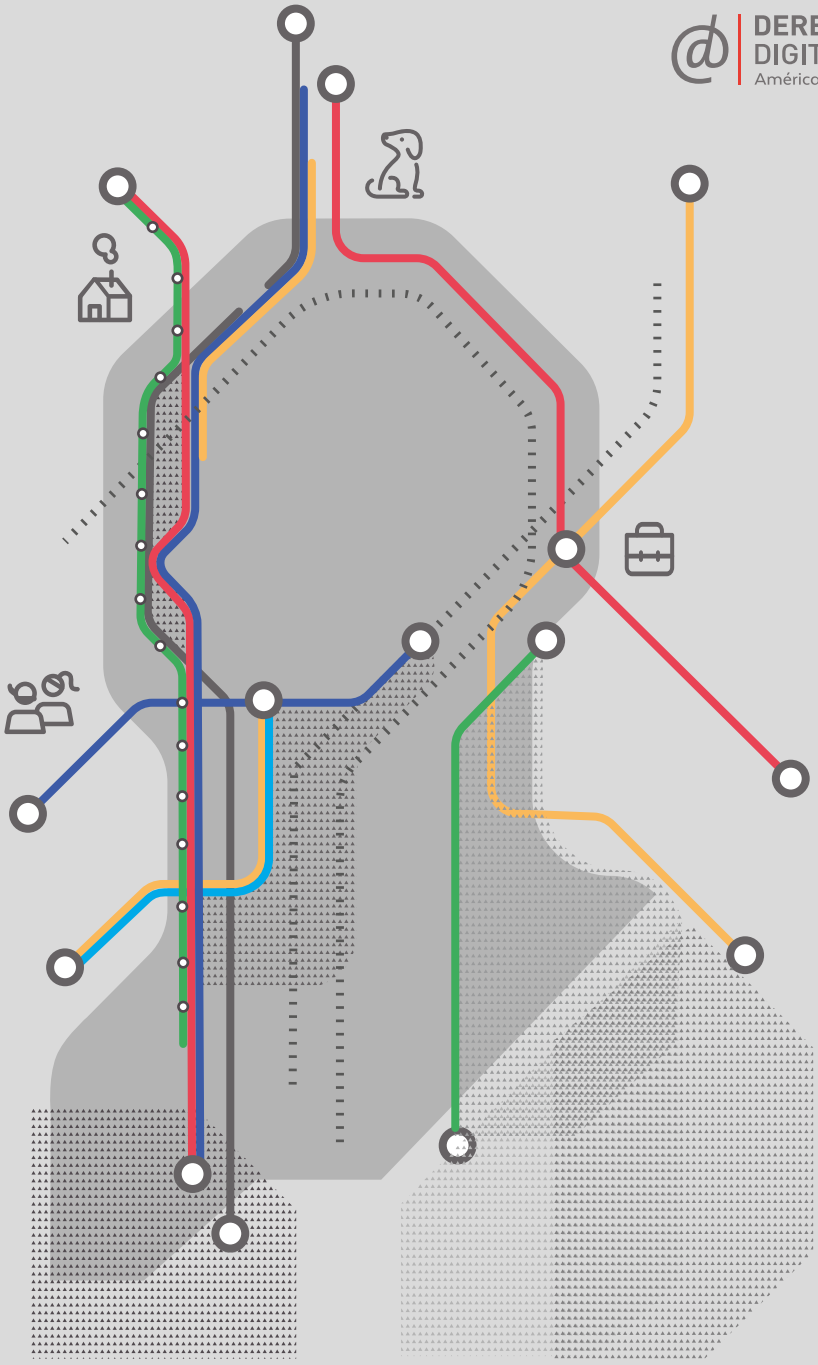
@ DERECHOS DIGITALES
América Latina



LA RECOLECCIÓN DE DATOS EN EL SISTEMA DE TRANSPORTE PÚBLICO DE SANTIAGO:

EL CASO DE LA TARJETA NACIONAL ESTUDIANTIL

POR MARTIN GUBRI

LA RECOLECCIÓN DE DATOS EN EL SISTEMA DE TRANSPORTE PÚBLICO DE SANTIAGO:

EL CASO DE LA TARJETA NACIONAL ESTUDIANTIL

POR MARTIN GUBRI

Diseño: Elisa Guzmán y Constanza Figueroa.
Edición y correcciones: Juliana Guerra y Vladimir Garay.

Esta publicación es posible gracias al apoyo de Privacy International.

Derechos Digitales, noviembre de 2016.

Agradecimientos especiales a Pedro Pablo Aste Kompen.

1. Introducción:

El siguiente es un análisis de Transantiago, el sistema de transporte público en Santiago de Chile, y las implicancias que su uso tiene sobre la privacidad de sus usuarios, específicamente a partir de los datos generados por ellos y colectados por el sistema.

Si bien esta es una idea que rondaba en el equipo hace un tiempo, y sobre la cual habíamos comenzado a realizar algunos esfuerzos, el punto decisivo fue un correo electrónico enviado por Pedro Pablo Aste Kompen donde nos informaba sobre una falla en el sistema que permitía a cualquier persona acceder al historial de uso del transporte público de cualquier estudiante de la capital.

Se trata de una violación importante del derecho a la privacidad de los estudiantes. La información en cuestión corresponde al registro histórico de geolocalización, con fecha y hora de los movimientos de cada estudiante durante los últimos tres meses, que puede ser fácilmente utilizado para rastrearlos. Además, para mucha gente, el uso del transporte público tiene cierta regularidad, por lo que a partir de estos datos es posible establecer rutinas y prever dónde estará una persona un día determinado, a una hora determinada.

No es difícil imaginar situaciones peligrosas que pueden generarse a partir del acceso a esta información, por ejemplo, por un acosador violento; pero nos resulta particularmente importante poner esta información en contexto: en la última década los estudiantes se han posicionado como una de las fuerzas políticas más importantes en Chile, utilizando como principal forma de expresión las manifestaciones masivas en la vía pública, las que usualmente han sido reprimidas de forma violenta por la policía.

Existen casos documentados en prensa respecto a la recolección que la policía ha realizado de los datos disponibles en internet sobre los estudiantes movilizados, por ejemplo, en sus cuentas de Facebook, por lo que no es descabellado pensar que una vulnerabilidad de este tipo podría ser utilizada para vigilar más de cerca de los manifestantes.

Vigilar los movimientos de uno (o de muchos) estudiantes, puede constituir una violación a los derechos humanos, por lo que nos parece muy importante estudiar cómo se maneja este tipo de datos, cuáles son las buenas prácticas que deber ser promovidas, así como las malas prácticas que es necesario evitar.

Con este trabajo queremos hacer un aporte que ayude a la elaboración de mejores políticas públicas, donde el respeto por la privacidad de las personas sea preocupación central.

2. Transantiago y la tarjeta Bip!

Transantiago es el nombre del sistema de transporte público que opera en el área metropolitana de la capital chilena, Santiago.

Destinado a cambiar por completo la organización del transporte colectivo por microbuses, Transantiago reformó la malla de recorridos de las antiguas micros, diseñando un sistema basado en el uso de servicios alimentadores y troncales, en conjunto con el Metro de Santiago.

Para ello, se desarrolló una enorme inversión en infraestructura y flota vehicular, y además se estableció un sistema tarifario integrado, por medio de una tarjeta de pago conocida como Bip!

2.1. La tarjeta Bip!

Existen cuatro tipos de tarjetas Bip!: la regular, la Tarjeta Nacional Estudiantil (TNE), la activación de las tarjetas bancarias como Bip! y una tarjeta personalizada.[1] Nuestra investigación se concentró en las dos primeras, pues son las más usadas.

La tarjeta de uso regular es anónima y puede ser adquirida en las estaciones de metro por alrededor de un dólar y medio. Es una tarjeta plástica, identificable solamente a partir de un número impreso en ella.

La TNE es la tarjeta utilizada por los estudiantes para acceder a la tarifa especial rebajada. A diferencia de la tarjeta Bip! de uso regular, es nominativa, personal e intransferible. La tarjeta posee el nombre de su titular, una fotografía (para evitar que sea utilizada por otros), el número de identificación nacional (RUN o RUT), la institución educativa a la cual pertenece y el número de tarjeta.

Si bien nos parece importante y correcto que por defecto la tarjeta utilizada en el transporte público sea anónima, una de las preguntas que gatilla esta investigación es qué tan anónima es realmente y si existe alguna posibilidad de utilizar la tecnología para facilitar la identificación del usuario de una tarjeta.

2.1.1. Tecnología de la tarjeta Bip!

La tarjeta Bip! utiliza tecnología RFID (*Radio Frequency IDentification,* identificación por radiofrecuencia), un sistema de almacenamiento y recuperación de datos remoto, cuyo propósito fundamental es transmitir la identidad de un objeto (similar a un número de serie único) mediante ondas de radio.

---

1 Para más detalles sobre las tarjetas, ver https://es.wikipedia.org/wiki/Tarjeta_bip!#Tipos_de_tarjetas y http://tarjetabip.cl/que-es-tarjeta-bip.php

Como explica Adrian Dabrowski,[2] RFID es un concepto y no una norma: hay diferentes implementaciones de este principio y diferentes tipos de tarjetas RFID, desde la más simple y barata, a la más compleja que también es la más cara.

Para el Transantiago, los reguladores chilenos podían escoger entre utilizar una tarjeta RFID simple, barata, con un sistema criptográfico débil, o una tarjeta compleja y más cara, con un sistema criptográfico serio.[3]

Como muchas organizaciones en el mundo,[4] la tarjeta Bip! usa la tecnología MIFARE Classic,[5] que emplea NFC, una tecnología de comunicación inalámbrica, de corto alcance y alta frecuencia, que permite el intercambio de datos entre dispositivos.

Las tarjetas regulares anónimas usan MIFARE Classic 1K y la tarjeta TNE usa MIFARE Classic 4K. La diferencia es la cantidad de memoria disponible: la primera tiene 1024 bytes de memoria y la segunda 4096 bytes,[6] probablemente, porque esta última necesita más espacio para guardar más datos personales.

Confirmamos el uso de estas tecnologías usando un lector NFC en USB.

El problema con esta tecnología es que, para economizar costos, la criptografía usada es mala. El sistema criptográfico se llama Crypto-1, es privado y cerrado. La seguridad de este algoritmo no está basada en investigaciones revisadas por pares, sino en mantener el secreto de su funcionamiento. Gracias a la ingeniería inversa, el algoritmo fue "quebrado"[7] y existen al menos tres maneras documentadas de explotar diferentes fallas, que pueden ser utilizadas para conocer las llaves usadas para leer y escribir en la tarjeta.[8]

---

2 Ver https://media.ccc.de/v/30C3_-_5334_-_en_-_saal_2_-_201312291400_-_rfid_treehouse_of_horror_-_adrian_dabrowski

3 Acá un esquema que ilustra las diferentes opciones https://framapic.org/BSwXM3RsZFES/gYnthr0bV9J5.png

4 Ver https://en.wikipedia.org/wiki/MIFARE#Places_that_use_MIFARE_products

5 Ver https://es.wikipedia.org/wiki/Tarjeta_bip!#Especificaciones_t.C3.A9cnicas

6 Ver https://en.wikipedia.org/wiki/MIFARE#MIFARE_Classic

7 Ver https://en.wikipedia.org/wiki/MIFARE#Security_of_MIFARE_Classic.2C_MIFARE_DESFire_and_MIFARE_Ultralight

8 Las maneras conocidas están acá:
- MFCUK https://github.com/nfc-tools/mfcuk
- MFOC http://nfc-tools.org/index.php?title=MFOC
- Crapto-1 https://github.com/Tilka/crapto1
- Libfreefare http://nfc-tools.org/index.php?title=Libfreefare
- Otros herramientas http://nfc-tools.org/index.php?title=MIFARE_Classic

Intentamos reproducir estos ataques usando un Dongle NFC USB, sin éxito.[9]

Sin embargo, la llave secreta completa de las tarjeta Bip! esta disponible en internet.[10] Estas vulnerabilidades se hicieron populares en octubre de 2014, luego de que alguien creara una aplicación para Android que permite añadir $10.000 al saldo de tu tarjeta, utilizando un teléfono que integra NFC.[11]

Se puede encontrar mucha documentación en línea sobre este *hack.* Por ejemplo, la parte del código Java usado en la aplicación. Esta vulnerabilidad se encuentra en pastebin.[12] Decidimos que no había mucho terreno que explorar en esta línea.

Antes de esta aplicación, Transantiago usaba los datos de las tarjetas para validar las transacciones en las estaciones de metro. En respuesta a los cargos ilegítimos, se cambió a un sistema en línea: la maquina verifica el saldo registrado en la tarjeta con los datos almacenados en el *backend* del sistema. Dado que los buses no tienen conexión con la red interna de Transantiago, implementar una solución similar ahí resulta mucho más costoso y, por ahora, no existe un sistema de verificación en línea para los buses.

No sabemos si en los buses existe una "lista negra" que impida el uso de las tarjetas ilegalmente cargadas.

Este *hack* tiene implicancias para la privacidad. Que la llave para leer y escribir en las tarjetas sea de público conocimiento significa que cualquier persona con un lector NFC puede atacar de distintas formas a una persona determinada. Por ejemplo, se puede seguir a alguien y obtener su número de tarjeta (y los datos personales de las tarjetas en el caso de la TNE).

Con un lector NFC estándar, la dificultad es que la distancia de lectura es corta. Pero en una situación de alto tráfico, como cualquier día de la semana en el Metro de Santiago, es muy viable (dentro del tren o en las escaleras mecánicas).

De lo contrario, es posible comprar lectores NFC más específicos, de mayor alcance. Con este aparato es posible seguir una persona y tener su número de tarjeta (más adelante veremos las

---

9 La tarjeta Bip! no tiene llave por defecto y los ataques *offline* paraban después de algunas horas, probablemente por un problema de hardware. No intentamos guardar un *dump* de translaciones legítimas para intentar otros tipos de ataques, lo que podría haber sido de ayuda, además de tener un equipo más costoso, como un Proxmark.

10 Como funciona el 'Glitch' de la Tarjeta bip! ? (Carga mediante NFC)". Disponible en https://www.youtube.com/watch?v=VUvSpyNg9OI

11 Ver http://www.cooperativa.cl/noticias/tecnologia/industria/informatica/expertos-analizaron-el-hackeo-a-la-tarjeta-bip/2014-10-23/155152.html y http://securityaffairs.co/wordpress/29491/cy ber-crime/chile-nfc-ticketing-hacked.html

12 Ver http://pastebin.com/reZ3MaUU

consecuencias que puede tener esta información en la privacidad).

Además, el atacante podría escribir en la tarjeta para dejarla inutilizable. Una acción como esta podría ser obra de un acosador, alguien que se oponga a que ciertos grupos se manifiesten, e incluso a la policía y los servicios de inteligencia.

Otra cosa que puede hacer un atacante es clonar una tarjeta para usarla. Eso se puede hacer comprando tarjetas NFC especiales en las que se puede escribir el UUID. En general, no se puede escribir el UUID de las tarjetas NFC vacías, (solamente de lectura), pero no es difícil adquirir legalmente tarjetas con el UUID reescribible en internet.

3. Metodología de trabajo

3.1. El problema de privacidad de las tarjetas TNE

Iniciamos nuestra investigación trabajando sobre el problema de privacidad de la tarjeta TNE. Tras recibir el correo electrónico de Pedro Aste, verificamos el problema e investigamos para qué sirven las plataformas que publican los datos. Nos reunimos con Pedro para obtener más informaciones y él hizo una demostración práctica de la vulnerabilidad.

Decidimos implementar de nuevo su *script* de demostración usando nuestra tecnologías (Python 3) para tener la posibilidad de hacer demostraciones nosotros mismos. En paralelo, investigamos los sitios web que publican los datos de las tarjetas de transporte, para realizar un reporte en caso de encontrar otras fallas.

Con investigaciones adicionales, escribimos un reporte técnico corto, que enviamos a los responsables de los sitios web involucrados. Explicamos los problemas, sus implicaciones en términos de privacidad y las soluciones que recomendamos.

Después, nos comunicamos con el Ministerio de Educación, responsable de uno de los sitios web. Dado que agendaron una reunión con nosotros sesenta días tras la solicitud, decidimos contactar al Ministerio de Transporte, responsable del otro sitio web.

En la reunión, explicamos los detalles del problema y entregamos copias del reporte técnico. El Ministerio de Transporte envió copias del reporte a los responsables del Ministerio de Educación.

Algunas semanas después descubrimos que el sistema de tne.cl había sido cambiado. Hicimos una evaluación de los cambios en relación a nuestras recomendaciones. Descubrimos algunas fallas importantes en la solución establecida y comunicamos por mail nuestras preocupaciones a nuestro contacto en el Ministerio de Transporte.

Durante todo este proceso, tuvimos mucho cuidado de limitar la comunicación de este problema de privacidad a un numero mínimo de personas, para que la falla no se hiciera pública antes de ser arreglada y evitar exponer a cientos de miles de personas.

3.2. Los sitios web

En paralelo, decidimos investigar el manejo de los datos generados por el sistema de transporte, buscando otras fallas que pueden tener un impacto sobre la privacidad de los usuarios. Estudiamos el funcionamiento normal de los dos sitios web que permiten el acceso a datos de usuarios de Transantiago para familiarizarnos con ellos: Bip! en línea y tne.cl. Mas específicamente, estudiamos los puntos siguientes:

- El detalle de las funcionalidades del sitio web.
- Los datos disponibles (tipos de datos, todos las variables, sus unidades y sus precisiones)
- El *login* (datos necesarios para ingresar (datos personales o no), formularios html, *javascript*, etc.).
- El *javascript* usado en el sistema.
- Las *requests* envidas y recibidas durante la carga y el uso de las paginas (tipos, archivos/*scripts* externos, etc.).
- El ingreso de datos (formularios html).
- La seguridad de la conexión entre el servidor y el navegador (TLS disponible).
- La protección contra el *scraping* masivo de datos (evaluación de la factibilidad).
- El formato de los datos (html, imagen, etc. en relación con el punto precedente).
- "*Hand-crafted requests*" para buscar fallas de seguridad.
- Versiones de los softwares usados para investigar la existencia de *exploits*.
- Algunas configuraciones malas del servidor web.

Para testear algunos de los puntos precedentes, programamos unos pequeños *scripts*. Un *script* más grande fue creado para investigar la posibilidad de hacer un *scraping* masivo de los datos. Es importante conocer qué tan fácil resulta almacenar masivamente los datos de los usuarios, porque los ataques a la privacidad son diferentes si el atacante puede obtener los datos de todos.

Entonces, decidimos intentar un *scraping* masivo, para ver si hay un sistema automático que limite el número de *requests* desde una misma dirección IP (un equivalente de *fail2ban*), si hay una lista negra manual o si no hay ninguno de los dos.

Además estábamos interesados en conocer la velocidad del *scraping*. Este parámetro es importante para determinar qué tipo de ataque se pueden realizar. Nos interesaba saber qué porción del historial de tres meses podemos tener simultáneamente para todas las tarjetas; por ejemplo, si el *scraping* necesita un mes entero para colectar la información, vamos a tener dos meses simultáneos de información de todos los usuarios del sistema. Además, la velocidad puede determinar la facilidad para guardar los datos.

Además, buscamos otros sitios web que pueden guardar informaciones asociadas con la entrega de un número de tarjeta, por ejemplo, el sitio web para cargar en línea.

Los resultados son detallados a continuación.

3.3. Las tarjetas físicas

Como se detalló en la parte anterior, la llave para leer y escribir en la tarjeta está disponible en internet.

Decidimos investigar los datos guardados en la memoria de 1k de la tarjeta Bip! y de 4k de la tarjeta TNE. Es importante saber qué tipo de datos están en las tarjetas porque, dado que la llave está disponible en internet, fácilmente cualquier persona podría crear un *sniffer* y acceder a estos datos.

Un *sniffer* es un dispositivo que es posible fabricar con un *hardware* barato (por ejemplo, usando un Arduino con una antena NFC[13] de 40 dólares) y que permite leer el contenido de las tarjetas próximas. Existen antenas más grandes y más caras que permiten leer tarjetas a mayor distancia.

El tipo de datos contenidos dentro de la tarjeta es muy importante, porque son accesibles por cualquiera que esté cerca, sin que necesariamente la víctima se de cuenta. Podemos pensar en muchas situaciones peligrosas, desde acosadores hasta investigadores policiales (realizando vigilancia potencialmente ilegal).

En nuestro caso, transcribimos la llave de acceso a un formato usable, y utilizamos mfoc[14] (que usa libnfc) obtuvimos un *dump* binario del contenido de la tarjeta. El *hardware* que usamos es el dongle NFC USB SCL3711 que cuesta 40 USD.

Para leer e investigar el *dump*, usamos el *script* python mfdread[15] y un editor de archivos binarios. Investigamos el contenido de las tarjetas BIP! anónima y de la tarjeta TNE.

No intentamos modificar el contenido de la tarjeta para ver si es posible hacerse pasar por otro número, sin modificar el UID.

3.4. Otros atentados a la privacidad

Buscamos otras maneras en que fuese posible atentar contra la privacidad de los usuarios del transporte público. Una de las metodologías usada fue ver los casos judiciales o de investigación policial documentados. Usamos también nuestra propia experiencia de usuarios de transporte público para imaginar otras manera de asociar un número de tarjeta a un nombre.

---

13 https://www.adafruit.com/products/789

14 https://github.com/nfc-tools/mfoc

15 https://github.com/zhovner/mfdread

3.5. El valor de los datos de movimientos en los transporte público

El sitio web de Transantiago permite conocer el saldo en la tarjeta y permite también ver todos los movimientos asociados a ella durante los tres últimos meses.

Eso nos dio la idea de investigar lo que podemos deducir sobre una persona solamente usando *data mining* de tres meses de datos de transporte público.

Para investigar esto, escribimos un *script* en Python 3 usando *scrapy* para guardar los datos de una manera fácilmente utilizable para analizarlos.

El *script* se configura con el (o los) número(s) de tarjetas de interés. El código ingresa al sitio web usando la misma *request* que un usuario regular de la pagina, guardando el código de sesión y algunos datos de la tarjeta (su saldo, la fecha de actualización del saldo y el tipo de contracto asociado a la tarjeta).

Después, usando la sesión iniciada, se envía un segundo *request* para obtener la página html con la tabla de los movimientos de los tres últimos meses. Por cada movimiento, guardamos fecha y hora (horas y minutos, segundos no disponibles), el lugar (estación de metro, o número de línea de bus y matrícula del vehículo del bus), el tipo de acción de la tarjeta (uso, carga, otros), su costo, el saldo después de este movimiento y el número de transacción asociado con la tarjeta.

Finalmente, el *script* envía una *request* de desconexión para no sobrecargar el servidor de llave de sesiones. El *script* se puede correr a través de Tor.

Otro *script* analiza los datos para obtener informaciones que pueden ser relevantes. Puede analizar algunas estadísticas sobre el uso de la tarjeta, los recorridos usuales y los recorridos inusuales. El objetivo es demostrar que estos datos son importantes, porque dan mucha información sobre una persona.

Hicimos una pequeña presentación de estos datos en una interfaz web, que permite mostrar esta información en un texto que explica y presenta los datos extraídos. Este parte funciona gracias a Flask.

4. Resultados

4.1. El problema de privacidad de las tarjetas TNE

Estudiamos el grave problema de privacidad en la tarjeta TNE, que funciona usando dos sitios web, Bip! en línea y tne.cl.

El problema es el siguiente: cualquiera que conozca el nombre de un (a) estudiante -actualmente o que haya tenido el beneficio en el pasado- puede acceder a toda la información personal de esa persona, usando la función "Estado de tu TNE",[16] que incluye:

- Nombre.
- Apellido.
- RUT.
- Nivel de estudio (superior, media o básica).
- Historial de todas las TNE de esta persona.

Para cada una de estas tarjetas es posible acceder a los siguientes datos:

- Número de tarjeta de la TNE, sin las dos primeras cifras.
- Código de referencia de la institución educacional.
- Región geográfica de Chile.
- Año.
- Estado de la tarjeta (activa/inactiva).

Esto es posible debido a dos vulnerabilidades: en primer lugar, el número de la tarjeta TNE está asociado al RUT, número de identificación personal usado en Chile.

Es posible acceder al RUT de prácticamente cualquier persona, a través de portales como Rutificador.[17]

Conociendo el RUT, es posible acceder al sitio web tne.cl y acceder a la opción "Estado de la TNE". Llenando el RUT, el sistema revela el número de la TNE, ocultando solo los últimos dos dígitos. El número máximo de combinaciones posibles es 100, por lo que es posible incluso probarlas todas de forma manual. Pero es sabido que el número de las tarjetas TNE es creado de forma consecutiva y que comienzan con 7 o 8.[18]

Esto permite solamente 20 combinaciones posibles. Por otra parte, alguien realizando una investigación puede ver que no todos los números en el rango de los 20 millones entre 70.000.000 y 89.999.999 están siendo utilizados, por lo que el número de combinaciones posibles es menor a 20.

---

16 http://sistema.tne.cl/reposiciones/estado_tarjeta_alumno

17 Ver http://datos.24x7.cl/

18 www.joshuaprovoste.com/analizando-la-tarjeta-bip/

Una persona con conocimientos muy básicos de programación puede automatizar el texto de todos los números de tarjeta posibles a través del sitio web de Transantiago.[19]

Con esta información es posible acceder a tres meses de historial de la tarjeta, incluyendo todas las actividades y transacciones realizadas (validaciones, cargas y otras).

Por cada validación realizada durante los últimos tres meses, es posible conocer:

- Nombre de la estación, si es que se realizó un viaje en Metro.
- La línea de bus y la patente del vehículo, si se realizó un viaje en bus.
- La fecha (DD-MM-AAAA) y la hora (HH:MM) en que se realizó el viaje.
- El costo de la validación.
- El saldo en la tarjeta.
- El número total de actividades realizadas.

Para cada carga realizada durante los últimos tres meses, es posible conocer:

- El lugar en que se realiza la carga : la estación de Metro o el nombre y dirección del local comercial asociado.
- La fecha (DD-MM-AAAA) y la hora (HH:MM) en que se realizó.
- La cantidad de dinero cargado.
- El saldo de la tarjeta.
- El número total de cargas realizadas.

Con esta información podemos saber mucho sobre los hábitos de una persona. Todavía más importante, cuando la gente usa el transporte público de forma regular, algunos patrones emergen claramente y se hace relativamente sencillo saber en qué sector estudia alguien (validaciones realizadas en la mañana), en qué sector vive (validaciones realizadas por la tarde). Alguna gente carga su tarjeta en el negocio más cercano a su hogar y con toda esta información podemos tener una idea más precisa de dónde vive.

Un problema mayúsculo podría ser, por ejemplo, que un acosador podría determinar con gran precisión dónde estará la persona acosada en un determinado momento, tan solo viendo su historial de viajes.

4.2 Soluciones recomendadas por Derechos Digitales

Ante los problemas expuestos más arriba, sugerimos las siguientes soluciones:

- Enmascarar, al menos, los cuatro últimos dígitos del número de la tarjeta TNE (dado que los números de tarjeta son asignados de forma consecutiva, es demasiado sencillo adivi-

19 www.metrosantiago.cl/contents/guia-viajero/includes/consultarTarjeta/xxxxxxxxor http://pocae.tstgo.cl/PortalCAE-WAR-MODULE/

nar los dígitos al inicio). Implementando esta medida, sigue siendo sencillo identificar el número de la tarjeta si se tienen conocimientos de programación, pero es casi imposible hacerlo de forma manual.
- No entregar información sobre el código de las instituciones educacionales pasadas y actuales del beneficiario (Código RBD) en la tarjeta. Y de forma más general: no mostrar más información que la estrictamente necesaria (principio de *data minimization*).
- No pedir información disponible en el sitio de la TNE para acceder a Bip! en línea.
- Establecer un sistema de contraseñas fuertes para acceder al sitio de TNE (solución a mediano plazo).

4.3. Descripción y evaluación de los cambios realizados por por la Junta Nacional de Auxilio Escolar y Becas (Junaeb)

Los responsables del sitio web de tne.cl hicieron algunos cambios tras la comunicación de nuestro reporte técnico. Queremos felicitar a esta institución por preocuparse de la protección de los datos personales de los usuarios del sistema.

El nuevo sistema esconde los 5 primeros números de la tarjeta, por lo que es imposible descubrir los dígitos faltantes a mano: en lugar de 20 combinaciones, tenemos ahora 20.000 números de tarjetas posibles.

Todavía es posible usar un *script* para conocer la tarjeta de alguien, pero necesita más tiempo para ejecutar. Según nuestras pruebas se necesitan 60 minutos para intentar los 20.000 números posibles. Eso significa que es posible conocer las tarjetas de solamente algunas personas.

La solución técnica para solucionar este problema es implementar un *captcha* en el sitio web "Bip! en línea".

Otro cambio que queremos destacar es la aplicación del principio de *data minimization*, presentar solamente los datos estrictamente necesarios para una funcionalidad.

En efecto, el nuevo sistema da solamente los datos siguientes:

- Primer nombre
- Apellido paterno
- Tres últimos números de la tarjeta
- Estado TNE

Ya no está el código de la institución, que puede permitir en algunas situaciones una violación importante a la privacidad, por ejemplo, contra menores de edad. Además, solamente está disponible la información sobre la última tarjeta y no las pasadas.

Sin embargo, descubrimos un problema grave. El antiguo sistema es todavía accesible en internet para cualquier persona que tiene la URL. La antigua forma de ataque con solamente 20 combinaciones posibles para conocer el número de la tarjeta y acceder a los datos todavía es posible.

Notificamos este problema, pero no teníamos respuestas y no vimos otros cambios en el sitio web.

4.4. Investigaciones complementarias sobre los sitios web

4.4.1. Bip! en línea

Funcionalidades

Este sitio web sirve para consultar el saldo la tarjeta Bip! y los movimientos de los tres últimos meses.

Se accede a través de la siguiente URL: http://pocae.tstgo.cl/PortalCAE-WAR-MODULE/

El sitio puede ser usado para contactar al servicio cliente. En este caso, el ingreso del RUT es necesario para enviar un mensaje. Este puede ser un medio para de-anonimizar una tarjeta.

Además, funciona para cargar la tarjeta en línea y permite recibir un newsletter, para lo cual pide datos adicionales.

Datos disponibles

Para cada tarjeta, los datos siguientes están disponibles:

- Saldo.
- Fecha de actualización del saldo.
- Tipo de contracto asociado con la tarjeta.
- Todas las actividades de la tarjeta durante los últimos 3 meses (validación en el metro y bus, cargas, y otros, como la activación de su tarjeta TNE).

Por cada validación, es posible conocer:

- Nombre de la estación, si es que se realizó un viaje en Metro.
- La línea de bus y la patente del vehículo, si se realizó un viaje en bus.
- La fecha (DD-MM-AAAA) y la hora (HH:MM) en que se realizó el viaje.
- El costo de la validación.
- El saldo en la tarjeta.
- El número total de actividades realizadas.

Para cada carga realizada durante los últimos tres meses, es posible conocer:

- El lugar en que se realiza la carga : la estación de Metro o el nombre y dirección del local

comercial asociado.
- La fecha (DD-MM-AAAA) y la hora (HH:MM) en que se realizó.
- La cantidad de dinero cargado.
- El saldo de la tarjeta.
- El número total de cargas realizadas.

Podemos saber muchas cosas sobre una persona solamente mirando estos datos con atención.

Intentamos modificar la *request* (*hand-crafted values*) que carga la página de los movimientos en busca de una falla que puede permitirnos saber si hay más de tres meses de información almacenada, pero siempre obtuvimos tres meses de datos (lo que no prueba que no hay más información).

El ingreso al sitio (login)

El sitio tiene un sistema de login propio. El formulario pregunta el número de la tarjeta. Si se trata de una tarjeta corriente, no es necesaria más información. Si el número corresponde a una tarjeta TNE, una función en *Javascript* añade un campo para pedir el RUT.

Detectamos una falla en el login. En el caso de la TNE, el contenido del campo RUT esta rescrito en *Javascript* para presentar el formato del número (que incluye una cifra verificadora separada por un guión).

Observamos que enviando la misma *request*, pero con un RUT igual a "0" el ingreso se hace de todos modos. En el sitio web, el ingreso 0 esta prohibido, pero hay que recordar que una medida de seguridad en el cliente no es seguridad. El servidor siempre tiene que verificar los datos enviados por el cliente. Sugerimos implementar una verificación en este caso.

Además, como lo vimos en la parte precedente, sugerimos implementar a corto plazo un *captcha* para evitar el problema de privacidad de la TNE.

A largo plazo, pensamos que es buena idea implementar un sistema de contraseña para acceder al sitio web, para proteger los datos de las tarjetas TNE y de las tarjetas anónimas.

En efecto, como lo vimos, acceder a estos datos solamente con el número de la tarjeta no es seguro, porque el número puede ser obtenido de muchas maneras.

Scraping masivo

Intentamos evaluar la factibilidad de realizar un guardado masivo de datos, pues estos datos pueden interesar a muchas empresas (marketing, data brokers, etc.). Además, conocer los movimientos de todos los usuarios permite asociar un número de tarjeta a una persona específica. Si conozco dos actividades realizadas por esa misma persona (específicamente, la estación de me-

tro y la hora en que realizaron las actividades) puedo obtener fácilmente el número de la tarjeta.

Dado que los números de tarjetas son creados de maneras consecutiva, es más sencillo realizar el *scraping*: Solamente se necesita conocer el rango de los números. Algunas personas ya lo han hecho.[20] Para evitarlo, recomendamos asociar de manera aleatoria los números de tarjetas.

En este caso, ejecutamos un *script* que ingresa al sitio web, guarda los datos y se desconecta. Así para cada tarjeta.

Lo ejecutamos durante todo una fin de semana. No tuvimos problema, pero el lunes la dirección de nuestro servidor fue bloqueada.

Eso nos dice que quizás existe una lista negra compilada manualmente. No hay una protección de tipo *fail2ban* para limitar las *requests* desde una misma IP.

Decidimos intentarlo de nuevo, pero trabajando más lento. Esta vez no encontramos problemas.

La velocidad del sitio web es un tanto lenta, entonces necesita mucho tiempo para guardar los datos de todas las tarjetas producidas durante años. Pero no es muy complejo obtener algunos millones de tarjetas (lo hicimos).

Esto puede ser muy útil para un servicio policial, para implementar silenciosamente una vigilancia masiva. Pensamos particularmente en la vigilancia de los usuarios de los transporte público en el marco de manifestaciones.

Configuración del servidor web

Pudimos observar que el servidor web no está bien configurado: en la raíz aparece la página de demostración de Tomcat.[21] Es importante desactivar este tipo de página, además de las funcionalidades no usadas de Tomcat.

Por ejemplo, se puede saber la versión exacta de Tomcat usada, que además es muy vieja (parece que no se ha actualizado desde 2009). Y para esta versión existen *exploits* publicados para penetrar el servidor, como CVE (no intentamos usarlos).

Queremos insistir sobre la importancia de actualizar este tipo de *software*, y en general todos los *softwares* de un servidor. Después de nuestro intento de *scraping* masivo, vimos que los responsables de la plataforma actualizaron Tomcat. Es una buena idea, pero esperar a tener problemas para hacerlo no lo es.

---

20 http://www.joshuaprovoste.com/analizando-la-tarjeta-bip/

21 http://pocae.tstgo.cl

Además es necesario ahora configurar el servidor para ofrecer https, por la seguridad de la conexión entre el servidor y el navegador.

Conclusión y recomendaciones

Además de las recomendaciones mencionadas más arriba, recomendamos lo siguiente:

- No guardar los datos de los movimientos de todos los usuarios.
- Si Transantiago necesita guardarlos, recomendamos usar un servidor desconectado de Internet (*backend*) para no exponer a la gente a diferente riesgos.
- Si Transantiago necesita poner estos datos en línea, recomendamos usar un sistema de *login* adaptado: *captcha*, contraseña, verificación suficiente del servidor (que la seguridad no dependa del cliente).

Además, en todos los casos, es una buena práctica:

- Siempre usar TLS para la conexión entre el servidor y el navegador (https forzado).
- Actualizar el servidor y sus softwares para evitar el uso de *exploits*.
- Configurar de manera correcta el servidor web.

4.4.2. TNE.CL

Funcionalidades

El sitio web sirve para que los estudiantes pueden consular el estado de su tarjeta TNE. Es útil para seguir el proceso de activación.

Datos disponibles

Asociados con el RUT, los siguientes datos están disponibles:

- Nombre.
- Apellido.
- Nivel de estudio (superior, media o básica) (ausente del nuevo sitio web).
- Historial de todas las TNE de esta persona (ausente del nuevo sitio web, al excepción de la ultima tarjeta).

Para cada una de estas tarjetas es posible acceder a los siguientes datos:

- Número de tarjeta de la TNE, sin las dos primeras cifras (cinco en el nuevo sitio web).
- Código de referencia de la institución educacional (ausente del nuevo sitio web).
- Región geográfica de Chile (ausente del nuevo sitio web).
- Año (ausente del nuevo sitio web).
- Estado de la tarjeta (activa/inactiva).

Podemos decir que la disminución de los datos disponibles en el nuevo sistema es una buena idea, porque no eran necesarios para la funcionalidad del sitio web.

Uno de los problemas es que el sistema hoy está oculto, pero sigue activo, con todos los datos disponibles.

El ingreso al sitio (login)

El sitio web no tiene un sistema de login propio. El formulario para ingresar el RUT carga en *Javascript* el contenido desde una URL que lo contiene. Más específicamente, el contenido se carga desde http://sistema.tne.cl/reposiciones/estado_tarjeta_alumno/tneEmitidas/<RUT>/<RUT_verif>

Scraping masivo

El servidor no tiene protección para detener un número excesivo de conexiones desde una misma dirección IP. Además, el *scraping* se puede hacer cargando solamente una pagina html que tiene los datos en variables *Javascript*. Entonces una *request* es suficiente por cada RUT. Cada *request* es rápida.

Al ser consecutivos los RUT, es bastante fácil definir un rango de números con los cuales trabajar, a partir de la edad. Por estas cuatro causas es muy fácil realizar un *scraping* masivo de los datos.

Con estos datos sería bastante fácil asociar un número de tarjeta TNE con un nombre: hay que buscar en la base de datos las tarjetas e intentar conectarse a "Bip! en línea" a cada una de los RUT asociados en nuestra base de datos, hasta que el *login* funciona.

Conclusión y recomendaciones

Además de los cambios realizados, recomendamos:

- Esconder más cifras.
- Poner una *captcha* para acceder al sitio de la TNE, para evitar la recopilación masiva de datos personales (solución a corto plazo).
- Usar un sistema de identificación (login) en el sitio web de la TNE.
- Establecer un sistema de contraseñas fuertes para acceder al sitio de TNE (solución a mediano plazo).
- Hacer las siguientes URL disponibles solo para usuarios que hayan iniciado sesión en el sitio, para evitar la recopilación masiva de datos personales: URL escondidas para proteger la privacidad.
- Ofrecer https para asegurar la conexión entre el servidor y el navegador (y forzar su uso).
- Implementar una protección contra el scraping masivo de datos de tipo *fail2ban*.

4.4.3 Carga Tarjeta bip! en Línea

Otro sitio web analizado es el que permite la carga de la tarjeta Bip! en línea.[22] Usa WebPay como sistema de pago y es posible que este sistema pueda ser usado como un medio para asociar un nombre a un número de tarjeta, usando los datos bancarios.

No sabemos cuál es la política del servidor de WebPay y del sitio web "RedBip" para guardar y proteger los *logs* generados. Si los servidores guardan la dirección IP, es posible asociar el número de tarjeta a la dirección IP usada durante el acceso al sitio web.

Además, notamos que este sitio web usa Google Analytics, que puede ser utilizado para obtener más información sobre algunas personas.

4.5. Las tarjetas físicas

Investigando el contenido de la memoria de las tarjetas Bip! y de las tarjetas TNE, confirmamos que las dos guardan el numero de la tarjeta.

Además, la tarjeta TNE guarda el primer nombre y el apellido paterno del titular. Entonces, si alguien tiene el equipo necesario para leer tarjetas a distancia, puede saber el nombre del estudiante (incluyendo menores de edad), y la información necesaria para acceder al historial de uso de la tarjeta de cualquier persona, a través del sitio "Bip! en línea".

En la segunda parte de este reporte vimos los peligros potenciales asociados a las tarjetas físicas. Principalmente, la publicación de la llave para leer y escribir en las tarjetas permite a cualquier persona que posea el hardware necesario, que es barato y relativamente fácil de conseguir, hacer un *sniffer* para acceder a los números de tarjeta.

Esto podría eventualmente ser utilizado, por ejemplo, por los servicios policiales para vigilar a alguien: en lugar de seguir a una persona, su número de tarjeta permite seguirlo de manera rápida, fácil y gratuita a través de internet.

4.6. Otros atentados a la privacidad

De nuestra experiencia como usuarios de Transantiago, vimos que las boletas de cargas de Bip! en tiendas tienen impreso el número de la tarjeta. Observamos que mucha gente no guarda el boleto y lo botan. Eso puede ser un medio de ataque *low-tech* para asociar un número de tarjeta con una persona.

---

22 Ver http://carga.tarjetabip.cl/

Además, el caso del atentado de Santiago en 2014[23] confirma que la policía es capaz de asociar un número de tarjeta a un sospechoso. En este caso, la policía encuentra los dos números de tarjetas Bip! de los atacantes usando las cámaras de seguridad de la entrada en el metro. Transantiago guarda las validaciones en el metro y puede tener los números de tarjetas que estaban a esa hora en la estación.

Gracias al sitio sabemos que la hora está guardada con una precisión de menos de un minuto. Pero quizás la base de datos tiene una precisión al segundo.

Si Transantiago tiene una precisión de un minuto, es posible saber el número de tarjeta de una persona usando las cámaras de seguridad estudiando otras validaciones en el metro. En efecto, en un momento de alta afluencia en el Metro, mucha gente va a validar durante un minuto. Eso implica que hay muchos números posibles. La solución es buscar otras validaciones de estas tarjetas y mirar las cámaras de seguridad en este momento, hasta que se encuentra a la persona.

4.7. Valor de los datos de movimientos en el transporte público

Podemos deducir muchas informaciones solamente estudiando con atención los datos proporcionados por los sitios web del sistema de transporte, sin usar otras herramientas.

En general, se ve fácilmente cuáles son las estaciones de Metro y las líneas de bus más usadas. Esto porque al recargar en alguna de las tiendas habilitadas para ello, se entrega la dirección de dicha tienda. Gracias a esta información se puede tener una estimación más precisa de los lugares frecuentados por el usuario (trabajo, casa, etc.).

Por ejemplo, intentamos estudiar el número de tarjeta siguiente al de un miembro del equipo. Así, pudimos saber que la tarjeta fue comprada en la misma estación de metro y deducir que su dueño vive cerca del miembro del equipo, pues repetidas veces recargó su tarjeta en una tienda muy cerca. Luego, con nuestro *script* que guarda y limpia los datos de los últimos tres meses de una (o más) tarjetas, pudimos computar fácilmente los siguientes datos:

Información general sobre el uso de la tarjeta:

- Número actividades durante los últimos tres meses; tarjeta usada o no usada.
- Número total de actividades en toda la vida de la tarjeta (no solamente durante los últimos tres meses); tarjeta con mucho uso, tarjeta con uso parcelario.
- Número de validaciones en el metro y/o en el bus; tipo de transporte público usado.
- Línea de bus más usada; en qué línea de bus es posible encontrar a esta persona. Debemos tener en cuenta que los recorridos de los buses están disponibles en líneas.[24]
- Número de carga de la tarjeta; dónde y cuándo la persona carga su tarjeta.

---

23 Ver https://es.wikipedia.org/wiki/Atentado_de_Santiago_de_Chile_de_2014#Tarjeta_Bip

24 http://www.transantiago.cl/mapas-y-recorridos/conoce-los-recorridos

- El valor de las recarga y el promedio.

Recorridos usuales, computando los patrones de movimiento:

- Ubicaciones más frecuentes (estación de metro-hora del día). Con esto es posible saber si durante los últimos tres meses, por ejemplo, una persona X tomó el metro a la estación A entre 8am y 9am, 34 veces; y 28 veces a la estación B entre 6pm y 7pm.

Del mismo modo se puede hacer seguimiento a las líneas de bus.

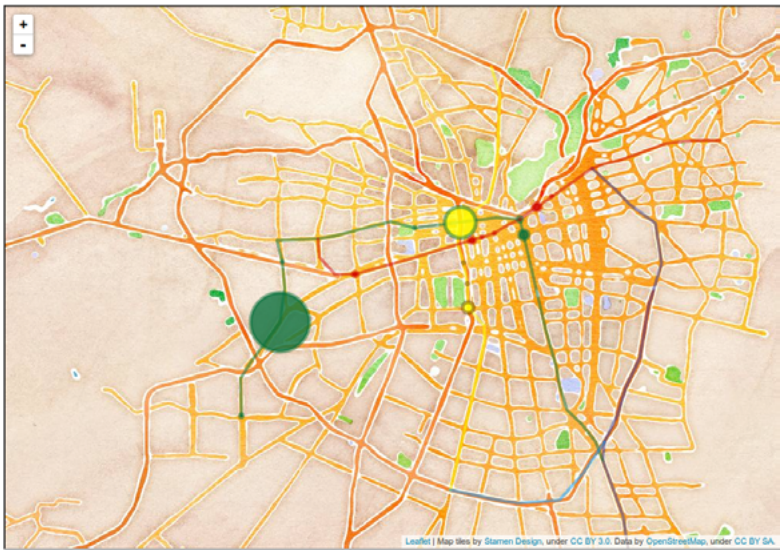
Recorridos inusuales, seleccionando algunos datos aislados, a partir de la validación de la tarjeta:

- La estación de metro menos usada.
- La hora menos frecuente.
- La línea de bus menos usada.
- La carga recarga de saldo más grande, si la cantidad es superior a 1.5 veces la carga media.

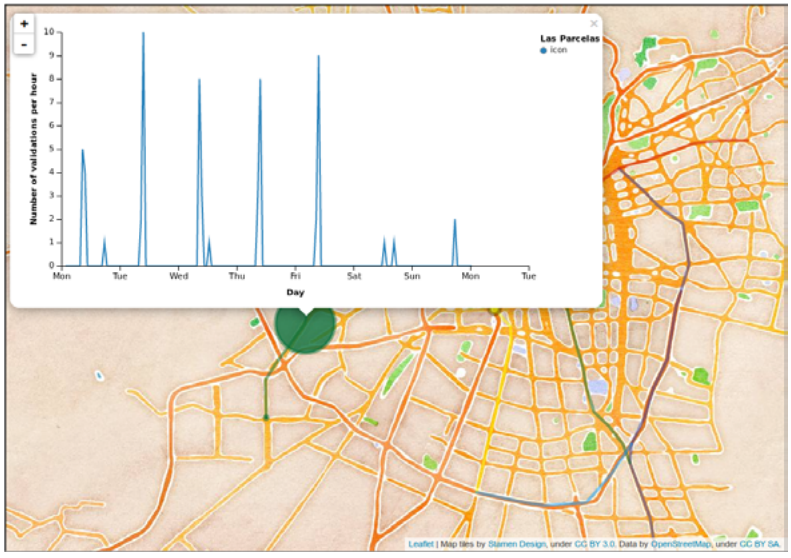
Visualizaciones

A continuación dos visualizaciones basadas en la línea de metro de la ciudad de Santiago de Chile tomando los datos de una tarjeta Bip!, que nos permiten visualizar recorridos usuales y extraer información tal como:

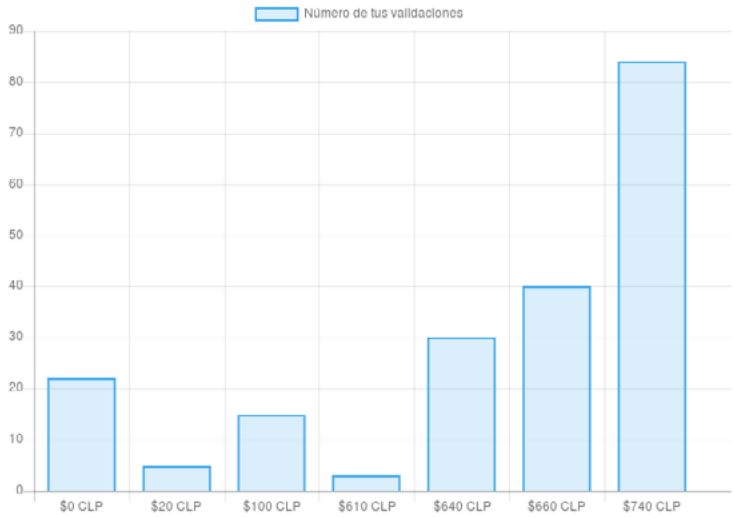
¿Dónde toma el metro?
Este usuario transita habitualmente por línea 1 (en rojo) la línea 5 (en verde) y la línea 2 (en amarillo)
El tamaño del círculo indica la frecuencia con que toma el metro una determinada estación. La estación que más usaba era "Las Parcelas", seguida de Santa Ana.



¿Cuándo?
Por cada estación de metro, es posible visualizar la serie temporal de una semana típica. Así se puede ver cuando la persona está en cada lugar.



¿Cuánto dinero gasta en transporte público?
Histograma de dinero invertido en el uso de transporte público



Con estos datos es posible saber muchas cosas sobre la gente y, casi automáticamente, sobre los hábitos, los lugares favoritos, las actividades profesionales o personales, entre otra información que puede ser relevante. A partir de las pruebas que realizamos, pudimos constatar que la computación de estos datos muestra cerca a qué estación de metro vive una persona, o trabaja o frecuenta. Además los movimientos son muy representativos del estilo de vida de alguien. Se puede inferir si la persona tiene un trabajo o no, si le se gusta salir tarde en la noche o en fin de semana, por ejemplo.

Es importante aclarar que utilizamos solamente estadísticas descriptivas para analizar estos datos, pero podemos pensar que es posible extraer más información personal sobre alguien, utilizando herramientas estadísticas más complejas.

Hoy en día los sistemas de transporte público modernos permiten recolectar una variedad de datos sobre sus usuarios: dónde viven, dónde trabajan, los recorridos diarios, los recorridos inusuales, etc. Por separado, estos podrían parecer irrelevantes, pero con un poco de análisis es posible obtener información sensible sobre los hábitos de una persona.

5. Conclusiones

El problema de privacidad causado por la exposición de los movimientos asociados a la tarjeta TNE de los estudiantes, demuestra la necesidad de una reflexión y un análisis minucioso del manejo de datos personales por parte de las instituciones públicas.

El transporte público es parte de la vida diaria de muchísimas personas, por lo que es importante no exponerlas a riesgos innecesarios, que pueden ser evitados aplicando buenas prácticas para manejar datos.

Además, este caso muestra que el sistema de datos es más critico cuando la tarjeta es nominal. Mantener el anonimato de la tarjeta es una buena solución para limitar el impacto sobre las personas en caso de falla o error.

Recientemente nos hemos enterado del cambio planificado del sistema de tarjeta de Transantiago.[25] Derechos Digitales espera que el ente regulador ponga atención a la privacidad de los usuarios del nuevo sistema.

Esperamos que las nuevas tarjetas sigan ofreciendo la posibilidad de viajar de manera anónima. Además, queremos insistir sobre el punto de que es mucho más fácil de diseñar un bueno manejo de datos desde el inicio que de cambiar un sistema vigente.

Alentamos a los responsables del diseño del nuevo sistema a tener en cuenta la privacidad de sus usuarios y aprender de los errores del antiguo sistema que están enumerados en este reporte.

Los resultados de este reporte fueron presentados en el marco de la Primavera Hacker, el 5 de noviembre de 2016 en Santiago de Chile. Durante la presentación nos concentramos en los diferentes peligros descritos aquí, incluyendo las demostraciones prácticas de los análisis que realizamos utilizando los *scripts* que creamos.

Algunos detalles han sido omitidos para proteger la privacidad de los usuarios.

25 http://www.emol.com/noticias/Nacional/2016/10/25/828123/Nueva-tarjeta-bip-del-Transantiago-permitira-comprar-paquetes-de-viajes-a-menor-precio.html


DERECHOSDIGITALES
Derechos Humanos y Tecnología en América Latina