Plataforma de Preferencias de Privacidad: Un caso práctico, el mundo

Manuel Cañón López

Octubre 2006

Resumen

Se pretende dar una visión global de la privacidad en la red: que es y porque es tan importante, centrándonos en el *standard* P3P [1]. Como ejercicio didáctico, se interpretará la política de privacidad de un sitio web real, la versión *online* del periódico el mundo.

1. Introducción

Lo primero que nos preguntamos es que entendemos por privacidad en el ámbito informático, nosotros proponemos una definición muy simple a la vez que descriptiva: el derecho a mantener en secreto nuestros datos personales y comunicaciones.

Este tema siempre ha tenido una importancia enorme en internet, por lo que ha sido objeto de gran debate puesto que que la falta de confianza en la privacidad de un sitio web comercial puede significar dejar de usar sus servicios y, por tanto, que este sitio deje de ganar dinero. Por otro lado, ganar la confianza del usuario puede significar ganar un número potencialmente elevado de clientes: no debemos olvidar que el usuario de a pie cada vez está más acostumbrado a las nuevas tecnologías, aunque no tenga claro lo que significan si está acostumbrado a lo que significan gracias a los artículos de divulgación [1]. Esto hace que pequeños detalles técnicos le puedan dar el empujón decisivo para decidirse por un portal o por otro.

No hace falta ir muy lejos para darnos cuenta que este problema es algo real, tan solo tenemos que preguntarnos cuantas veces no hemos dado nuestros datos personales reales por miedo a un posible uso fraudulento de estos. Y, si vamos un poco más allá, si se nos pide algún dato verdaderamente sensible (como podría ser el número de tarjeta de crédito) nuestro miedo crece exponencialmente. ¿Cuántos profesionales de la informática no compran por internet porque no se fían de que se hará con sus datos?. Un claro ejemplo de esto son las *cookies*, muchos usuarios las tienen desactivadas, pero ciertos sitios web requieren que estén activas. Suponer que el usuario va a saber hacer algo tan

[1] Platform for Privacy Preferences

simple se convierte en un arte adivinatorio, más aún cuando se está hablando de actividades comerciales, donde la pérdida de un potencial comprador tiene mucha importancia.

Unido a lo anterior tenemos la situación política actual o las nuevas leyes impulsadas por el gobierno americano, que nos hace preguntarnos que pasará si visitamos páginas que no son "políticamente correctas", si no tendremos que dar explicaciones a alguna autoridad. La vigilancia en internet puede significar para muchas personas ver sus derechos constitucionales pisoteados, y esto los convierte por tanto en otro grupo de personas interesados en la privacidad: en como sus datos personales son tratados.

Lo primero que debería un desarrollador preocupado por que pensará el usuario debe ser definir una política de privacidad, algo que haga que el usuario se sienta tranquilo con lo que se van a hacer con sus datos, tanto en el presente como en el futuro. Usualmente, el principal contenido de la política de privacidad es:

- Que información almacena el servidor
- Que uso se le dará a dicha información
- Cuanto tiempo se almacenará esta información

Y su acceso es a través de un link que nos muestra un texto más o menos largo que el usuario no se molestará en leer. Evidentemente esta solución no hará que el usuario medio tenga más confianza en el sitio, puesto que prácticamente nunca lo legará a leer. Curiosamente, el usuario suele ser desconfiado respecto al uso que se harán de sus datos. Esto hace que nos preguntemos que alternativas tendrá el usuario para mantener a salvo su privacidad sin que le suponga la molestia de tener que leerse un aburrido texto listando estas libertadas. Una solución a este problema se tratará en el siguiente punto.

2. El standard P3P

Una vez que observamos que definir una política de privacidad utilizando lenguaje natural y publicarla tal cual no suele ser suficiente para lograr la confianza del usuario, tenemos que preguntarnos que podemos hacer, ¿existe algún *standard* que permita al usuario controlar que se va a hacer con sus datos?. La respuesta a esta pregunta es sí: P3P, el cual es promovido por el W3C [2], oficialmente recomendado a partir de Abril del 2002 y que permite al usuario controlar la información personal en los sitios web que visitan.

Un sitio web que contemple la utilización del *standard* P3P declararía que tipo de información almacena y para que es almacenada: esto no es más que definir su política de privacidad. Lo que ocurre es que esta información es almacenada tal y como nos indica el *standard*, de forma que un navegador con

[2] World Wide Web Consortium

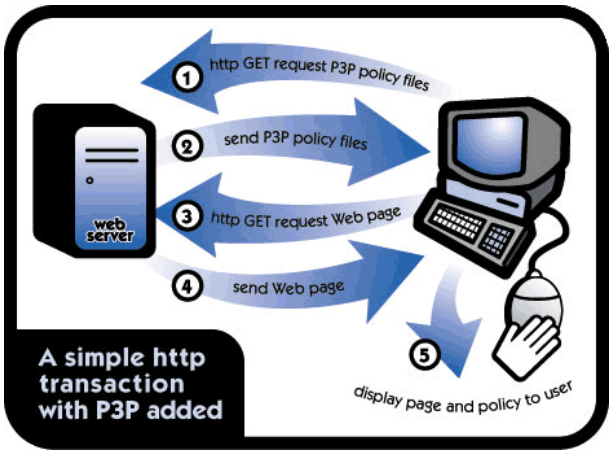

Figura 1: Funcionamiento básico P3P

la capacidad P3P activada puede decidir que hacer comparando la política del sitio web visitado con las preferencias del usuario. Así mismo, también puede mostrar una versión con un formato visual para que el usuario decida que hacer: si dar sus datos o no.

Toda la teoría anterior no tendría sentido si los clientes no soportaran este *standard*, siendo curiosamente *Microsoft Explorer 6* el primero en hacerlo pero de forma parcial: tan sólo permite gestionar las preferencias de privacidad con respecto a las *cookies*. No obstante, si un usuario deseara que su navegador examine la política de privacidad de un sitio web, podría instalar *privacy bird*[3], el cual no es más que una herramienta gratuita que añade esta característica al Internet Explorer. Pero no sólo Explorer nos proporciona la posibilidad de gestionar nuestras preferiencias, *Mozilla Firefox* a través la extensión P3P[4]. También está previsto incluirlo a corto plazo en la distribución habitual de *Firefox*. No obstante, todos tienen un funcionamiento similar, partimos de una configuración predeterminada que el usuario puede adaptar a su gusto respondiendo a preguntas del tipo que datos estaría el usuario dispuesto a compartir o que tipo de *cookies* quiere el usuario almacenar. A esto último es lo que se presta una atención especial, pudiendo restringir que tipo de *cookies* aceptar: sólo temporales, sólo que se refieran a ciertos datos...

Ahora trataremos de explicar la interacción básica entre un cliente y una aplicación que sigue el *standard* P3P apoyándonos en la figura anterior. Lo primero que hace el cliente es buscar y leer las políticas de privacidad de un sitio, tras ello el navegador compara las directivas de privacidad introducidas por el usuario con las que el sitio tiene: en caso de que sean compatibles, se haría una petición HTTP correspondiente y se mostraría la página solicitada; si las preferencias del usuario son distintas se preguntará a éste que desea hacer. De

[3] `http://www.privacybird.com/`
[4] `http://www.mozilla.org/projects/p3p/`

esta forma, el usuario podrá elegir si desea seguir o bien si desea cancelar la acción. Como consecuencia de esto, nos podríamos preguntar si el rendimiento de la aplicación disminuirá debido al aumento del número de peticiones: antes de acceder a una página debemos pedir su política de privacidad para posteriormente poder procesarla por el cliente. La respuesta es que no, que no aumenta realmente el tiempo de espera, puesto que aunque hay que hacer una petición ésta suele ser resuelta rápidamente llegando a calcularse que su coste es inferior al de traer una simple imagen de una página. La potencia de P3P es el grado de detalle al que podemos llegar, puesto que aparte de definir que datos queremos y cuales no dar, podemos definir que ciertos datos se darán de acuerdo al posterior uso que se le vaya a dar.

3. Funcionamiento a nivel técnico

Una vez explicado en que consiste P3P, trataremos ahora de explicar como se ha llevado a cabo su implementación, que tendríamos que hacer para que nuestro sitio web fuese *P3P compliant* [6]. Lo primero que haremos será decidir que política de privacidad tendrá nuestro sitio web, para posteriormente traducirla a un archivo XML (policy.xml).

De esta forma, tendremos que crear dos archivos: por un lado, un archivo **p3p.xml** que tendrá la misión de referenciar los distintos archivos con la política de privacidad del usuario); por otro un archivo **policy.xml** que contiene realmente la política de privacidad de nuestro sitio web. Actualmente, existe el convenio de que, al menos, el archivo p3p.xml esté colgando de un directorio llamado **w3c**. De esta forma podremos denominar como queramos a nuestro archivos con las preferencias de privacidad, puesto que será *p3p.xml* quien defina que archivos tocar.

Procedemos ahora a examinar la configuración P3P de un sitio real y correcto [8] (concretamente la versión *online* del periódico el mundo) reducida por razones de espacio, dicho ejemplo contiene comentarios que serán de utilidad para entender los distintos atributos y etiquetas utilizados [7]. Lo primero que examinaremos será su fichero **p3p.xml**:

Listado 1: p3p.xml

```
<META>
  <POLICY-REFERENCES>
    <!-- Indica, en segundos, cuanto tiempo es válida esta
         política
     desde que se recibió en el cliente: 24 h. El minimo-->
    <EXPIRY max-age="86400" />
    <POLICY-REF about="/w3c/general.xml#general">
      <!-- informacion general, se incluye todo por defecto
       exlucyendo el resto -->
      <INCLUDE>/*</INCLUDE>
        <!--
         excluimos aquello que require
```

```
        informacion adicional del usuario
        -->
      <EXCLUDE>/tiempo/*</EXCLUDE>
      <EXCLUDE>/perl/*</EXCLUDE>
      <EXCLUDE>/porras/*</EXCLUDE>
      <EXCLUDE>/debate/*</EXCLUDE>
      <EXCLUDE>/debates/*</EXCLUDE>
    </POLICY-REF>

    <POLICY-REF about="/w3c/cookies.xml#cookies">
      <INCLUDE>/medscape/*</INCLUDE>
      <!-- datos del tiempo - personalizacion -->
      <INCLUDE>/tiempo/*</INCLUDE>
      <!-- cgis que pueden depender de cookies -->
      <INCLUDE>/perl/*</INCLUDE>
      <INCLUDE>/debate/*</INCLUDE>
      <INCLUDE>/debates/*</INCLUDE>
      <INCLUDE>/charla/*</INCLUDE>
      <EXCLUDE>/porras/*</EXCLUDE>
      <!-- las cookies en si, solo será aplicable esta
          política
       a aquellas cookies que tengan como dominio elmundo.es --
          >
      <COOKIE-INCLUDE domain=".elmundo.es" name="*" value="*"
          path="*"/>
    </POLICY-REF>
  </POLICY-REFERENCES>
</META>
```

En este fichero se definen definen las políticas tanto general como para *cookies* a través de la etiqueta policy-ref. Si nos fijamos, se hace referencia a unos ficheros (general y *cookies*) que tendrán las políticas en sí para las *cookies* y para el sitio. En este fichero, se asocia que recursos están incluidos y cuales excluidos de la política de la que descienden jerárquicamente a través de las etiquetas *include* y la etiqueta *exclude*. De esta forma, podemos ver que las políticas generales no cubren las porras, los debates o el tiempo, en cambio, estas si que son cubiertas por las políticas de *cookies*.

Ahora veremos uno de estos archivos de privacidad, no veremos el resto por cuestiones de espacio. Al igual que en el otro ejemplo, recomendamos leer los comentarios detenidamente, pues en ellos se irá explicando parte del formato:

Listado 2: general.xml

```
<POLICIES>
    <POLICY name="general" discuri="http://www.elmundo.es/
        privacidad/">
      <ENTITY>
        <!-- Datos acerca del elemento que ofrece la politica
            -->
        <DATA-GROUP>
```

```
        <DATA ref="#business.name">
          Mundinteractivos - El Mundo</DATA>
          <DATA ref="#business.contact-info.postal.stateprov">
 9        Madrid</DATA>
          <DATA ref="#business.contact-info.postal.postalcode"
              >
          28002</DATA>
          <DATA ref="#business.contact-info.online.email">
          webmaster@el-mundo.net</DATA>
14        <DATA ref="#business.contact-info.telecom.telephone.
              number">
          915864800</DATA>
        </DATA-GROUP>
      </ENTITY>
      <!-- Especificamos si el sitio
19    proporciona acceso a varios sitios de informacion-->
      <ACCESS>
        <!-- En este caso vemos que el mundo no recoge
        información proporcionada
        a través del protocolo HTTP-->
24      <nonident/>
      </ACCESS>
      <!-- Indica que procedimientos debemos seguir si
          tenemos alguna duda con la politica
          de privacidad adoptada por el sitio web-->
29    <DISPUTES-GROUP>
        <!-- A través del atributo resolution-type vemos que
            podemos
            hacer si las politicas de privacidad de rompen.
            en este caso podemos quejarnos a la propia web-->
        <DISPUTES resolution-type="service" service="http://
            www.elmundo.es/">
34        <!--Aqui podemos indicar que posibles remedios hay
              cuando
          se rompe la politica de privacidad -->
          <REMEDIES>
            <!-- El propio servicio se encarga de errores o
                acciones incorrectas -->
            <correct/>
39          <!--Rupturas de la política de privacidad
            se determinarán segun la propia ley -->
            <law/>
          </REMEDIES>
        </DISPUTES>
44    </DISPUTES-GROUP>
    <STATEMENT>
      <!-- Para que se utilizan los datos procesados por la
          web -->
      <PURPOSE>
        <admin/>          <!-- Administracion -->
```

```
49          <develop/> <!-- Seguir con el desarrollo -->
            <current/>
            <!-- Completar la actividad para la que se
                introdujeron los datos -->
          </PURPOSE>
          <!-- Quien utilizara la informacion -->
54        <RECIPIENT>
            <ours/> <!-- El propio servicio -->
          </RECIPIENT>
          <!-- Cuando tiempo se retendrá esta información -->
          <RETENTION>
59          <!-- Sin fecha de eliminacion -->
            <indefinitely/>
          </RETENTION>
          <!-- Describe que datos van a ser transferidos o
              inferidos -->
          <DATA-GROUP>
64          <!-- Se guardaran los típicos datos que se encuentran
                en el log del servidor: ip, host
            del cliente, la URI del recurso pedido.., el navegador
                ,
            de donde viene el cliente,
            con que cadena de busqueda y las cookies -->
            <DATA ref="#dynamic.clickstream"/>
69          <DATA ref="#dynamic.http.useragent"/>
            <DATA ref="#dynamic.http.referer"/>
            <DATA ref="#dynamic.searchtext"/>
            <DATA ref="#dynamic.cookies" optional="yes">
                    <!-- Esta marca pretende dar al agente
                        información sobre
74                   el uso que se dara a los data padres--!>
              <CATEGORIES>
                    <!-- En este caso es un identificador unico,
                        de ámbito no financiero
                    con la única intención de reconocer al
                        individuo. Esto incluye identificadores
                    que nos pueden dar tras registrarnos-->
79                <uniqueid/>
              </CATEGORIES>
            </DATA>
          </DATA-GROUP>
        </STATEMENT>
84    </POLICY>
    </POLICIES>
```

Examinando este fichero, podemos ver como este sitio no recoge ninguna información proporcionada por el protocolo utilizado por el cliente (esto viene definido por el elemento ACCESS), y que si se produce alguna ruptura en la política de privacidad el propio servicio del *mundo.es* se encargará de solucionarlo de acuerdo con las interpretación de las leyes adecuadas (se define a través

del elemento DISPUTES-GROUP). Por otro lado, también sabemos la información recogida se utilizará para mejorar la administración, desarrollo del sitio o completar los procesos para los que se introdujeron (etiqueta PURPOSE) y que será leída tan solo por los propios trabajadores del mundo (etiqueta RECIPENT). También tenemos que tener claro que esta información será almacenada indefinidamente (etiqueta RETENTION). Por último, podemos saber a través de la etiqueta DATA-GROUP que se almacena del usuario: que navegador usa, desde donde vino, que cadena de búsqueda utilizó o donde hizo *click*.

Por supuesto, hoy en día no se crean estos ficheros manualmente, sino que disponemos de herramientas que hacen esto sea más fácil, concretamente el W3C nos recomienda el editor de IBM (el cual no parece estar mantenido muy frecuentemente), *privacyBot.com*, *IAJapan* o *P3PEdit*, los cuales son curiosamente comerciales. Su funcionamiento básico es similar a otras herramientas de generación: se van haciendo una serie de preguntas en lenguaje natural y finalmente las traduce a los recursos ya explicados.

4. Conclusiones

Como se ha comentado al inicio del artículo, la privacidad es un tema que interesa a todo internauta, nadie quiere ver sus datos personales comprometidos. Pero, curiosamente, nunca se tiene tiempo para leer las políticas de privacidad. Tal y como hemos visto, con una simple ojeada a un fichero P3P real, se puede codificar toda la información de la política de privacidad de un sitio web cómodamente; sabiendo que esta información puede ser procesada de forma fácil por el navegador, podemos contar con que nuestro usuario llegue a tener constancia que nuestra política de privacidad se adapta a su gusto sin que esto le suponga ninguna atención extra. Esto debería ser un elemento diferenciador frente a otros portales que no ofrezcan una política de privacidad automatizada.

Curiosamente, a día de hoy, casi nadie tiene activadas las preferencias de privacidad o en su defecto, instalado el *plugin* que permita gestionarlo. Esto hace que, iniciativas como la estudiada, no tengan el impacto que deberían. No obstante, desde aquí creemos que esta es una tecnología emergente que poco a poco irá siendo más conocida y, por tanto, utilizada.

El futuro directo de P3P va ligado a dos proyectos nuevos, por un lado la investigación de unir privacidad con la gestión de la identidad a través del proyecto PRIME[5]. Este proyecto pretende lograr un prototipo plenamente funcional, que nos permita gestionar automáticamente distintos roles de usuario, y asociados a ellos distinta información privada. Otro punto muy interesante de este proyecto consiste en el estudio de las obligaciones que adquiere el *back-end*, que obligaciones tiene quien ofrece el servicio y por tanto su política de privacidad.

Por otro lado, se está investigando en otro proyecto para la creación de los fundamentos técnicos y legales para definir claramente la transparencia y responsabilidad para la recuperación de datos para sistemas heterogéneos de

[5]Privacy and Identity Management for Europe, `https://www.prime-project.eu/`

gran tamaño en un proyecto llamado TAMI[6]. La idea existente tras este proyecto es desarrollar lenguajes de reglas que permitan expresar o deducir que hacen exactamente a través de todos los procesos de recuperación de información. La incorporación de transparencia y responsabilidad se muestra crítico para ayudar a la sociedad a administrar los riesgos provenientes de la explosión en las comunicaciones así como almacenamiento y búsqueda de la información.

Una última vía de investigación, mucho más utópica, sería lograr una gestión semántica de la información a través de la incorporación de metadatos así como una forma efectiva de de expresar las obligaciones con el acuerdo de gestión de la privacidad.

[6]Transparent Accountable Datamining Initiative, `http://dig.csail.mit.edu/TAMI/`

Referencias

[1] **P3P ofrece intimidad a medida**, Artículo en el periódico del mundo http://www.elmundo.es/navegante/98/mayo/25/ecomerciop3p.html

[2] **Privacidad**, Artículo en la wikipedia sobre la privacidad http://es.wikipedia.org/wiki/Privacidad

[3] **Guía Breve de Privacidad y P3P**, Artículo introductorio del W3C sobre la privacidad http://www.w3c.es/Divulgacion/Guiasbreves/PrivacidadP3P

[4] **P3P**, Artículo en la wikipedia sobre la P3P http://en.wikipedia.org/wiki/P3P

[5] **Platform for Privacy Preferences (P3P) Project**, Página web oficial del proyecto P3P http://www.w3.org/P3P/

[6] **El reto de la P3P, Amenaza u Oportunidad ?**, Revista digital Click and tips http://www.clickandtips.com/articulos/ct0039.htm

[7] **The Platform for Privacy Preferences 1.0 (P3P1.0) Specification**, Publicación onlice del W3C http://www.w3.org/TR/P3P

[8] **P3P W3C Validator**, Validador online de ficheros de configuracion P3P http://www.w3.org/P3P/validator.html