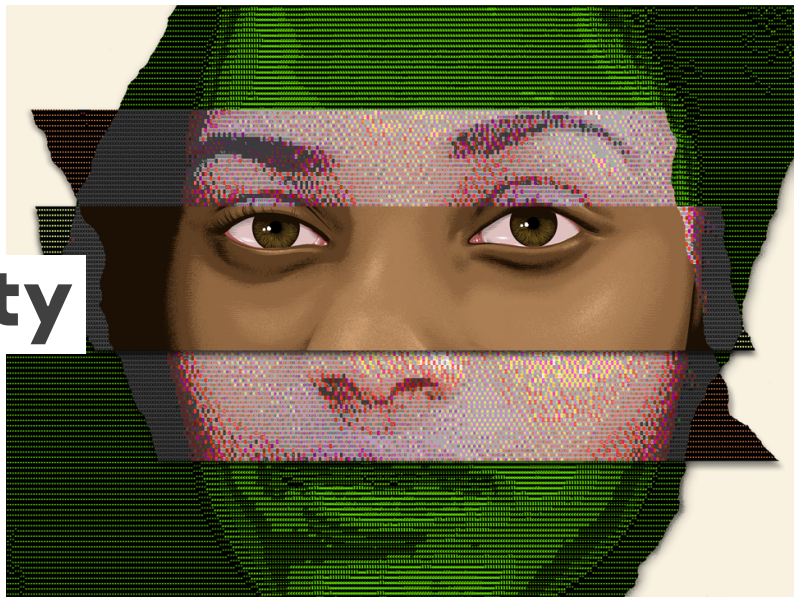
Los riesgos del Big Data: ¿Una nueva commodity o un problema para el futuro?



Todo y cualquier dato colectado tiene una gran posibilidad de ser robado. O por lo menos todos los datos que tu empresa retiene a lo largo del tiempo. Es por eso que el especialista Cory Doctorow nos desafía a repensar el uso de los datos de los consumidores.

AUTOR
Cory Doctorow

> Todo y cualquier dato colectado tiene una gran posibilidad de ser robado. O por lo menos todos los datos que tu empresa retiene a lo largo del tiempo.

Esa es una afirmación que fue controversial en el pasado, pero que hoy se ha vuelto sentido común. Si el Equifax, la CIA, la NSA, el Departamento de Gestión de Personas (la entidad que administra las jubilaciones en los EEUU), Facebook, las aplicaciones de citas y muchos otros no logran mantener nuestros secretos, tu negocio tampoco va a poder hacerlo.

En verdad, toda la confianza ciega que la industria le pone a la seguridad de datos siempre ha sido un ejemplo de un raciocinio enfocado solo en los negocios. Recolectar datos y almacenarlos se ha vuelto algo tan barato y fácil que muchos analistas, inversores y hasta especuladores salieron por ahí afirmando que "los datos son el nuevo petróleo". Con todo eso, dejar de recoger y guardar para siempre todo lo que puedas obtener se volvió algo irresponsable y casi ignorante desde un punto de vista fiscal.

¿Quién diría que la información se volvería algo tan rentable? Era dinero cayendo del cielo y las empresas vieron eso como una oportunidad para llenar los bolsillos. Aún que no supieran exactamente qué hacer con esos datos, no restaban dudas de que un mercado para ellos surgiría en un horizonte cercano.

Hoy, claramente, el tiempo nos mostró que esas personas estaban equivocadas, y vemos filtraciones todo el tiempo, en olas cada vez más grandes de ataques virtuales. Con eso, toda la lógica anterior ha cambiado. Ahora, en lugar de discutir si las filtraciones son o no inevitables, la salida es decir que el robo de datos no es algo tan relevante. Con eso, a cada vez que un robo de datos ocurre, vemos al portavoz de la compañía afectada recitando la misma frase de siempre, casi que religiosamente: "Llevamos la privacidad de nuestros consumidores muy en serio, pero garantizamos que ninguno de los datos filtrados es comprometedor".

Una parte de esa reacción viene de un cierto "nihilismo sobre la privacidad": toda va a salir al público en un momento u otro, entonces qué diferencia hace? Pero existe una versión aún más engañosa de ese discurso que es la defensa de que datos filtrados no son un gran problema porque los criminales no tienen mucho qué hacer con aquello. O sea, más que nihilistas, las compañías también son negacionistas.

Los que piden perdón por las filtraciones argumentan aún que los datos no son comprometedores o peligrosos porque son anónimos, o porque tuvieron sus identificadores removidos. Pero eso en realidad es una desinformación absurda y profunda sobre cómo los datos son usados.

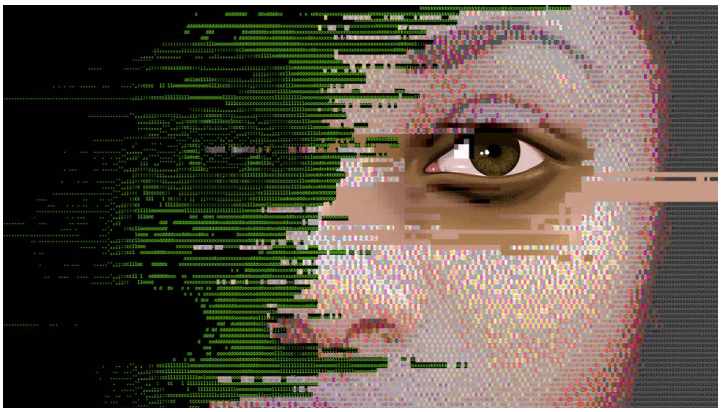
La reidentificación de paquetes de datos es uno de los tópicos en alta en la ciencia de la computación hoy en día, con especialistas creando herramientas automáticas que pueden juntar piezas de diferentes paquetes de datos para identificar a quién le pertenece esa información. Por ejemplo, se pueden fundir datos anónimos y perecibles de una autoridad de salud, como un médico consultado, medicamentos indicados, fecha y horario, con otro paquete de datos robado de una empresa de taxis que incluye los viajes hechos a un hospital en particular. Con eso, se pueden descubrir cosas como quienes están tomando medicamentos para la depresión, antirretrovirales o haciendo un tratamiento para el cáncer.

Muchos proveedores de protección de datos prometen que sus sistemas irán inyectar ruidos a los paquetes de archivos para evitar que ocurra la reidentificación de datos, pero esas medidas casi nunca superan las pruebas hechas por especialistas.

Ya hace años que el primer trabajo significativo sobre la reidentificación teórica fue publicado y las cosas siguen empeorando para aquellos que insisten que mantener datos anónimos aún es posible.

Pero los métodos de reidentificación nos dicen mucho sobre cómo los criminales digitales operan y sobre su increíble creatividad y autocontrol.

> Como nossos ancestrais da década de 1930, que foram perseguidos pela miséria da era pós-Depressão, os ladrões de identidade nunca jogam nada fora e sempre encontram maneiras de usar cada pedacinho solto para criar algo novo.



Nombre de usuario y contraseña pueden ser reciclados y utilizados para invadir cámaras de seguridad de plataformas modernas como Ring y Nest, pedir comida por aplicaciones o rastrear y movilizar flotas enteras de vehículos corporativos. Datos de usuario filtrados pueden ser usados para sobrecargar procedimientos reglamentarios con comentarios falsos, pero plausibles, y hasta para crear decenas de cuentas falsas en Twitter.

Criminales operan combinando y recombinando paquetes de datos, usando la filtración de una empresa mezclada con una fuente de datos públicos y además datos anónimos de una tercera empresa para causar daños en proporciones gigantes. Ellos pueden incluso lograr juntar fragmentos suficientes para obtener una copia de la escritura de una casa y así venderla a otra persona mientras te vas de vacaciones.

No es posible apuntar para un dato específico y decir: "este es el dato que va a hacer que pierdas tu casa" o "este es el dato que va a permitir que ladrones para que limpien tu cuenta de ahorros". Pero eso no importa. Tampoco es factible llegar a una fábrica, ver un humo raro saliendo de una de sus chimeneas y decir: "¡Ese es! Eso es lo que va a hacer que aquella mujer, la madre de tres hijos que vive a 10 kilómetros de la fábrica, tenga cáncer". Pero eso no impide que las compañías que contaminan el aire o el agua de una región sean responsabilizadas por sus irregularidades. Daños causados por filtraciones son impredecibles y difíciles de determinar.

No hay manera de estar seguros sobre qué dato puede causar cada problema.

> Pero sabemos que esos daños son inevitables y que aumentan de acuerdo con el tamaño de la filtración.

Hasta el momento, las compensaciones para los afectados por filtraciones de datos han sido extremadamente limitadas, pero eso mejora poco a poco. La filtración ocurrida en Home Depot en el 2014 tuvo un costo final de apenas US $0,34 en indemnizaciones por cada consumidor afectado. Pero eso pasó hace seis años. Clientes de Yahoo! que tuvieron sus datos filtrados hace poco deben recibir valores cercanos a los US $100 cada uno. Facebook es otra compañía que acaba de recibir una multa de increíbles US $5 mil millones. Y la fiesta apenas comienza.

Los daños causados por filtraciones son acumulativos. Como la basura tóxica liberada en la naturaleza, las filtraciones generan un acúmulo de informaciones sueltas y se vuelven prácticamente inmortales en capacidad de generar problemas. Mientras el público, y la Ley, empiezan a notar esos efectos, seguimos en el camino para ver indemnizaciones cada vez más altas para aquellos que ven sus datos privados robados y sueltos de forma definitiva para el mundo.

Es una pena que cuando eso pase ya sea demasiado tarde. Los datos que tu empresa almacena hoy en días tienen una gran probabilidad de haber sido filtrados y robados de tu red sin que lo notaras, hasta que uno de sus consumidores descubra de la peor manera posible que tu negocio comprometió su privacidad y con eso decida demandar a tu empresa.

Tu aseguradora tampoco va a ayudarte, ni va a crear nuevas pólizas de seguro para tu empresa o protecciones contra errores y omisiones para los miembros de tu consejo, porque tú estás guardando un material frágil y pasible de filtraciones. Y esa ayuda se volverá aún menos probable cuando las penalidades por perder el control sobre esos datos empiecen a transformarse en prejuicios financieros.

Tal vez aún puedas justificar ese gran riesgo si las ganancias obtenidas con esos datos alcanzaran la misma magnitud. Pero como han descubierto los especialistas, los beneficios de la obtención y del almacenamiento de datos suelen ser muy sobrevalorados, principalmente porque se descubrió que la eficiencia de los anuncios basados en el comportamiento de los usuarios es casi idéntica a la de aquellos basados solo en el contenido de los sitios donde aparecen.

Pero si eres de una agencia de publicidad data-driven o de una de las gigantes de la tecnología, como Facebook o Google, todas la mística que existe sobre la capacidad de transformar datos en conversión permite que vendas tu producto como un servicio Premium. Mientras eso, intimidas a posibles competidores que ni siquiera se atreven a entrar en el mercado por miedo de no lograr cosechar la misma cantidad de datos de aquellos que ya dominan el sector.

Al final, los que afirman que los datos son “el nuevo petróleo” son los mismos que los venden. Y las afirmaciones hechas por ellos de que datos como esos permitirán que hagas cosas increíbles, aún son apenas un truco de vendedor, y no algo comprobado.

Los datos nunca fueron el nuevo petróleo y si los nuevos residuos tóxicos: potentes e inmortales, pero imposibles de contener. Si fuera usted, no trataría de extraer más de ellos, pero si de librarme de esa cantidad enorme de información acumulada sin criterio, ni propósito, y que se puede convertir en un problema.

Minimizar tu acumulo de datos nos es solo una buena práctica, pero también una buena decisión de negocios. Recolecta solo lo que realmente necesites y guarda esos datos por el menor tiempo posible. si tu política de privacidad cabe en una servilleta, entonces eso quiere decir que estás acumulando y procesando una cantidad mínima pero específica de datos y borrándolos luego en seguida. Y eso quiere decir que estás en el camino correcto.

Este es un artículo de opinión y no refleja necesariamente las opciones, ni las posiciones de Kaspersky.

kaspersky