que lo borre. Algunos clientes de correo eliminan automáticamente los correos del servidor en cuanto se descargan en el ordenador o smartphone del usuario.

Seguridad y correo electrónico Los eMails pueden ser interceptados por terceros mientras están siendo enviados de un servidor de correo a otro. Hay dos maneras de impedirlo: usar una comunicación segura entre servidores o cifrar el contenido de los mensajes. La comunicación segura entre servidores de correo funciona de la misma manera que el protocolo HTTPS para proteger la comunicación HTTP (descrito anteriormente).

En el caso de los correos electrónicos, la desventaja es que tu ordenador no se comunica directamente con el servidor de destino, lo que significa que si uno de los servidores de correo intermedios no usa cifrado para enviar tu mensaje, podría ser interceptado en ese punto.

Esta vulnerabilidad hace que sea más recomendable cifrar el mensaje. Un método muy popular y gratuito para cifrar correos electrónicos es PGP (Pretty Good Privacy), también disponible como OpenPGP y GPG.

DEEP PACKET INSPECTION

UN VISTAZO A TU TRÁFICO EN INTERNET

En internet, los datos se envían en “paquetes", que son básicamente pequeños bloques de datos. Cada paquete tiene una cabecera con información sobre su origen y su destino (igual que un sobre tiene la dirección del destinatario y la del remitente). Esta información permite al hardware de red determinar la mejor ruta para enviar un paquete en un determinado momento.

En el pasado, el hardware de red solo se fijaba la información sobre el origen y el destino, pero ante el rápido aumento de la actividad fraudulenta, los administradores de redes decidieron que necesitaban analizar más datos de cada paquete para distinguir los paquetes "seguros” de los que eran parte de ataques de hackers o de denegación de servicio.

Al principio, los programas de seguridad de red ("cortafuegos”) solo podían bloquear un paquete que viajara de un origen concreto a un destino concreto y a un servicio concreto. Siguiendo estos criterios, podías bloquear todas las conexiones entrantes a la red de tu oficina y permitir las conexiones salientes para acceder a otros servicios en Internet.

Puede que en algún momento decidas montar un servidor web en tu red para publicar documentos. Necesitarías cambiar la configuración del cortafuegos para permitir las conexiones entrantes, pero solo para el servicio web. Sin embargo, hay numerosos ataques contra servidores web que parecen bastante inocentes desde el punto de vista del cortafuegos. Es imposible distinguir los paquetes legítimos de los maliciosos basándose únicamente en la información de origen y destino.

Los ingenieros de redes pronto se dieron cuenta de sería más fácil detectar los ataques si el hardware de red inspeccionara más a fondo los paquetes. En teoría es sencillo - las cabeceras de los paquetes no están "separadas” más allá de la definición lógica de límites. Se trata de analizar unos cuantos bytes más de los que estábamos analizando hasta ahora, por ejemplo para enrutamiento de datos. O ir un poco más profundo y mirar dentro del bloque de datos del paquete.

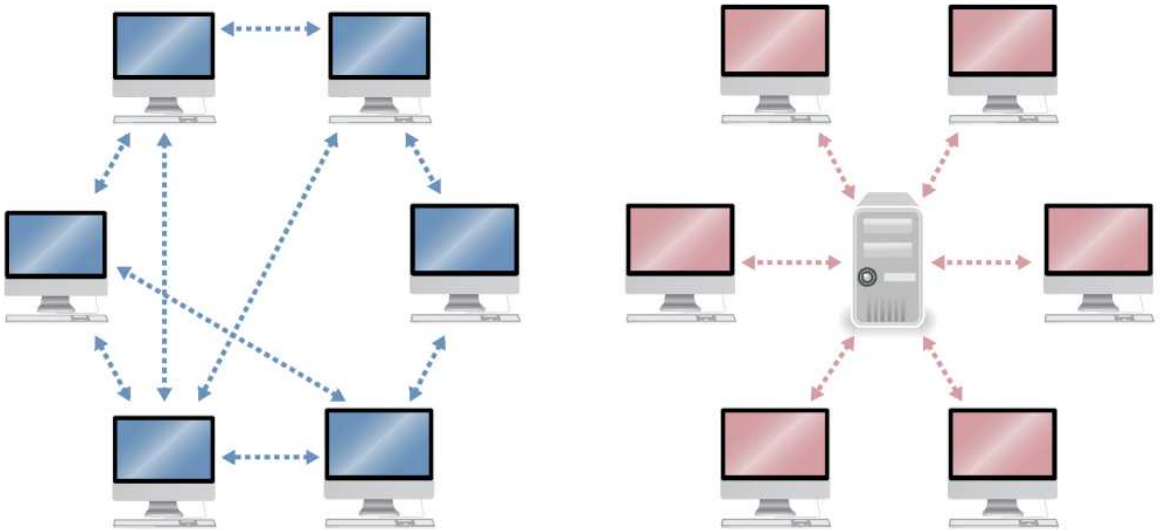
Los dispositivos que empezaron a hacer eso se denominaron Sistema de Prevención de Intrusiones (IPS) y pronto, la gran mayoría del hardware de red incorporaba esta funcionalidad. Mientras se utilizaba para bloquear ataques de piratas informáticos, no hubo controversia.

Sin embargo, con el tiempo, los gobiernos, los proveedores de contenidos y las operadoras de red empezaron a darse cuenta de que la técnica - conocida como deep packet inspection (DPI - inspección profunda de paquetes) les da un control mucho mayor de los datos de los usuarios de la red. Las autoridades ya utilizan técnicas de DPI para labores de vigilancia, bloqueo, etc. y se está valorando utilizarlas para el cumplimiento de los derechos de autor. Otras de sus aplicaciones son la segmentación de mercados, la segmentación publicitaria, el cumplimiento de acuerdos de nivel de servicio...

El usuario, por su parte, puede bloquear las técnicas DPI usando cifrado; los contenidos "profundos” de un paquete cifrado son totalmente opacos para el operador.

PEER-TO-PEER

DE MÍ PARA TI SIN NADIE DE POR MEDIO



Las redes Peer-to-Peer o P2P están compuestas por dispositivos (grandes servidores y ordenadores personales) que participan en un tipo de comunicación en igualdad de condiciones. Cada nodo (dispositivo) puede comunicarse con otros nodos y no hace diferenciación entre consumidores y productores, clientes y servidores, etc. Sencillamente, son muchos dispositivos comunicándose con muchos dispositivos.

Es lo contrario al modelo cliente-servidor, en el que un ordenador responde a las peticiones de muchos clientes. Un ejemplo sería un sitio web que da contenido a muchos usuarios (un dispositivo que se comunica con muchos dispositivos).

En Internet, las aplicaciones P2P usan protocolos P2P sobre IP.

Las redes P2P presentan una serie de ventajas:

- No tienen punto único de fallo porque no hay entidades centralizadas. En una red cliente-servidor, si el servidor falla, el sistema falla. En una red entre pares, si uno de los nodos falla, el daño general es mínimo;
- Pueden crecer fácilmente porque cada nuevo nodo añade más recursos (capacidad de tráfico, almacenamiento, capacidad de procesamiento) a la red;
- No están administradas porque no hay una autoridad central.
- Los fallos tienen un impacto mínimo porque los recursos están descentralizados y tiende a haber un alto nivel de duplicación de recursos.

- Dan libertad a sus usuarios. No solo los dispositivos están en igualdad de condiciones, sino también los usuarios de la red P2P.
- Una de las tareas más importantes de una aplicación P2P es organizar la red y localizar recursos en ella.

En parte, los servidores de correo electrónico son un ejemplo precoz de aplicaciones P2P: cualquier servidor que utilice el protocolo SMTP puede enviar un correo a otro servidor. El Sistema de Nombres de Dominio también puede devolver una lista de servidores capaces de procesar el correo entrante de un dominio concreto, lo que aumenta la fiabilidad del sistema.

En las redes de intercambio de archivos, los nodos no conocen directamente la dirección IP de los otros nodos ni saben qué archivos (o partes) tiene cada nodo. Esto se solventa mediante un proceso en el que los usuarios comparten información sobre los contenidos que tienen. Los archivos se identifican mediante archivos de "hash", que básicamente son huellas digitales que permiten que cada archivo sea único e identificable. Las Tablas de Hash Distribuidas (DHT) permiten a los usuarios saber qué usuarios tienen el archivo (o algunas de sus partes) que quieren descargar.

Los usuarios de la red P2P necesitan conseguir los hash de los archivos que buscan. Por ejemplo, las páginas web desde las que se pueden descargar versiones del sistema operativo Ubuntu publican estos archivos. Existen diccionarios que convierten las huellas hash en descripciones comprensibles para los humanos, lo que permite buscar archivos en la red P2P.

Páginas como thepiratebay.org y mininova.org proporcionan estos diccionarios, pero las huellas también pueden distribuirse por correo electrónico, chat y redes sociales, lo que significa que no hay un sistema centralizado.

También existen redes P2P que mantienen el anonimato de sus usuarios.

PUBLICIDAD COMPORTAMENTAL

ES ALGO PERSONAL

La ('behavioural advertising' en inglés) es una técnica basada en rastrear en internet los hábitos de los usuarios. Se utiliza para crear perfiles de internautas y mostrarles anuncios que, si el perfil es correcto, serán más relevantes para ellos y por tanto más efectivos.

La publicidad comportamental sigue un principio muy simple: si por ejemplo un internauta visita por primera vez un sitio web sobre fútbol, un pequeño archivo llamado cookie se guardará en su navegador (Firefox, Internet Explorer, Chrome...). Normalmente, una página web tiene contenido de diversas fuentes; el texto y las imágenes pertenecen al sitio que estás visitando, pero otros contenidos, como la publicidad, se cargan desde otras fuentes (que pueden no tener relación con el sitio web). Cada vez que se carga nuevo contenido, la petición también puede enviar información de las cookies desde tu ordenador.

Las cookies utilizadas en publicidad comportamental suelen incluir un número de identificación. Si luego ese usuario lee una noticia sobre coches, las empresas de publicidad podrán hacer suposiciones sobre alguien que lee artículos sobre coches y fútbol. En este ejemplo, podrían dar por supuesto que el usuario se mostrará receptivo a los anuncios de cerveza. También podrían dar por supuesto que no sería buena idea mostrarle ofertas de seguros de coches porque probablemente sea muy joven.

Cuantas más páginas visite el usuario que formen parte de la red de rastreo de los servicios de publicidad comportamental (como la mayoría de los sitios web de noticias, entre otros muchos), más datos se añadirán a su perfil. Tras leer los hábitos online de un usuario durante relativamente poco tiempo, puede crearse un perfil muy detallado que podría servir para identificarle, aunque en teoría sea "anónimo".

A medida que se van incorporando nuevos datos al perfil del usuario, el grupo en el que se incluye se va acotando hasta un alcanzar un número muy pequeño de individuos que encajan en ese patrón. Hace unos años, un buscador publicó un lote de búsquedas "anónimas" realizadas a través de su servicio. Tras analizar los datos, algunos periodistas pudieron identificar a personas, lo que demuestra que los datos "anónimos" resultaron no serlo tanto.

Se desconoce si la publicidad comportamental usa datos de otras fuentes. Muchas compañías activas en el negocio de la publicidad comportamental, como Google y Yahoo!, ofrecen también otros servicios, como buscadores. El cruce de bases de datos produciría enormes cantidades de datos de carácter personal que permitirían identificar a una persona de forma relativamente sencilla.

La publicidad comportamental ha sido uno de los motores del éxito económico de la publicidad online en los últimos años. La técnica también se usa de forma experimental para enviar otros contenidos a los internautas, como noticias, por ejemplo.

Este uso de la información de carácter personal del usuario se realiza sin su consentimiento. La industria publicitaria sostiene que este tipo de rastreo beneficia al usuario porque ayuda a que solo reciba publicidad "relevante". También proponen un procedimiento de cancelación (opt-out) que afirman que cumple los requisitos de la Directiva europea sobre la Privacidad y las Comunicaciones Electrónicas, que obliga a obtener el consentimiento informado de los usuarios.

El quid de la cuestión es si tener el navegador configurado para aceptar cookies (que habitualmente es la opción activada por defecto) puede considerarse una aceptación válida por parte del usuario. El Supervisor Europeo de Protección de Datos[3] dice que no. Muchos internautas ni siquiera saben qué son las cookies o nunca han modificado las opciones de privacidad. Técnicamente, la solución propuesta también plantea dificultades, ya que el régimen "opt-out" no incluye a todos los anunciantes. Es más, el actual sistema "opt-out" usa cookies, así que al borrar las cookies también se borra el opt-out.

Además, los navegadores actuales y las extensiones (llamadas plug-ins, como Flash) disponen de muchas más formas de almacenar y obtener datos aparte de las tradicionales cookies. Estos datos adicionales son difíciles de gestionar por el usuario medio, y no siempre están incluidos en las preferencias sobre aceptación de cookies de los navegadores.

Actualmente, hay una ley europea para proteger a los ciudadanos ante estas prácticas, pero continúan desprotegidos de facto por falta de voluntad para aplicar la legislación.

[3] http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2009/09-01-09_ePricacy_2_EN.pdf

EL BUSCADOR

UN ÍNDICE DE INTERNET

La navegación por la Red se realiza a través de hipervínculos (texto o imágenes que, al hacer clic en ellos, llevan a otro sitio web).

Cualquiera que cree una página web puede añadir enlaces a otros contenidos online. Gracias a esta práctica de añadir enlaces, todos los internautas contribuyen a organizar la información en una Red de recursos interconectados.

Hay que tener en cuenta que la Web no tiene un índice centralizado que registre todo lo que está disponible en la red, por lo que los buscadores se hacen indispensables para navegar por Internet de forma más eficiente.

Existen varios tipos de buscadores. El más importante es el buscador jerárquico, que utiliza software (conocido como "robot" o "spider") para buscar contenidos en la Web e indexarlos sistemáticamente. La complejidad y eficacia del robot determina el tamaño y la actualización del índice, dos datos importantes para medir la calidad de un buscador. Para que nos entendamos: el spider sigue todos los enlaces de todas las páginas, indexa las páginas enlazadas y luego sigue los enlaces de esas páginas, los indexa y así sucesivamente.

La operación más importante que realiza un buscador es cotejar la búsqueda que realiza el usuario con la información indexada.

Por lo general, el resultado es una lista de referencias ordenadas según su relevancia. Suelen incluir un título, un fragmento de información e hiperenlaces a las páginas que la tecnología del buscador considera que pueden ser relevantes.

Junto con los "resultados orgánicos" (es decir, las páginas encontradas por el buscador), los buscadores comerciales incluyen resultados patrocinados, determinados por un proceso de subastas de palabras clave a los anunciantes. El proceso de búsqueda de resultados orgánicos es complejo y los principales buscadores comerciales protegen sus algoritmos de clasificación como secretos de empresa. PageRank, propiedad de Google, es uno de los algoritmos de posicionamiento más conocidos. Predice la relevancia de las páginas web en el índice analizando la estructura de enlaces en la Red (es decir, el tipo de páginas que enlazan con esa página).

Otras técnicas utilizadas para que los resultados sean más relevantes para el usuario son el análisis del contenido de los sitios web y el análisis de los datos de usuario. Los buscadores comerciales utilizan cookies para saber las búsquedas que realiza cada usuario, en qué enlaces hace clic y mucho más. Esta información se guarda en formularios individuales durante largos periodos de tiempo.

Los buscadores "verticales" o temáticos están especializados en búsquedas sobre un tema concreto, como viajes, compras, artículos académicos, noticias o música. Los grandes buscadores jerárquicos incluyen también buscadores especializados como parte de sus servicios. Un metabuscador es un motor de búsqueda que no genera su propio índice y resultados de búsqueda sino que utiliza los resultados de uno o más buscadores externos. Un "directorio" es un repositorio de enlaces clasificados por categorías. Ejemplos famosos son el directorio de Yahoo! y el Open Directory Project.

CLOUD COMPUTING

INTERNET SE CONVIERTE EN TU ORDENADOR

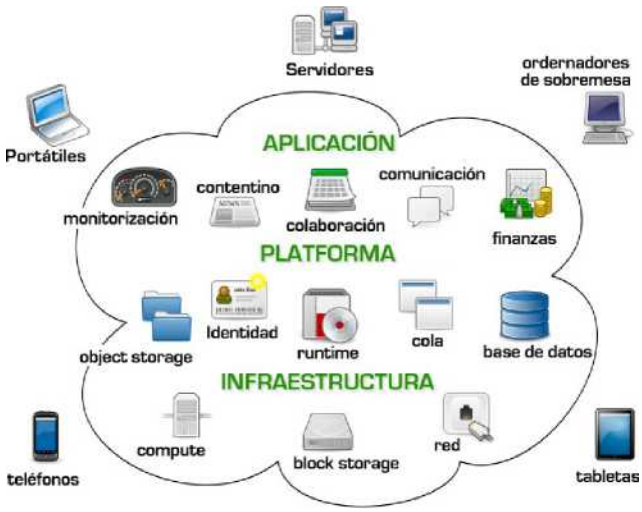
La expresión “en la nube" se ha puesto de moda últimamente. El concepto en sí no es nada nuevo, aunque en los últimos tiempos se ha producido un aumento espectacular de las aplicaciones disponibles.

En los diagramas para representar una red de comunicaciones se usa una nube para mostrar la red que está fuera de la red del usuario. "Informática en la nube” hace referencia a cualquier servicio informático que se ejecuta dentro de la red en vez de en el ordenador del usuario final.

Uno de los primeros ejemplos de computación en la nube es el correo web ("webmail”). Los usuarios pueden acceder a su correo electrónico desde cualquier dispositivo conectado a Internet en vez de desde una sola máquina. Correo Yahoo!, Hotmail y Gmail son algunos de los servicios de correo web más populares.

Con el aumento constante de la velocidad de conexión a Internet, la gama de servicios en la nube se ha multiplicado en los últimos años. Ahora, por ejemplo, podemos almacenar grandes cantidades de datos en la "nube” usando discos duros virtuales como el de Microsoft Live.

Igualmente, también está creciendo la oferta de software de ofimática en la nube, como procesadores de texto y bases de datos.



INFORMÁTICA EN LA NUBE

El proyecto de sistema operativo Google Chrome es un paso más en la evolución hacia la informática en la nube. El sistema será accesible desde el navegador web y planea incorporar por defecto tecnologías en la nube, lo que significa que los programas que necesitas tener instalados en el ordenador serán mínimos, con una fuerte dependencia de los servicios disponibles online. Esto es justo lo contrario del modelo tradicional, en el que los programas están instalados en el ordenador y la dependencia del software en la nube es baja o nula.

MEDIOS DE COMUNICACIÓN SOCIALES

DONDE NOS JUNTAMOS

Los medios de comunicación sociales son un conjunto de herramientas de comunicación online que permiten a sus usuarios crear y compartir contenidos.

Los medios sociales se diferencian de los medios de comunicación convencionales en que no solo informan sino que interactúan contigo mientras te dan la información.

La interacción puede ser algo tan sencillo como permitirte escribir comentarios o votar artículos o marcar como "Me gusta" o "Ya no me gusta" cualquier acción de otros usuarios. El usuario no es un mero espectador, sino que forma parte de los medios, ya que otros usuarios también pueden leer sus comentarios o reseñas.

La gente se está acostumbrando a poder responder a lo que otros escriben y a expresar su punto de vista. Esto hace que aumente la participación de la comunidad en los debates sobre temas de actualidad. El número de usuarios de medios sociales no para de crecer y con ello su influencia, lo que los hace cada vez más poderosos.

Cualquier página web que permita a sus visitantes interactuar con el sitio y con otros visitantes puede ser considerada un medio de comunicación social. En líneas generales, los medios sociales pueden dividirse en seis tipos:

1. Proyectos colaborativos (como Wikipedia), en donde los usuarios interactúan añadiendo artículos o editando artículos existentes;
2. Blogs y microblogs (como Twitter);
3. Comunidades de contenidos (como YouTube o Flickr), en donde los usuarios interactúan compartiendo y comentando fotos o vídeos;
4. Redes sociales (como Facebook, Myspace, Hi5, google+), en donde los usuarios interactúan añadiendo amigos, comentando perfiles, uniéndose a grupos y participando en debates;
5. Mundos de juego virtuales (como World of Warcraft);
6. Mundos sociales virtuales (como Second Life).

La protección de los usuarios de los medios sociales es otro tema importante, especialmente la protección de la privacidad. Aunque generalmente los usuarios pueden decidir si publican datos de carácter personal, la configuración predeterminada y la protección de los menores de edad no están exentas de polémica. Por si fuera poco, algunas páginas, como Facebook, han cambiado unilateralmente los ajustes de privacidad de sus usuarios en varias ocasiones.

GOBERNANZA DE INTERNET

DEMOCRACIA DIGITAL

Los primeros intentos de definir el término “gobernanza de internet" (internet governance) tuvieron lugar en las reuniones preparatorias de la Cumbre Mundial sobre la Sociedad de la información de las Naciones Unidas.

El Grupo de Trabajo sobre Gobernanza de Internet, creado por el Secretario General de la ONU e integrado por distintas partes interesadas, propuso la primera definición comúnmente aceptada, que fue incluida en el Programa de Acciones de Túnez para la Sociedad de la Información:

"preparación y aplicación por parte de los gobiernos, el sector privado y la sociedad civil, en el desempeño de sus respectivas funciones, de principios, normas, reglas, procedimientos de toma de decisiones y programas comunes que dan forma a la evolución y a la utilización de Internet.”

Esta definición pone énfasis en la participación de todos los actores de Internet de forma abierta, transparente y responsable.

Para lograr este objetivo, se creó el Internet Governance Forum (IGF), un foro en el que las partes interesadas debaten aspectos políticos relacionados con los factores clave de la gobernanza de Internet. El foro, que ya ha tenido seis ediciones (entre 2006 y 2011) originó la creación de foros similares a escala nacional y regional (como la EuroDIG, diálogo europeo para la gobernanza de Internet). Es importante señalar que estos foros no tienen poder de decisión pero sí influyen en las políticas.

¿Qué abarca la gobernanza de Internet?

- Infraestructura y estandarización;
- Temas técnicos relativos al funcionamiento de Internet: infraestructura de telecomunicaciones, estándares y servicios de Internet (Protocolo IP, DNS...), estándares para contenidos y aplicaciones (como el HTML);
- Temas relacionados con salvaguardar la seguridad y la estabilidad de Internet: seguridad, encriptación, spam;
- Temas legales: legislación y normativas nacionales e internacionales aplicables a temas relacionados con Internet (copyright, delitos informáticos, privacidad y protección de datos...);
- Temas económicos: comercio electrónico, fiscalidad, firma electrónica, pago electrónico;
- Temas de desarrollo: brecha digital, acceso universal a Internet;
- Temas socioculturales: derechos humanos (libertad de expresión, el derecho a buscar, recibir y divulgar información), política de contenido, privacidad y protección de datos, multilingüismo y diversidad cultural, educación, protección de los menores.

¿Quién participa en la gobernanza de Internet?

- Gobiernos: elaboran y aplican políticas públicas y regulaciones relativas a Internet;

- Sector privado: proveedores de servicios de Internet (ISP), proveedores de red, registradores de dominios, empresas de software, empresas de contenidos;
- Sociedad civil: organizaciones no gubernamentales que representan a los internautas;
- Organizaciones internacionales: Unión Internacional de Telecomunicaciones, la UNESCO, Programa de las Naciones Unidas para el Desarrollo (PNUD);
- Comunidad técnica: Internet Society, Internet Engineering Task Force, Internet Architecture Board, ICANN (Corporación de Internet para la Asignación de Nombres y Números).