Entendiendo la privacidad hoy

Alejandro Hevia
Profesor Asistente, DCC, Universidad de Chile; Ph.D. Computer Science, University of California, San Diego (2006); Ingeniero Civil en Computación, Universidad de Chile (1998). Director del Grupo de Respuesta a Incidentes de Seguridad Computacional, CLCERT.
ahevia@dcc.uchile.cl

En mi época de estudiante de computación, uno de mis profesores tenía colgada en su puerta una caricatura que mostraba dos perros hablando, uno de ellos sentado frente al computador: *"On Internet, nobody knows you are a dog" (o "En Internet, nadie sabe que eres un perro".)* le decía un perro al otro [22]. Para quienes vivieron online esa época, el chiste era claro: en Internet era fácil pretender que se era alguien o algo distinto pues nuestras comunicaciones no tenían mecanismos de autentificación de ningún tipo: tu email o tu página Web era para todos los efectos prácticos indistinguible de la de millones de otras personas.

Hoy en día, la situación no ha cambiado demasiado y todavía no hay autentificación de ningún tipo[1]. Sin embargo, nadie se reiría del chiste anterior. ¿Por qué? El concepto de "quién soy" en Internet ha sido redefinido por el colectivo denominado "Web 2.0", las redes sociales y sistemas de contenido generado por usuarios. "En Internet, tus amigos y seguidores de Facebook, Google y Twitter saben a qué le ladras, cuándo ladras, con quién ladras, qué comes y por dónde paseas. Es irrelevante si eres un perro o no". Probablemente con este título la caricatura ya no sería tan graciosa.

Creemos que el anonimato es prevalente online, cuando de hecho no lo es, o al menos, es muy difícil ser realmente anónimo para un usuario honesto promedio. Es vox populi que en los últimos años, grandes bases de datos con información de cada uno de nosotros han sido amasadas con nuestro conocimiento e incluso con nuestro beneplácito. Esta tendencia de hecho parte fuera de Internet ofreciéndonos beneficios a cambio de un poco de información.

1 A pesar de los deseos de Google, quienes en su red social Google+ inicialmente han restringido el acceso a quienes firmen con "nombres reales" y no seudónimos, algo que ha causado gran controversia y probablemente será cambiado.

Aceptamos la existencia de DICOM como una manera de mejorar el acceso al crédito, o aceptamos revelar nuestra lista de compras a cambio de un beneficio económico (puntos del supermercado). Ya en el mundo online, por un beneficio económico y/o de entretención por ejemplo cedemos gustosamente nuestros datos a cambio de una cuenta gratis en Facebook, donde podemos ver a nuestros amigos y dejarnos ver por ellos. Así, DICOM sabe quiénes somos, a quiénes no les pagamos, y por cuánto. La cadena de supermercados, por otra parte, sabe quiénes somos, qué compramos y cuándo. Facebook, sabe algo más importante: qué tipo de cliente somos, qué queremos y qué quieren nuestros amigos. Y sorprendentemente, toda esta información está en manos de privados. En comparación, el Estado pareciera saber en mucho menos: el Servicio de Impuestos Internos (SII), uno de los mayores recolectores de datos estatales en Chile, sabe esencialmente sólo cuánto ganamos, aunque seguro querría saber aún más. Cierto, los organismos de seguridad estatales legalmente pueden saber mucho más: nos dejamos escanear con rayos X o registrar en los aeropuertos (¡y sorprendentemente aún en las calles!), e incluso nos dejamos escuchar, fotografiar y filmar en la vía pública por el beneficio de la "seguridad" que la policía nos puede proveer en ese ambiente. En suma, aceptamos (incluso aplaudimos) las reglas del juego para obtener los supuestos beneficios. Sin embargo, a muchos todavía les incomoda o incluso se quejan cuando ven reportajes sobre las injusticias que datos incorrectos en DICOM pueden causar, o se quejan de la pérdida de privacidad cuando en las noticias aparecen hackers comprometiendo sitios Web y revelando datos de millones de clientes, o cuando se hacen públicas conversaciones por celular entre congresistas (y más recientemente, entre defendidos y sus abogados), o se publican datos o fotos íntimas de personas sin su consentimiento en portales públicos online. ¿Cuánto realmente queremos tener privacidad? ¿O ya la perdimos? ¿Vale la pena el (aparentemente necesario) costo monetario y social de recuperarla y mantenerla?

¿QUÉ ES PRIVACIDAD?

Entender privacidad es difícil[2]. Según Van der Berg [24], es "probablemente uno de los conceptos más complicados, malentendidos y altamente debatidos en ciencias sociales, en ámbitos legales, filosóficos, y tecnológicos, durante las últimas décadas en el mundo". Esto quizás porque la definición de privacidad no es trivial de entender. Por ello, ruego al lector me permita una sección algo más teórica del tema, con el compromiso de aplicarlo a temas tecnológicos más adelante.

Según Warren y Brandeis [27] privacidad es "the right to be let alone" ("el derecho a ser dejado tranquilo o solo"). Esta definición es notable por dos razones. Primero, por lo adelantada a su tiempo: en ese entonces, a dichos autores les preocupaba la aparición de la fotografía y de las grabaciones como herramientas periodísticas. Y segundo, por tocar un nervio en casi todos los seres humanos: el valor de nuestra individualidad. Sin embargo, es una definición limitada, pues no da demasiadas luces de qué implica, por ejemplo, cuando decidimos dejarla de lado voluntariamente (esto es, cuando a propósito buscamos "dejar de estar solos" y nos contactamos con otros). Una segunda definición la da Burgoon y otros (citado en [17]): "La habilidad para limitar físicamente, vía interacción, sicológicamente e informacionalmente el acceso a mi individualidad o la individualidad de un grupo". Esta definición pone énfasis en las distintas dimensiones de la privacidad: existe en términos físicos (por ejemplo alguien mirando por mi ventana); existe al interactuar con gente (uno la considera al conversar con otra persona); existe en su rol sicológico (expresada en la libertad de tomar decisiones personales, religiosas, de orientación sexual sin presión de otros); y puede dejar de existir cuando la información acerca de una persona es transmitida a otros sin su consentimiento o conocimiento. Una perspectiva distinta la entrega Hildebrant [10] quien la define como "la libertad o carencia de limitaciones irrazonables sobre la construcción de mi propia identidad". Esta definición nos recuerda la íntima relación entre privacidad e identidad. Aquí, una información es privada o sensible si "revela algo respecto a quién soy". Privacidad es algo dinámico y abierto, algo que, tal como nuestra identidad, puede cambiar en el curso de nuestra vida. Finalmente, la definición de Westin [29] se focaliza en un aspecto más preciso, el control del acceso a la información sobre la individualidad de una persona: "La necesidad de individuos, grupos o instituciones de determinar por ellos mismos, cuándo, cómo y hasta qué nivel de información acerca de ellos es comunicada a otros".

Ahora bien, de las decenas de distintas definiciones de privacidad propuestas en las últimas décadas -muchas muy amplias o muy restringidas-, hay una en particular que sobresale en el contexto de tecnologías computacionales: privacidad como "integridad de contexto", propuesta por Nissenbaum [15,16]. Su punto de partida es la existencia de los distintos "mundos" o contextos donde la gente se mueve. *"Al observar la textura de la vida de las personas, vemos que ellos salen y entran, y se mueven dentro de una pluralidad de ámbitos o mundos distintos. Están en sus casas con sus familias, van a trabajar, buscan atención médica, visitan amigos, consultan siquiatras, hablan con abogados, van al banco, asisten a misa, votan, compran, y mucho más. Cada una de dichas esferas, ambientes o contextos involucra, o incluso es definida por, una serie de normas, las cuales gobiernan sus distintos aspectos tales como los roles, esperanzas, acciones y prácticas" [16].* Nissenbaum argumenta que la privacidad de una persona se desprende de su habilidad de compartimentalizar su vida (social), de manera que la información sobre ella, que pudiera ser dañina o vergonzosa cuando es compartida fuera del contexto donde se entregó, se mantenga protegida, circunscrita a las reglas del contexto donde se entregó. Por ejemplo, la gente usualmente no considera privada la información compartida con su doctor en el proceso de obtener atención médica, pero pudiera molestarse si dicha

2 La presentación de las definiciones, historia y varios ejemplos del presente artículo se basan en el excelente reporte titulado "Privacy Enabled Communities" del Primelife Project [18] de la Comunidad Europea.

El problema surge cuando la información es compartida fuera del contexto original, sin preservar las normas (explícitas e implícitas) referentes a cómo se comparte información en ese contexto.

información va a parar a su empleador o a un conocido del colegio de sus hijos. Esta distinción es notable, pues en otras palabras, los datos no son privados o públicos *per se,* sino que lo son sólo respecto a un contexto. No es el contenido de la información, sino el contexto donde fue compartido y, en particular, la audiencia que tiene acceso a dicha información. Esto explica la aparente contradicción vista en personas que comparen información íntima y personal sin mucha preocupación, al tiempo que se sienten profundamente perjudicados cuando ven compartida otra información de ellos aún si esta última no es personal ni sensible. Más aún, entender privacidad bajo la definición de Nissenbaum nos enseña que privacidad es una característica social y no sólo informacional. Compartir información *per se* no es el problema; muchos de nosotros compartimos información con otros todo el tiempo sin estar necesariamente preocupados de las repercusiones de esta conducta en términos de privacidad. Sin embargo, el problema surge cuando la información es compartida fuera del contexto original, sin preservar las normas (explícitas e implícitas) referentes a cómo se comparte información en ese contexto. De hecho, Nissenbaum lo resume en forma excelente: *"Si un ítem de información es considerado apropiado para una situación en particular, típicamente es compartido sin problemas. Más aún, la información puede ser guardada o usada en una situación en particular sin gatillar objeciones de ningún tipo. La gente no objeta el tener que entregarle a los doctores los detalles de su condición médica, o discutir los problemas de sus hijos con los profesores de sus hijos, ni divulgar información financiera a un ejecutivo de cuentas para pedir un préstamo, o compartir con amigos cercanos los detalles de una relación romántica. Para el sinnúmero de transacciones, situaciones y relaciones en las cuales la gente se aboca, hay normas -explicitas e implícitas- que gobiernan cuánta información y de qué tipo es apropiada para ellos. Donde dichas normas son respetadas, diremos que la integridad de contexto se mantiene. Donde no, diremos que la integridad de contexto ha sido violada"* [15].

DE VUELTA A LA PRÁCTICA

Entender privacidad como mantener la integridad de contexto de (por ejemplo) la información asociada a nosotros mismos, puede ayudarnos a entender nuestro comportamiento ante los eventos mencionados al comienzo. Por ejemplo, respecto a los datos en DICOM, la regla implícita (supuesta) es que dichos datos serán fieles representantes de nuestro crédito, que serán preservados (no modificados arbitrariamente), y que serán comunicados apropiadamente (por ejemplo sólo a quienes debieran revisar nuestro crédito y no a otras entidades que aparentemente no debieran necesitarla, como un futuro empleador, o mis amigos o mi doctor). Datos incorrectos en DICOM son una violación a la privacidad en el sentido que vulneran una parte de esta regla, la integridad de los datos en particular. En el caso de las fotos privadas reveladas online, claramente dicha información ha cruzado a un contexto para el cual nunca fueron orientadas (lo cual explica por qué la gente sigue tomándose este tipo de fotos, aún cuando son aconsejados de lo contrario). Dichas fotos tuvieron su contexto, íntimo quizás, en el cual no se consideraban privadas. Y en el caso de los petabytes de información en Facebook, la mayoría de los usuarios implícita o explícitamente creen que ciertas normas se respetarán en el contexto formado por sus amigos. O bien, no entienden cuál es ese contexto. En muchos casos, hay una desconexión con la realidad. Por ejemplo, aunque técnicamente sea posible que un futuro empleador mire las fotos compartidas de una persona buscando trabajo, frecuentemente dicha persona no considera como una violación de su privacidad el compartirlas, pues cree (sin razón) que dichas fotos sólo permanecerán en el contexto (real o imaginado) de sus amigos. Obviamente, esa persona siente su privacidad violada cuando aparecen en el computador del entrevistador o, peor aún, ventilados en una reunión familiar.

PRIVACIDAD VERSUS SEGURIDAD

Lamentablemente, es frecuente ver discusiones donde privacidad es presentada como algo transable por seguridad pública o el bien común. Gobiernos de todo tipo frecuentemente nos quieren convencer que para garantizar la seguridad pública es necesario recolectar datos de los ciudadanos en forma de monitoreos masivos y extensa minería de datos. Se argumenta que los ciudadanos honestos no debieran temer, pues "quien nada hace nada teme".

Éste es un falso dilema (discutido por muchos, entre ellos Solobe [20,21]), pues presupone que privacidad sólo existe para "esconder cosas malas". El argumento es que gente honesta y sana, no oculta nada, sólo aquellos en los límites de la decencia o la legalidad lo hacen. Con ello se justifica plenamente reducir la privacidad de todos para buscar a las "manzanas podridas". Ejemplo de ello es la instalación de cámaras en las calles de las ciudades, para detectar y perseguir delitos, o, por ejemplo, en la instalación de

escáner de cuerpo en los cruces fronterizos del norte para detectar traslado de droga. Sin embargo, pese a que los objetivos puedan ser loables, esta dicotomía es falsa. Primero, si privacidad es integridad de contexto, todos tenemos algo que ocultar (¿por qué no vivimos en casas de vidrio?). No quiero que mi información médica deje la consulta de mi doctor o que el monto de mis ahorros lo sepa mi vecino. Como vimos, qué es privado depende del contexto: hay veces que información muy íntima (la última cirugía plástica de una actriz famosa) no es considerada privada e información muy pública (su número de carnet) sí lo es. La dicotomía entre privacidad e ilegalidad, por un lado, y seguridad y legalidad, por otro, simplemente desconoce la percepción de la privacidad de la gente en el mundo real. Peor aún, tal argumento (privacidad sólo es para quienes quieren ocultar algo malo) no considera aspectos más fundamentales, como la necesidad de su existencia para tener una sociedad realmente democrática y libre. Valorar y respetar la privacidad permite a la gente hablar libremente, sin temor a manifestar ideas contrarias al credo imperante, y finalmente relegarse, si es necesario, a un espacio que pueden llamar privado, donde la intervención (estatal o privada) no es permitida. Por ende, argumentar que la privacidad de los ciudadanos debe ser disminuida en favor de su (supuesta) seguridad pública va en contra de valores intrínsecamente democráticos. En palabras de Kee Hinckley, criticando la falta de seudoanonimato en la red social Google+: *"El foro de discusión pública ya no es la plaza del pueblo, el diario ni la calle. Es aquí, en Internet, y está sucediendo en comunidades como ésta, hosteadas por compañías del sector privado"*.

Por supuesto, el hecho que mis fotos íntimas o el vídeo de la cámara de vigilancia donde aparezco saliendo de una tienda de ropa látex para adultos aparezca publicado en Facebook o Youtube no tiene nada que ver con la validez de mis opiniones políticas. Sin embargo, todos sabemos que tal hecho ciertamente puede provocar un daño serio a mi reputación, indirectamente descalificándome como interlocutor válido. El punto a recordar es que no sólo debemos pedirle a entidades privadas respetar la integridad de contexto (mi privacidad) sino también a entidades públicas. Este tema de la relación entre privacidad y libertades democráticas es largo y probablemente requiera su propio artículo.

PRIVACIDAD Y TECNOLOGÍAS MODERNAS: ¿QUÉ CAMBIÓ?

Las tecnologías de información han posibilitado como nunca la recolección, manipulación, distribución y mantenimiento de información en una escala masiva. Al usar empresas como Facebook, Google o el Servicio de Impuestos Internos como intermediarios en muchas de las acciones de nuestra vida diaria, les hemos permitido recolectar dicha información a escalas sin precedentes. En palabras de Vint Cerf, uno de los creadores de Internet: *"Nunca en la historia de la humanidad hemos tenido acceso a tanta información tan rápido y tan fácilmente"*. Tal información puede agregarse, copiarse, enlazarse, correlacionarse y distribuirse en forma barata y masiva. En comparación, antiguamente, lograr recuperar información confiable – o armar un "dossier"- respecto a una persona era una labor investigativa mayor: el "investigador" debía visitar hospitales, escuelas, oficinas, municipalidades, bancos, casas, iglesias, etc., todos aquellos lugares donde la persona había estudiado, trabajado, interactuado y vivido. En cada lugar, el investigador debía conversar con quien estuviera en control de los datos y convencerlo de lo apropiado de compartirlos, justificando de paso su autoridad para solicitarlos. Luego debía hacer copias manuales (con suerte fotocopias) de los documentos con los datos. Hough [9] (citado en [18]) argumenta que *"por ineficiente que pareciera, el almacenamiento de registros en papel, en ubicaciones diversas, en realidad creaba un colchón de protección, asegurando que los datos no fueran revelados sin un esfuerzo considerable y sólo con una causa justa"*. No sólo servicios "gratuitos" recolectan esta información, también hay casos emblemáticos de empresas que derechamente los venden. Solove en su libro "The digital person" [20] reporta a una compañía llamada "Regulatory DataCorp" (RDC) la cual ha "creado una base de datos masiva para investigar gente que abre nuevas cuentas bancarias" y que en su base de datos la información es recolectada "desde más de veinte mil fuentes distintas en el mundo". Es discutible cuánto de esa información es estrictamente necesaria para

evaluar el riesgo de un potencial cliente y cuánto de ella simplemente es un "dossier" de la persona.

Quizás todo esto fue lo que llevó a Scott McNealy, ex CEO de Sun Microsystems, a argumentar *"you have zero privacy anyway. Get over it" ("tienes cero privacidad de todas maneras. Resígnate.")* ¿Es cierto que hemos perdido toda privacidad? McNealy fue altamente criticado por su postura pues aunque correcta en los hechos, fallaba en la reacción: resignación no es la acción adecuada. Ustedes y yo podemos haber perdido nuestra privacidad online, pero ¿nuestros hijos y nietos deben también perderla?

Paradójicamente, las características que más beneficios han traído a la manera de organizar la información en el mundo son aquellas con el mayor riesgo de perjudicar nuestra privacidad. En el proceso de combinar, mover, copiar y editar información desde lugares dispares es que violaciones de integridad de contexto pueden producirse. De hecho, Nissenbaum [15] distingue dos niveles de problemas: (1) los tipos de violaciones ocurridos en el proceso de mover información de un contexto a otro, y (2) aquellos ocurridos en el proceso de combinar distintos trozos de información. Ejemplos del primer tipo son las clásicas violaciones producidas al copiar información médica o financiera a otros contextos donde su uso no fue contemplado. Situaciones del segundo tipo ocurren, por ejemplo, cuando compañías de seguro solicitan exámenes médicos creando perfiles sobre los clientes con información irrelevante al simple análisis de riesgo, pero que vulneran su privacidad. El problema puede ocurrir también por subestimar las capacidades actuales de "enlazar" datos. Es el caso de Netflix, compañía que en 2007 reveló su base de datos de recomendaciones de películas hechas por más de 500 mil clientes, con la esperanza de obtener mejores sistemas de recomendación. Y aunque dijeron tomar especial cuidado de "anonimizar" los datos (eliminando datos personales y reemplazando nombres con identificadores al azar), Narayanan y Shmatikov, dos investigadores de la University of Texas at Austin [14], mostraron que podían "desanonimizar" a muchos de los clientes simplemente comparando información de rankings y fechas/horas con aquellas disponibles en Internet Movie Database (imdb.com), un sitio de recomendación de películas donde los usuarios sí entregan sus nombres. Éste y otros casos similares (como cuando gente fue individualizada a partir de logs "anonimizados" publicados por AOL en 2006 [19], o experimentos que mostraron que el 80% de los ciudadanos en Estados Unidos pueden ser individualizados a partir de los datos del censo, su código de área, su género y su fecha de nacimiento [23, 8]) mostraron que los mecanismos básicos para anonimizar datos no son suficientes para prevenir violaciones de privacidad.

De hecho, en el contexto de recopilación masiva de datos, hay quienes argumentan la necesidad de recuperar nuestro "derecho a ser olvidado" (ver Solove [20]). Biológicamente, nuestra memoria nos ayuda a recuperarnos y comenzar de nuevo, luego de errores o experiencias traumáticas simplemente dejándonos olvidar dichos eventos. *"Olvidar es la norma, recordar es la excepción"*, pero con herramientas ya disponibles podemos perfectamente terminar en *"olvidar es la excepción y recordar es el default"* [12]. ¿Cuántos de nosotros podríamos vivir teniendo "un registro detallado y públicamente consultable" [21] de nuestras acciones, datos y vida desde nuestra infancia? Tal como argumenta Solove [21], no conviene olvidarnos de las ventajas sociales de poder "comenzar de nuevo" y "partir de cero" al considerar la lista de requisitos de nuestra vida digital.

REDES SOCIALES Y REPUTACIÓN

Un tema aparte lo constituyen los posibles problemas de privacidad (y de reputación) derivados del contenido generado por otros usuarios en redes sociales. La identidad de cada usuario no está sólo definida por información entregada por el mismo usuario (su foto, su nombre, sus gustos) sino por información entregada por otros, que puede perfectamente permanecer inmutable en el tiempo pese a ser incorrecta o, aún siendo incorrecta, perjudicar/humillar públicamente a alguien. Un caso emblemático de esto, es el llamado caso de la *"the dog poop girl" o "niña del excremento de perro"* ocurrido en Corea del Sur en 2005. Allí, una mujer en el metro de Seúl se rehusó a recoger el excremento de su perro, lo cual fue fotografiado por otro pasajero. Rápidamente, la foto empezó a circular ampliamente por las redes sociales en el mundo. Luego, su nombre y detalles personales fueron revelados por otras personas en represalia. Al final, su reputación fue arruinada y abandonó la universidad.

¿Podría haber pasado lo mismo sin el apoyo de la tecnología moderna (redes sociales, Internet)? Los rumores y habladurías han existido desde siempre, pero históricamente su alcance ha sido limitado, típicamente al grupo donde se generan. Es sólo con el advenimiento de las redes sociales de gran escala donde este tipo de casos terminan alcanzando audiencias de millones de personas. Se puede argumentar que el supuesto "anonimato" detrás de un nombre de usuario disminuye la inhibición de su dueño y lo hace menos socialmente conciliador en sus comentarios y críticas, pero estudios muestran que comentarios negativos de este tipo aún surgen cuando el autor realiza sus comentarios en forma pública y completamente identificado[3].

Concluyo esta sección poniendo énfasis en cómo las redes sociales cambian el tipo de problemas de privacidad existentes. Por bastante tiempo, mucho esfuerzo técnico (y legal) fue puesto en desarrollar mecanismos de control de información para limitar la filtración de información almacenada y procesada por parte de empresas y organizaciones. Sin embargo, hoy en día poco de ello es aplicable al escenario social, pues en este contexto, quienes comprometen la privacidad de un usuario no son las empresas ni las organizaciones, sino otros usuarios del sistema. Y aunque mecanismos técnicos que nos permitan "contar un secreto" sin temer a la falta de discreción de nuestro confidente en teoría son posibles [18] (la mayoría derivados de "Zero Knowledge", una técnica criptográfica

3 Quizás porque el medio desconecta al autor de su "víctima", algo mucho menos frecuente en comentarios cara a cara.

bellísima y elegante pero poco práctica), no es claro que sean efectivos: muchos de estos esquemas se basan en evitar "generar evidencia" que soporte la indiscreción (el indiscreto no puede probar su aseveración). Es claramente cuestionable si la falta de evidencia ha disminuido la distribución de información incorrecta (pero "jugosa") alguna vez en Internet.

> Gobiernos de todo tipo frecuentemente nos quieren convencer que para garantizar la seguridad pública es necesario recolectar datos de los ciudadanos en forma de monitoreos masivos y extensa minería de datos.

BIG BROTHER, MANY LITTLE BROTHERS

Hoy dejamos muchos rastros "sin movernos del escritorio". Sitios Web nos monitorean en forma distribuida para darnos el dudoso beneficio de mejores avisos (más focalizados, target advertising) o productos gratis. Pero, ¿es tan así? ¿Somos realmente monitoreados? En palabras de Andrew Lewis: *"Si no estás pagando por algo, entonces no eres el cliente; eres el producto en venta"*. Dado lo extenso de este tema, simplemente le sugiero al lector testearlo por sí solo: para saber quiénes lo siguen diariamente, le recomiendo usar un par de días el plugin "collusion" para Firefox [25]. ¡Los resultados son sorprendentes! El nivel de colaboración entre sitios distintos en Internet sólo con el propósito de seguir usuarios es abismante. Aún con esa evidencia a mano, a mucha gente no le importa. Quizás han comprado la excusa de que tal comportamiento es "el precio de obtener productos gratis". El problema más bien pareciera ser el desconocimiento o la falta de concientización respecto a "las posibilidades técnicas para recolectar, guardar y procesar datos acerca de esta persona" [26]. De hecho, estudios muestran que la mayoría de la gente, aunque expresa preocupación por su privacidad online, lo que entiende por "privacidad" es poco claro. Típicamente se refiere "a temores diversos en la red como por ejemplo encontrarse con virus, troyanos y programas espías, atraer spam o ser atacados por un hacker" (ver Paine [17]). Peor aún, Paine reporta que aún aquellos con mejor entendimiento de las amenazas carecían de las herramientas efectivas para protegerse. De hecho, ¿qué es posible hacer para protegerse hoy? Lamentablemente poco, pues los incentivos económicos están en recolectar información. Por ejemplo, para cada nueva técnica de limitar o borrar las "cookies" usadas para seguirnos, surge una nueva técnica para saltarse tal limitación [11]. Tímidamente, por otra parte, iniciativas legales como "Do-Not Track List" [7,28] o sus mecanismos asociados ("Do-Not-Track headers" [6,13]) han tomado fuerza, pero su efectividad está por verse.

Otros tipos de seguimiento "offline" más clásicos, como por ejemplo con cámaras públicas en las calles, se han masificado en las últimas décadas. Para que decir, el sinnúmero de veces que somos captados digitalmente por turistas o cámaras privadas al movernos por la ciudad. Aquí, en términos de privacidad quizás el tema es algo más claro, pues la pregunta no es simplemente si tal desarrollo es deseable o no, sino si tal desarrollo es conveniente para nuestra sociedad. Por un lado podríamos plantearnos que potencialmente el costo del tracking pudiera ser lo suficientemente bajo como para justificar económica y socialmente los beneficios (búsqueda eficiente en la Web, correo y repositorios de vídeos gratuitos con Google, o disminución del crimen en las calles). Pero, la evidencia muestra lo contrario. Aunque el costo directo es pequeño, el costo indirecto es alto comparado con los beneficios. Por ejemplo, en Inglaterra desde hace unos años se cuestionan si el gasto en cámaras es justificable económicamente en términos de disminución efectiva de delincuencia[4], sobre todo considerando los altos costos de estos sistemas [1]. Por otro lado, ni las políticas públicas ni la ley parecieran hacer mucho para disminuir el problema de los miles de "little brothers" que capturan nuestra imagen (física) sin pedirnos permiso. El tema en particular no es simple. Nos gustaría que limitaran legalmente la captura de imágenes de nosotros en la vía pública, pero tal limitación hoy en día es difícil de asegurar técnicamente. A nadie le gustaría, por ejemplo, ser demandado porque la cámara de seguridad de su casa tomó la foto del vecino al pasar.

Volviendo a las redes sociales, los "little brothers" aquí son conocidos: mis amigos y familiares son posiblemente mi mayor fuente de problemas de privacidad. Hoy en día, son ellos, nuestros amigos, conocidos y familiares quienes revelan la mayor cantidad de información respecto a nosotros. Y aunque algunos mecanismos tecnológicos (criptográficos) permiten limitar quiénes pueden acceder a datos sensibles (como el plugin Scramble! para Facebook [2]), el tema está en su infancia.

PRIVACIDAD Y BIG DATA

Un aspecto interesante en la discusión de privacidad es cómo se inserta en la discusión del "Big Data": esta plétora de "masivas cantidades de información generadas por (y acerca de) gente, cosas y sus interacciones" a las cuales "computines, físicos, economistas, matemáticos, cientistas

4 Y no simplemente de una adaptación de la delincuencia a la ubicación y método de uso de las cámaras.

> Estudios muestran que la mayoría de la gente, aunque expresa preocupación por su privacidad online, lo que entiende por "privacidad" es poco claro. Típicamente se refiere "a temores diversos en la red como por ejemplo encontrarse con virus, troyanos y programas espías, atraer spam o ser atacados por un hacker".

políticos, bioinformáticos, y sociólogos están reclamando desesperadamente acceso" [3]. Y aunque hoy son comunes las discusiones acerca de los pros y contras de usar las grandes bases de datos de Twitter, Google, Facebook, Wikipedia y cualquier otro donde la gente deja rastros, para resolver problemas relevantes a nuestra sociedad (por ejemplo, si "¿la disponibilidad de mejores técnicas de análisis permitirá dar acceso más eficiente a información efectiva a la gente? ¿O será usada para monitorear a manifestantes en las calles?" como señala Boyd [5]) es poca la discusión clara respecto a cómo proceder como sociedad en forma integral en este tema. Según Dana Boyd, "Big Data se ve como pura oportunidad: agencias de marketing como un medio para avisos focalizados más efectivos, agencias aseguradoras como una manera de optimizar sus ofertas, y bancos como una manera de interpretar mejor un mercado complejo". Sin embargo, tal discusión se realiza en un ambiente dinámico, donde "la cantidad de almacenamiento no tiene una cota superior clara y donde las decisiones que se tomen hoy pueden impactar seriamente nuestro futuro" y nuestra privacidad en él [3].

Boyd y Crawford [3] insisten en un punto: hoy debemos preguntarnos qué significa tener acceso a estos datos, quiénes tienen acceso, cómo se establece este acceso y con qué fin. Para ello proponen preguntarse en forma crítica respecto al fenómeno "Big Data", sus supuestos y sus potenciales "biases" o precondicionamientos. Interesantemente, uno de sus cuestionamientos surge del tema privacidad: *"Just because it is accesible doesn't make it ethical" (o "solo porque sea accesible no significa que sea ético (accederlo)")*. Los autores cuestionan la libertad con que investigadores publican análisis de datos que, aunque disponibles online, nunca las personas a quienes se refieren accedieron a tal uso. Como ejemplo mencionan un proyecto de 2006 el cual analizó 1.700 perfiles de Facebook (aparentemente públicos) recolectados desde una "universidad norteamericana del noreste" para hacer un seguimiento de los estudiantes por varios años [30]. El estudio utilizó perfiles "públicamente asequibles" de estudiantes de una universidad (luego identificada como Harvard por terceros). Sin embargo, tales perfiles fueron recolectados por asistentes de investigación (estudiantes) de la misma institución, lo que cuestiona su calidad de perfiles públicos. Además, el proceso de anonimización fue cuestionado y con ello, surgieron quejas por la violación a la privacidad de los participantes del estudio[5]. En particular, Boyd y Crawford critican la falta de cuestionamiento de investigadores respecto a la admisibilidad de usar un conjunto de datos "públicos": *"¿Pueden ser simplemente usados sin pedir permisos? ¿Cuál debiera ser la norma ética que rige tales estudios? Las respuestas no son fáciles pues frecuentemente las violaciones de privacidad no pueden medirse en "daños" específicos al momento de publicarse los datos o incluso dentro de 20 años"* [3]. ¿Debieran los datos de un individuo ser incluidos en un conjunto de datos agregados? Por ejemplo, ¿qué tal si un comentario de un blog de una persona es tomado fuera de contexto y analizado públicamente en un estudio altamente publicitado, sin la persona saberlo? ¿Quién es responsable de que el (o los) dueño(s) de los datos no sean afectados al hacer análisis y publicar el estudio? ¿Qué significa que el dueño de un dato dé su consentimiento, sobre todo si es para cierto contexto? ¿Cambia el contexto dependiendo de los resultados del estudio? (Tal consideración me recuerda una broma televisiva en EE.UU. en la cual se les pedía a hombres jóvenes en la playa, dar un saludo para un vídeo de televisión, a lo cual accedían gustosos, pero luego, cuando se les indicaban que debían mandar saludos al "Gay Channel" muchos de ellos huían). La alternativa de solicitar consentimiento a cada uno de los dadores de los datos utilizados en un estudio/análisis es obviamente impráctica. Sin embargo, no se puede legitimar éticamente su uso simplemente porque los datos son asequibles. *"No porque los datos sean públicamente asequibles significa que fueron pensados para ser consumidos por cualquiera"* [4].

Boyd [5] sugiere a quienes deben analizar datos en el ámbito de "Big Data" los siguientes principios: (1) "Seguridad por 'oscuridad' es una estrategia razonable" (para quienes generan datos), lo que significa que la gente comparte sus datos aún sin mecanismos técnicos de protección efectiva bajo el supuesto implícito que *"nadie grabará públicamente esto y lo ventilara"*. Por lo mismo, Boyd propone respetar tal deseo atendiendo el contexto donde fue hecho. El principio (2) es que *"no todos los datos fueron hechos públicos pensando en que serían publicitados"*, lo cual debiera ser obvio; (3) *"Quienes publican información PII no necesariamente rechazan su privacidad"*, donde PII significa "Publicly Identifiable Information" o Información que identifica públicamente a su donante[6]. El principio (4) *"Agregar y distribuir datos fuera de contexto es una violación de privacidad"* debiera ser obvio para el lector a estas alturas, puesto que se sustenta en la justificación de privacidad como "integridad de contexto". Y finalmente, (5) *"Privacidad no es equivalente a control de*

5 Estos datos inicialmente fueron dados en el contexto que serían asequibles sólo para miembros de la universidad.

6 Segun Boyd, "PII se revela todo el tiempo en redes sociales. Lo que si quieren evitar es 'PEI', 'Personally Embarrasing Information' o Información qué avergüenza personalmente a su donante".

acceso". Nuevamente, este último principio es evidente si diferenciamos la "regla" (quién debe acceder a una información según el contexto) del "mecanismo" (quién puede acceder a la información según el mecanismo técnico empleado, el cual puede fallar o estar mal configurado).

Al lector interesado le recomiendo leer la transcripción de la presentación de Dana Boyd [5].

CONCLUSIÓN

La privacidad, en particular en un mundo digital, es un tema fascinante pues su mera definición no es trivial; sus implicaciones políticas, sociales y culturales pueden ser controversiales, pero su valor es inmenso. Como lo hacía notar en broma un amigo, es uno de los pocos temas donde abogados e ingenieros pueden tener una conversación de genuino interés mutuo. Ello pues la definición de sus límites y consecuencias, y qué mecanismos legales y tecnológicos disponibles pueden ser usados para protegerla no pueden sustraerse de las características de las comunidades mismas donde se intenta preservar. He allí el desafío: nuestra privacidad en el futuro no puede construirse en privado, debemos todos colaborar para lograrla. BITS

REFERENCIAS

[1] BBC report "1,000 cameras 'solve one crime'", Disponible en: http://news.bbc.co.uk/2/hi/uk_news/england/london/8219022.stm, 2009.

[2] F. Beato, M. Kohlweiss, K. Wouters, "Scramble! Your Social Network Data," in Proc. of the International Symposium on Privacy Enhancing Technologies (PETS), 2011.

[3] D. Boyd, K.Crawford, "Six Provocations for Big Data". Disponible en: http://ssrn.com/abstract=1926431, 2011

[4] D. Boyd, A. Marwick, "Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies," paper given at Oxford Internet Institute Decade in Time Conference. Oxford, England. [Citado por Boyd & Crawford, 2011].

[5] D. Boyd. 2010. "Privacy and Publicity in the Context of Big Data" (notas de charla), WWW 2010. Raleigh, North Carolina, abril 2010. Disponible en http://www.danah.org/papers/talks/2010/WWW2010.html

[6] Do Not Track Project, "Do Not Track - Universal Web Tracking Opt Out", Disponible en http://donottrack.us/

[7] Federal Trade Commision, "FTC Testifies on Do Not Track Legislation", Disponible en: http://www.ftc.gov/opa/2010/12/dnttestimony.shtm, 2010.

[8] P. Golle, "Revisiting the Uniqueness of Simple Demographics in the US Population", In proc. of WPES 2006. En http://crypto.stanford.edu/~pgolle/papers/census.pdf

[9] M.G. Hough. "Keeping it to ourselves: Technology, privacy, and the loss of reserve". Technology in Society Vol. 31 (4): 406-413. 2009

[10] M. Hildebrandt. "Technology and the end of law. In Facing the limits of the law", En Claes, W. Devroe and B. Keirsbilck editores. Springer, 2009.

[11] Samy Kamkar, Descripcion del mecanismo "Evercookie", Disponible http://samy.pl/evercookie/

[12] V. Mayer-Schönberger, "Delete: The virtue of forgetting in the digital age.", Princeton University Press, 2009.

[13] Mozilla Do Not Track Project, Disponible en: http://dnt.mozilla.org/

[14] A. Narayanan, V. Shmatikov. "Robust De-anonymization of Large Sparse Datasets" (How to Break Anonymity of the Netflix Prize Dataset). Security &Privacy, Oakland. Disponible en http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf, 2008

[15] H. Nissenbaum, "Protecting Privacy in an Information Age: The Problem of Privacy in Public". Law and Philosophy vol. 17(5-6), pp.559, 596. 1998.

[16] H. Nissenbaum, "Privacy as Contextual Integrity". Washington Law Review. Vol. 79 (119): pp.119-159. 2004.

[17] C.Paine, U-D. Reips, S. Stieger, A.N. Joinson, and T.Buchanan, "Internet users' perceptions of 'privacy concerns' and 'privacy actions'". International Journal of Human-Computer Studies. Vol 65(6), pp. 526-536. 2007.

[18] "D2.4.1 - Final report on mechanisms", Primelife.eu report, Disponible en: http://www.primelife.eu/results/documents/144-241d

[19] SecurityFocus, "AOL search data identified individuals", http://www.securityfocus.com/brief/277, 2006

[20] D.J. Solove, "The digital person: Technology and privacy in the information age". New York. New York University Press.

[21] D.J. Solove, ""I've got nothing to hide" and other misunderstandings of privacy". San Diego Law Review. Disponible en "http://ssrn.com/abstract=998565", vol. 44, pp. 745-772.

[22] Peter Steiner, "On the Internet, nobody knows you're a dog", New Yorker Magazine, publicada 5/Jul/1993, Disponible en http://www.cartoonbank.com/invt/106197, 1993.

[23] L. Sweeney, "Simple Demographics Often Identify People Uniquely". Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh. Disponible desde http://dataprivacylab.org/projects/identifiability/index.html, 2000.

[24] B.van der Berg, R. Leenes (editores), "Privacy Enabled Communities", Reporte de Privacy and Identity Management in Europe for Life, Disponible en: http://www.primelife.eu/images/stories/deliverables/d1.2.1-10.04.23-privacy_enabled_communities-public.pdf, 2011.

[25] Atul Varma, "Collusion Plugin", v.011, Disponible https://secure.toolness.com/xpi/collusion.html

[26] A. Vedder, "Privacy, een conceptuele articulatie", Filosofie & Praktijk, Vol 30(5), pp. 7-19, 2009. Citado en [17]

[27] S. Warren, L. Brandeis, "The Right to Privacy", Harvard Law Review, vol. 4(5), 1890.

[28] Washington Post, "Sen. Rockefeller introduces 'do not track' bill for Internet", Disponible en: http://www.washingtonpost.com/blogs/post-tech/post/sen-rockefeller-introduces-do-not-track-bill-for-internet/2011/05/09/AF0ymjaG_blog.html

[29] A.F. Westin, "Privacy and Freedom", 1st edition, New York, Atheneum, 1967.

[30] M. Zimmer, "On the 'Anonymity' of the Facebook Dataset", Disponible en: http://michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook-dataset/, 2008.