Fuga de datos por rastreo publicitario

Enseñanzas del análisis del sitio web “Tullave” y de la aplicación de la DIAN

Fundación Karisma <K+LAB>

Bogotá, Colombia 2019

En un esfuerzo para que todas las personas tengan acceso al conocimiento, Fundación Karisma está trabajando para que sus documentos sean accesibles. Esto quiere decir que su formato incluye metadatos y otros elementos que lo hacen compatible con herramientas como lectores de pantallas o pantalla braille. El propósito del diseño accesible es que todas las personas puedan leer, incluidas aquellas que tienen algún tipo de discapacidad visual o de dificultad para la lectura y comprensión.

Más información sobre documentos accesibles en:
http://www.documentoaccesible.com/#que-es

Fundación Karisma <K+LAB>

Autor: Stéphane Labarthe
Revisión: Pilar Saenz
Diagramación: Daniela Moreno
Coordinación editorial: Diego Mora Bello

Contenido

<K+LAB>

K+LAB es el Laboratorio de “seguridad digital y privacidad” de la Fundación Karisma. Como lo dice su título, en el K+LAB se unen dos temas que frecuentemente van separados porque creemos que en realidad están muy conectados. Así por ejemplo el rastreo publicitario usando cookies y otros mecanismos similares[1] suele asociarse con riesgos de privacidad pero es menos frecuente que se hable de los riesgos que pueden generar su utilización para la seguridad de la información. De hecho las multinacionales de marketing digital han hecho grandes esfuerzos para difundir la idea y la narrativa de que el rastreo publicitario que hacen en la web para darnos “ofertas adaptadas” y “mejorar nuestra experiencia” es un “rastreo anónimo” que poco impacta nuestra vida privada y que no afecta la seguridad de nuestros datos.

En un informe de SHARE LAB[2], se afirmaba que: “Es importante notar que los TPC [contenidos de terceros][3] pueden solo acceder a los metadatos que por defecto son una categoría de datos pública.”

Pensamos al contrario que la privacidad y la seguridad digital están íntimamente vinculadas. De hecho, en el caso del rastreo publicitario, los “terceros” pueden acceder a mucho más que metadatos e incluso ser la causa de fugas de datos personales.

---

1 Las cookies HTTP son informaciones que se pueden inscribir o leer a distancia en el terminal del usuario. Permiten al sitio y a los terceros mencionados proveer funcionalidades técnicas pero también pueden ser un herramienta para rastrear en su navegación en Internet. Existen también otras técnicas de rastreo cómo las cookies flash (local share objects), el finger printing o el local storage (HTLM5).

2 Traducción nuestra del texto siguiente: “Now, it is important to note that TPC can only access metadata, which by default is a somewhat public category of data.”, Invisible Infrastructures: Mobile permissions, SHARE LAB, marzo 2015, https://labs.rs/en/invisible-infrastructures-mobile-permissions/

3 Los “TPC” se refieren a los third party contents, contenidos de terceros, como empresas de publicidad, analítica de datos, botones de redes sociales, vídeos externos como los de YouTube, servicios de CHAT, etc. que se incluyen o interactúan con los sitios Internet, las redes sociales y las aplicaciones de smartphones/tabletas. Técnicamente se suelen materializar en una porción de código fuente externo que se “trasplanta” en el código original del sitio o de la aplicación.

El acceso a datos personales derivado del uso de redes sociales cobró vigencia de forma dramática en los últimos años, en particular debido al escándalo de Cambridge Analytica con el que se evidenció como se obtuvo un acceso a datos personales detallados de más de 87 millones de usuarios de la red social Facebook[4] vía la aplicación de terceros GSRApp (a veces llamada "thisisyourdigitallife"). Esta información fue utilizada para generar perfiles de votantes influenciables en la elección de Donald Trump. Esto -entre otras cosas- condujo a la Federal Trade Commission de Estados Unidos (FTC) a demandar a Cambridge Analítica y a poner la histórica multa de 5 billones de dólares a Facebook este mes de julio[5]. Más recientemente, la empresa Twitter tuvo que disculparse porque fueron compartidos datos personales con empresas de analítica de datos y de publicidad sin el consentimiento de sus usuarios[6]. Estos incidentes han demostrado que incluso las grandes empresas de redes sociales no siempre tienen control de los datos personales ni de lo que comparten con terceros.

Sin embargo, estos problemas no se limitan a las redes sociales. Son más amplios y este artículo apunta a mostrar cómo y porqué la inclusión incontrolada y sin precauciones de servicios de terceros genera riesgos de privacidad, de seguridad digital e incluso de fuga de datos personales que se extienden también a las aplicaciones para smartphone/tabletas y a los sitio web. Es importante precisar el sentido que le damos a "fuga de datos". Se trata de una transmisión de datos - en los casos analizados personales - a un tercero que no está autorizado para acceder a ellos y recibirlos. Por lo tanto se aplica pero no se limita a una publicación de estos datos en Internet.

Esta publicación contiene argumentos teóricos pero se apoya principalmente en dos análisis recientes que hizo "K+LAB": el primero a la aplicación de la DIAN y el segundo al sitio web de "Tullave", la tarjeta de transporte público de Bogotá. Estos hallazgos fueron presentados en la Conferencia RightsCon 2019 en Tunéz[7].

---

4 Según un comunicado de Facebook de abril 2018: "In total, we believe the Facebook information of up to 87 million people — mostly in the US — may have been improperly shared with Cambridge Analytica.", ver comunicado completo de la empresa aquí: https://newsroom.fb.com/news/2018/04/restricting-data-access/

5 Para detalles, se pueden leer las dos publicaciones de la FTC a respecto: https://www.ftc.gov/news-events/press-releases/2019/07/ftc-sues-cambridge-analytica-settles-former-ceo-app-developer y https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions

6 Se puede leer el comunicado aquí en el sitio de la empresa: https://help.twitter.com/en/ads-settings

7 Detecting personal dataleaks to third party actors, RightsCon 2019, https://rightscon2019.sched.com/event/Pvwj/tech-demos-blowing-the-whistle-data-leaks-and-digital-dropboxes

Los análisis de sitios webs y aplicaciones del Gobierno: del IMEI al Transmilenio

Desde hace casi tres años K+LAB intenta generar conciencia y conocimiento sobre privacidad y seguridad digital a través del análisis de sitios webs y de aplicaciones del Gobierno[8]. Hemos analizado los sitios de IMEI Colombia, la Unidad de víctimas, los sitios de los cinco principales candidatos para la presidencia de la República de Colombia en 2018, el sitio y la aplicación de la DIAN y recientemente el sitio de TullavePlus, la tarjeta del sistema de transporte público de Bogotá[9]. Estos análisis se han hecho con el fin de contribuir a mejorar la información, la seguridad digital y la privacidad de éstos sitios y aplicaciones, y para el beneficio de la ciudadanía y de las entidades responsables. Este tipo de ejercicios han demostrado ser valiosos y han llevado a mejoras importantes[10]. También cabe mencionar que estos análisis se hacen de una forma técnica pero no intrusiva, siempre buscando quedarse en un marco legal y ético. Por ejemplo se analizan los flujos de datos que salen y entran de nuestro equipo (con herramientas de código abierto[11]) pero nunca se hacen pruebas de intrusión (pentesting) en los servidores.

A través de la socialización de estos análisis, hemos intentado demostrar que una mala gestión de la privacidad puede generar problemas de seguridad digital e incluso fugas de datos personales. En la Conferencia RightsCon 2018 en Toronto, presentamos el análisis de la aplicación para smartphones y tabletas de la DIAN. En esta ocasión, pusimos en evidencia que una mala implementación de servicios de terceros en esta aplicación había generado fugas de todos los datos personales que los usuarios ingresaban en el formulario de inicio de chat (nombre, apellido, cédula, telefono, correo y dirección). Estos datos se iban hacia un tercero, Google, que facilitaba un API de interconexión con sus servicios[12]. En este caso, Google no estaba autorizada ni tenía necesidad de recibir estos datos personales destinados únicamente a la DIAN y a la empresa de CHAT contratada para ello. En la imagen siguiente, extraída de nuestra presentación en RightsCon, se ve en la parte izquierda los datos entregados en el formulario de la aplicación y en la parte derecha cómo estos datos se encuentran también en el flujo enviado por nuestro teléfono a los servidores de Google:

---

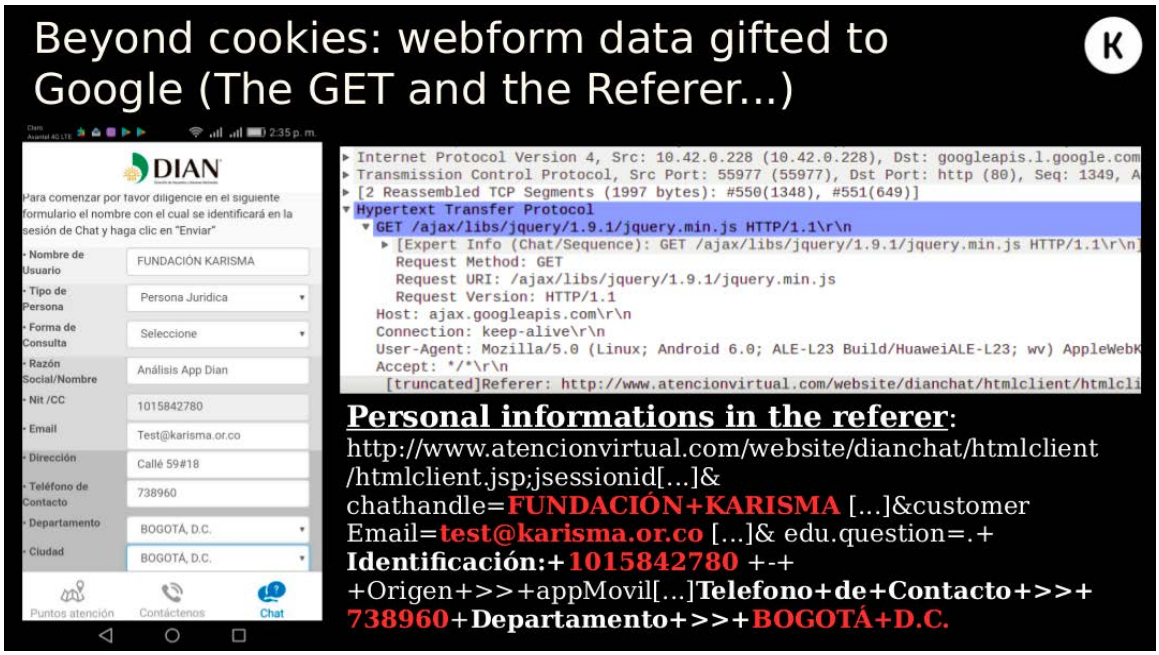
8 https://stats.karisma.org.co/klab-un-espacio-abierto/

9 https://www.tullaveplus.com/

10 Ver por ejemplo aquí: https://stats.karisma.org.co/analisis-de-imeicolombia-com-co-cronologia-de-un-dialogo-con-el-gobierno/ y aquí: https://karisma.org.co/la-corresponsabilidad-en-accion/

11 Para estos análisis, usamos el programa Wireshark (Mozilla Public License, version 2.0, https://www.wireshark.org/) y la extensión de navegador LiveHTTP Headers (Licencia pública GNU, versión 3.0, http://livehttpheaders.mozdev.org/index.html) con el navegador Waterfox (https://www.waterfox.net/). Pero son sólo posibilidades y existen otras alternativas.

12 La aplicación de la DIAN usaba el servicio Google Maps para localizar puntos de la DIAN.



Más adelante explicaremos en detalle como puede suceder esto.

En el último análisis que realizamos, el del sitio web “Tullave” [7], encontramos hechos similares.

Todavía no haremos público el informe del análisis de este sitio dado que algunos cambios aún están por realizarse[13] pero nos pareció importante hacer una publicación respecto a un problema que hubiera podido transformarse en una fuga de datos que incluía contraseñas y que ilustra muy bien el tema de este artículo. Afortunadamente, este problema se corrigió poco después de la presentación privada de nuestro análisis a los responsables de este sitio web. Para entender cómo se podía generar la fuga de datos, hace falta una explicación inicial.

13 En meses pasados entregamos a las personas responsables del tema en el Distrito de Bogotá y al Ministerio TIC los resultados completos de nuestro análisis del sitio web “Tu Llave”. Tras hacer el seguimiento nos indicaron que están trabajando en un plan de mejoras para resolver los hallazgos identificados en el informe. En consideración con su necesidad de más tiempo antes de nuestra publicación hemos pospuesto la divulgación del informe completo para después.

En el principio existían los sitios web, después llegaron los terceros

Cuando visitamos una página web o usamos una aplicación, nos conectamos por lo general[14] con el servidor web del sitio (que en realidad es un computador) que nos entrega el contenido solicitado. Al inicio de internet, los sitios webs sólo tenían contenidos propios y entonces la conexión se establecía sólo entre el computador de quien solicitaba la página y el servidor correspondiente. Pero poco a poco aparecieron espacios y servicios dentro de los sitios webs que no dependen directamente de ese servidor al que se está consultando sino que son externos, entregados por un tercero. Hoy la gran mayoría de los sitios webs y de las aplicaciones usan contenidos y funcionalidades externas, por ejemplo, contenidos publicitarios, vídeos, contenidos de redes sociales o las estadísticas de las visitas y el uso del sitio. Muchas veces estos contenidos y funcionalidades de terceros son gratuitas y las empresas que los desarrollan los ofrecen a cambio de los datos personales de los usuarios del sitio o de la aplicación. Esta información en general es utilizada posteriormente con fines publicitarios[15], así el pago por su uso no se hace con dinero sino con datos. La recolección de datos personales que hacen estas empresas directamente o a través de otros servicios les permiten entregar publicidad dirigida, más adaptada y eficiente, y por lo tanto más costosa. Muchas veces el responsable del sitio web o de la aplicación no es consciente de las implicaciones del uso de estos servicios de terceros o no mide los impactos que se pueden generar por su utilización, sólo ve el beneficio de una funcionalidad gratuita o la recompensa monetaria a cambio del alquiler de un espacio publicitario. Es algo que comprobamos cada vez que entregamos nuestros informes, los hallazgos suelen generar sorpresa.

Cuando nos conectamos a un sitio web o usamos una aplicación nuestro computador se conecta e intercambia información no sólo con el servidor web del sitio o de la aplicación sino - en la gran mayoría de los casos - con varios otros terceros que ofrecen sus servicios y también con aliados de estos terceros[16]. Si pensamos en una analogía con una casa, no solo se le abre la puerta a los amigos, sino a los amigos de los amigos, personas que no conoces y que terminan entrando en la casa libremente.

---

14 Con excepción de las aplicaciones fuera de línea que no necesitan conectarse al Internet.

15 Es por ejemplo el caso del servicio de vídeos Youtube o del servicio de analítica Google Analytics que permiten a Google captar datos de los visitantes para usos publicitarios ulteriores.

16 Es el caso por ejemplo cuando el sitio propone una publicidad digital basada en los modelos de subastas en tiempo real (RTB, Real Time Bidding, https://en.wikipedia.org/wiki/Real-time_bidding). En un caso así, el editor del sitio web inserta en su código fuente el de la plataforma de subastas, la cual transmite informaciones a varios candidatos a la impresión de la publicidad que son desconocidos del editor.

Para que esto suceda técnicamente el que gestiona el sitio web o la aplicación tiene que abrir a estos terceros una “puerta de entrada”. Esto se hace generalmente mediante la inclusión en el código fuente (HTML) de la página web o en el código fuente de la aplicación, de porciones de códigos externos entregados por estos terceros [13]. En los sitios webs, estos códigos de terceros se pueden observar al mirar el código fuente de la página web (CTRL+”U” en Firefox o Waterfox). Por ejemplo este es un extracto del código fuente que permite la inserción del botón Twitter tal cómo aparece en la página del sitio web Tullave:

```
<script type=”text/javascript” src=”http://platform.twitter.com/widgets.js”></script>
```

El código llama una función javascript externa de la plataforma de Twitter llamada “widget.js”. Es importante entender que el código de esta función, escrito en lenguaje javascript, puede ser modificado por Twitter dinámicamente y en cualquier momento sin que el editor del sitio se dé cuenta. El código se puede consultar en la URL http://platform.twitter.com/widgets.js pero es muy largo y difícil de entender. Generalmente este tipo de códigos suele ser ofuscado para hacer su entendimiento difícil para un humano, lo que dificulta su análisis.

Nuestra metodología, no analiza directa o detalladamente el código sino el efecto de su ejecución a través el flujo de datos generado[17]y la instalación de cookies[18]. Éstas tendrán un impacto en la privacidad del usuario mientras permanezcan en su computador lo que a veces puede durar semanas, meses o incluso años.

Hasta el momento hemos hablado de la privacidad, pero para analizar lo que implica el uso de servicios y funcionalidades de terceros en materia de seguridad de la información debemos ir un poco más profundo para entender quién tiene el control de estos códigos y cómo se transmite la información.

Sigamos con analogía de la casa. Es contraintuitivo porque cuando hablamos de visitar imaginamos que vamos a otro lugar y por tanto que visitar un sitio web debe ser ir a ese sitio pero en realidad cuando “visitamos” una página web, estamos recibiendo la visita en nuestra casa. Abrimos la puerta de nuestro dispositivo al sitio web que visitamos por medio del navegador. En el caso de la DIAN por ejemplo, podemos imaginar que invitamos empleados de la DIAN en nuestro salón para una reunión privada. Pero cuando estos empleados de la DIAN llegan, entran en la

---

17 En particular, buscamos en los flujos de datos entrando y saliendo de nuestra computador y capturados con herramientas como Live HTTP Headers o Wireshark, los datos personales entrados en los formularios.

18 Ibid.

casa y en el salon con amigos desconocidos e incluso amigos de estos amigos. Aunque ellos no tienen porque escuchar nuestra conversación privada y las respuestas personales que les damos a la DIAN, el riesgo que lo hagan es alto porque están en nuestra casa y en el salón. Tienen por lo tanto muchas posibilidades para hacerlo si lo quieren e incluso sin quererlo pueden oír pedazos de conversaciones privadas. Además para seguir la analogía hay que entender que la intención real de la mayoría de estos invitados sorpresas que se han juntado a la reunión es observar la casa y sus dueños para conocerlos mejor y luego ofrecerles productos a través de publicidad o vender la información de sus intereses para que otros les ofrezcan productos.

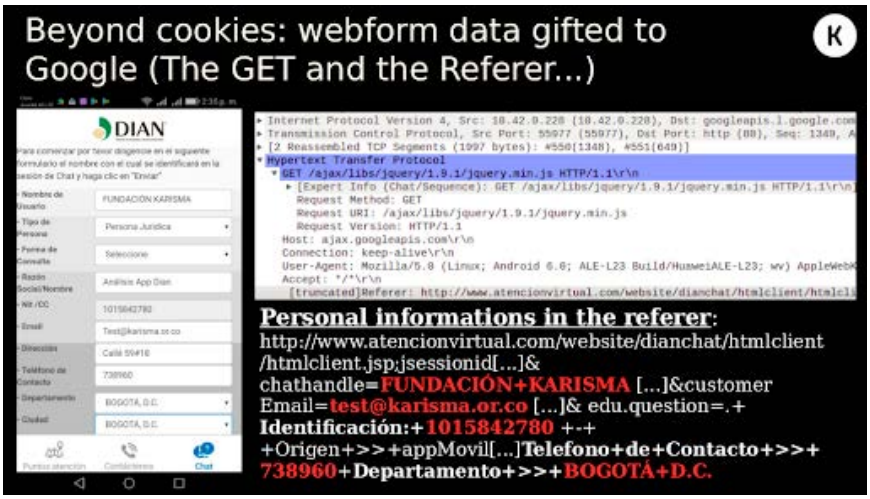
Ahora volvamos a una explicación más técnica. ¿Porque la DIAN viene a nuestra casa (con los amigos) y no al contrario? Cuando nuestro navegador se conecta al servidor web de la DIAN descarga un código (que incluye pedazos de código de terceros) que nuestro navegador interpreta y ejecuta en nuestro dispositivo, con todas las posibilidades vinculadas con los protocolos y los lenguajes de programación involucrados (HTTP(S), HTML y Javascript en particular). Las aplicaciones para los teléfonos inteligentes funcionan de la misma manera, su código se ejecuta en nuestro teléfono. Esta ejecución genera además flujos de datos en ambos sentidos: de mi dispositivo hacia el servidor web de la DIAN y los servidores de los amigos y en el sentido contrario también.

Un ejemplo técnico del invitado que escucha una conversación sin quererlo es cuando una información se transmite al servidor enviada en la URL (esto sucede cuando se envía la información con el método GET en vez del método POST[19]). Esto no es obvio pero este método de envío de información genera adicionalmente una transmisión de estos datos, que están visibles en la URL, a todos los terceros presentes en la página o en la aplicación, que están enviando solicitudes a sus servidores respectivos. Es una consecuencia de la presencia de un metadato, el Referer, en la cabecera del paquete HTTP[20]. Esta información es la que se ve en la imagen de la presentación que hicimos sobre el caso de la DIAN. Se ve una solicitud hacia la url externa “googleapis.l.google.com” en la cual se enviá, por el medio del Referer, todos los datos personales del formulario. En resumen, la combinación de una mala

19 El protocolo HTTP usa varios métodos para hacer sus peticiones, entre los cuales están GET y POST. La diferencia entre los métodos GET y POST radica en la forma de enviar los datos a la página cuando se pulsa el botón “Enviar”. Mientras que el método GET envía los datos usando la URL, el método POST los envía de forma que no podemos verlos, “dentro de paquete HTTP” o sea “ocultos” para la persona que usa la página. El método POST es el que está recomendado cuando se trata de ofrecer una mejor seguridad. Su uso es complementario del uso del protocolo cifrado HTTPS.

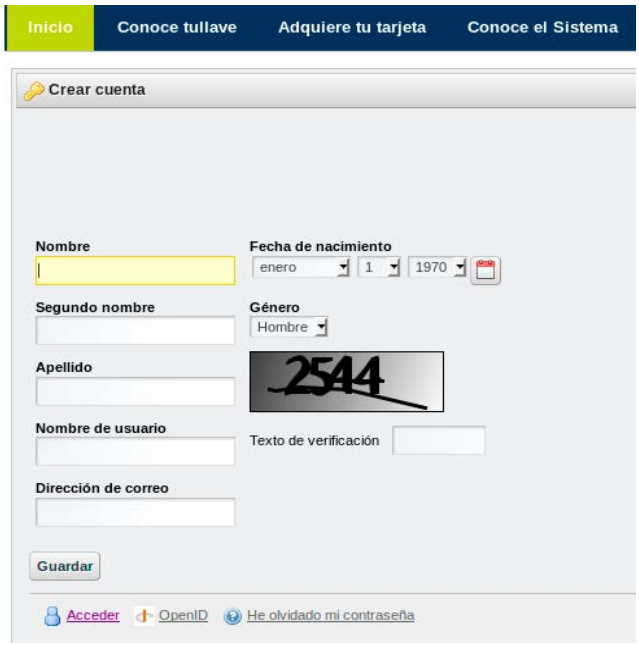
20 El Referer es una de las partes de la cabecera del protocolo HTTP que indica la dirección URL de donde proviene la petición. En el caso de la URL principal de la página es la página anterior. En el caso de solicitudes hacia servidores terceros que se originen desde una página web, el Referer es la URL de esta página. Por lo tanto, si esta URL contiene datos personales, se encontrarán también en este Referer.

práctica (el envío de los datos del formulario dentro de la URL) y del uso del servicio de tercero de Google API[21] provoca esta fuga de datos hacia la empresa Google.



Un formulario web que enviaba contraseñas: cómo se evitó una catástrofe

Cuando hicimos el análisis del sitio web de la tarjeta Tullave, del sistema de transporte de Bogotá (www.tullaveplus.com), encontramos páginas con formularios que permitían crear una cuenta, acceder y cambiar contraseña. Aunque eran indexados por Google, es decir, aparecían cuando se hacía una búsqueda sobre los sitios asociados a tullave, estos formularios parecían ser de un espacio privado de la empresa Recaudo Bogotá, encargada del sistema[22], y no hubiera tenido que estar disponible para nosotros. Encontramos por ejemplo este formulario de creación de cuenta.



21 Google API es un conjunto de interfaces de programación de aplicaciones (Application Programming Interface de su sigla en inglés) desarrollados por Google y que permiten comunicaciones con servicios de Google (como maps, mail, etc.) y su integración en otros servicios. En el caso de la aplicación de la DIAN se usaba el servicio de mapas de Google.

22 La empresa Recaudo Bogotá es concesionario del SIRCI (Sistema Integrado de Recaudo, Control, Información y Servicio al Usuario) y encargada del sitio web www.tullaveplus.com

Cuando hicimos el análisis técnico de estos formularios, simulando envíos de datos y capturando los flujos de datos generados (entrantes y salientes de nuestro equipo[23]), descubrimos que además de usar el protocolo inseguro HTTP estos formularios generaban fugas de datos vinculadas con la presencia de terceros en la página. Primero, en algunos casos el correo electrónico de estos formularios (que sirve de nombre de usuario para el ingreso) se enviaba a terceros que no debían recibirlos: Google, Facebook, Twitter, ShareThis y la empresa que provee el servicio de chat para esta plataforma Emtelco. Esto se debía al uso del método GET en el envío de esta información en la URL[24]. En un caso como éste, el uso del protocolo cifrado HTTPS no hubiera ofrecido protección contra este riesgo. Había también que usar el método POST para enviar los datos personales.

Más grave, el análisis mostró también que en otro formulario de conexión, se enviaban conjuntamente el login y la contraseña a la empresa ShareThis, una empresa estadounidense que propone un servicio para compartir contenidos en las redes sociales y que también de paso capta datos para comercializarlos como lo explicita su política de privacidad[25].A pesar de que no se había cometido el mismo error y los datos se enviaban sólo con el método POST, el entrar y enviar el login y la contraseña a través del formulario generaba el envío a ShareThis:


```
http://wd.sharethis.com/api/sharer.php?fpc=d4ba-
de4-165e4f96748-54b9806e-7&sessionID=1537145059104.71162&host-
name=www.tullaveplus.com& [...] &_58_struts_action=/login/
login&_58_login=test@karisma.org.co#sthash.A121jSBx.dpuf&sharUR-
L=&buttonType=custom&destination=copy&source=copy&description=Z4f@
LtN9r?V&ts=1537145110755   [21]
```

---

23 Esta parte del análisis se hizo con el navegador Waterfox y la extensión LiveHTTP Headers.

24 Por ejemplo, en el caso de Facebook, aquí está la transmisión del correo electrónico vía el Referer:

> GET/plugins/like.php?href=$request.attributes.CURRENT_COMPLETE_URL&layout=standard&show_faces=-true&width=450&action=like&colorscheme=light&height=80HTTP/1.1
> Host:www.facebook.com [...]
> Referer:http://www.tullaveplus.com/web/public/inicio?p_p_id=58&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&saveLastPath=0&_58_struts_action=%2Flogin%2Flogin&_58_login=test%40karisma.org.co

25 En la parte privacidad del sitio web de la empresa (https://sharethis.com/es/privacy/) se puede leer lo siguiente: "ShareThis también recopila datos sobre los usuarios de Internet y cómo interactúan con el contenido, los sitios web y los anuncios, lo que permite ShareThis y nuestros editores, anunciantes, clientes y socios de datos para facilitar la entrega de publicidad relevante y dirigida en línea a estos grupos."

Sin embargo, tenemos que reconocer que no estuvimos en capacidad de analizar en detalle el script de ShareThis, es demasiado largo y complejo para entenderlo. Sólo pudimos constatar su efecto: captar datos de un formulario, en este caso el login y la contraseña.

Antes de ir más adelante, cabe mencionar que poco después de recibir nuestro informe el Distrito pidió la desactivación de esta funcionalidad, neutralizando así este riesgo. Además nos contestaron lo siguiente:

> "Se validó el formulario y este fue un formulario que inicialmente se planteaba para una aplicación de tarjetas que no se dio uso correspondiente. Sin embargo, aunque el proyecto está activo y se desea poner en funcionamiento en el mediano plazo, se deshabilitó el botón de crear cuenta y se eliminará el acceso a la autenticación OpenID. [...] Este ítem ya no aplica dado que el formulario no se estaba utilizando."

Entonces, parece que se evitó una fuga de contraseñas que hubiera podido ser muy grave. Además del hecho de que ShareThis no era un tercero autorizado para recibir estos correos y contraseñas, dos elementos hubieran podido agravar las consecuencias:

1. ShareThis precisa en sus condiciones que puede compartir los datos colectados con sus "editores, anunciantes, clientes y socios de datos[26]", pero no se sabe cómo ni con quién se comparten;

2. En Julio del 2018 ShareThis sufrió un incidente de seguridad grave en el cual se robaron datos que la empresa poseía sobre más de 41 millones de usuarios, incluyendo correos y en ciertos casos fechas de nacimiento y hash de contraseñas. Estos datos se vendieron en el 2019 en el mercado negro[27].

Llegado a este punto esperamos haber mostrado porqué y cómo un uso inadecuado y sin precauciones de servicios de terceros en páginas webs puede generar riesgos no sólo en privacidad sino también en seguridad digital. En este artículo no hablamos de códigos maliciosos que pueden inyectar cibercriminales en sitios webs, ni de ataques XSS o de otros tipos. Existen muchas publicaciones de seguridad digital al respecto.

---

26 En la parte privacidad del sitio web de la empresa (https://sharethis.com/es/privacy/) se puede leer lo siguiente: "ShareThis también recopila datos sobre los usuarios de Internet y cómo interactúan con el contenido, los sitios web y los anuncios, lo que permite ShareThis y nuestros editores, anunciantes, clientes y socios de datos para facilitar la entrega de publicidad relevante y dirigida en línea a estos grupos."

27 Ver el artículo de Wikipedia y la página de la compañía en la cual anuncia el incidente de seguridad: https://en.wikipedia.org/wiki/ShareThis y https://sharethis.com/data-privacy-incident/

¿Y ahora, qué podemos hacer ? (Algunas recomendaciones)

Si soy un editor/desarrollador de un sitio web o de una aplicación:

- Hacer un análisis global de los riesgos y beneficios de la inclusión de cada código asociado a un servicio externo de terceros en el sitio web o en la aplicación (publicidad, analytics, vídeo, botones de redes sociales, CHAT de ayuda, etc.). Esta decisión no debe ser tomada únicamente por las personas responsables de comunicaciones o marketing de la entidad sino involucrar también las personas del equipo técnico y de seguridad digital (si las hay) y hasta las personas en altos cargos directivos;
- ciertas páginas del sitio que involucran datos sensibles (formularios con datos personales detallados o de conexión por ejemplo) no necesitan a priori ni deberían incluir publicidad ni otras funcionalidades ni códigos de terceros;
- los formularios de envío de datos deben usar siempre el protocolo HTTPS y el método POST;
- las cookies internas que contienen datos sensibles (ID de usuario, autenticación de sesión, etc.) deberían tener la etiqueta (flag en inglés) "HTTP Only" que ofrece una protección impidiendo su lectura por una función javascript y también la etiqueta "Secure[28]" ;
- implementar una política de seguridad de contenidos (content security policy) de manera que se permitan scripts y otros contenidos sólo cuando tengan un origen autorizado[29];
- usar metodologías de análisis para comprobar el comportamiento real de su página web o aplicación y hacer auditoría de código.

Además se pueden mencionar investigaciones y propuestas recientes de protección del lado del servidor web como la del proyecto del INRIA "Control What You Include ! Server-Side Protection against Third Party Web Tracking[30]". Por fin pensamos que hay una exigencia que se debería tener para que los terceros tengan mayor transparencia en cuanto a lo que hacen realmente sus códigos, para que ellos no sean "cajas negras". Pensamos que la responsabilidad respecto a los riesgos mencionados en este artículo son más del lado de los editores/desarrolladores y de los terceros.

Sin embargo, podemos recomendar a las personas usuarias que quieren protegerse en actividades sensibles como llenar formularios de autenticación o de entrega de datos, utilizar un bloqueo de los scripts externos. Esto se puede hacer mediante la configuración de su navegador o el uso de extensiones como

---

28 Un cookie con atributo "secure" sólo se transmite con el protocolo HTTPS.

29 https://developer.mozilla.org/en-US/docs/Web/HTTP/CSP

30 https://www-sop.inria.fr/members/Doliere.Some/essos/index.html

“No Script”[31]. Esta ya viene pre-instalada en ciertos navegadores como TOR Browser.

NOTA final: Como todas las publicaciones de K+LAB, este reporte está abierto a comentarios para eventuales correcciones, mejoras o inclusión de complementos. Este documento está bajo licencia Creative Commons “Reconocimiento, CC BY”[32].

31 https://noscript.net/

32 https://creativecommons.org/licenses/?lang=es_ES

Fuga de datos por rastreo publicitario

Enseñanzas del análisis del sitio web “Tullave” y de la aplicación de la DIAN

<K+LAB>

karisma.org.co
Twitter: @Karisma
Facebook: @fundacionkarismaa
Instagram: @karismacol