Big Data: uma nova commodity ou apenas um problema em potencial?

Todo e qualquer dado coletado tem grandes chances de ser vazado. Pelo menos todos os dados que você reter, com o passar do tempo. É por isso que o especialista Cory Doctorow nos desafia a repensar o uso dos dados de consumidores.

AUTOR
Cory Doctorow

> Todo e qualquer dado coletado tem grandes chances de ser vazado. Pelo menos todos os dados que você reter, com o passar do tempo.

Essa é uma afirmação que já chegou a ser controversa tempos atrás, mas que hoje se tornou senso comum. Se o Equifax, a CIA, a NSA, o Departamento de Gestão de Pessoas (uma espécie de INSS estadunidense), o Facebook, os sites de paquera e vários outros não conseguem manter nossos segredos, seu negócio também não vai conseguir.

Na verdade, toda a confiança cega que a indústria coloca sobre a segurança de dados sempre foi um exemplo de raciocínio voltado para o negócio. Coletar dados e armazená-los se tornou algo tão barato e fácil, que inúmeros analistas, investidores e até aproveitadores saíram por aí afirmando que "os dados são o novo petróleo". Com tudo isso, deixar de coletar e guardar para sempre tudo que você puder obter se tornou algo, de um ponto de vista fiscal, irresponsável e ignorante.

Quem diria que a informação se tornaria algo tão rentável? Era dinheiro caindo do céu e as empresas viram isso como uma oportunidade para encher os bolsos. Ainda que elas não soubessem exatamente o que fazer com esses dados, não havia dúvidas de que um mercado para eles surgiria num horizonte próximo.

Com um panorama tão otimista como esse, não foi surpreendente que as pessoas que se dedicaram a coletar e reter esses dados se convencessem de que essas atividades poderiam ser feitas de maneira segura.

Hoje, é claro, o tempo provou que elas estavam erradas, e vemos vazamento atrás de vazamento em ondas cada vez maiores de ataques virtuais. Com isso, a lógica mudou. Agora, ao invés de argumentar que os vazamentos são inevitáveis, a saída é dizer que esses vazamentos não são algo tão relevante assim. Com isso, cada vez que um roubo de dados acontece, vemos o porta-voz da empresa afetada recitando a mesma frase, quase que religiosamente: "Levamos a privacidade de nossos consumidores muito a sério, mas garantimos que nenhum dos dados vazados é comprometedor".

Um pouco dessa reação vem de um certo "niilismo sobre a privacidade": tudo vai vazar uma hora ou outra, então que diferença faz? Mas existe uma versão ainda mais ardilosa desse discurso, que é a defesa de que dados vazados não são um problema porque os criminosos não têm muito o que fazer com aquilo. Mais do que niilistas, as empresas também costumam ser negacionistas.

Os que pedem desculpas pelos vazamentos ainda argumentam que os dados vazados não são comprometedores ou perigosos porque são anônimos, ou porque tiveram os identificadores removidos. Mas isso na verdade é uma desinformação profunda sobre como os dados são usados. E abusados.

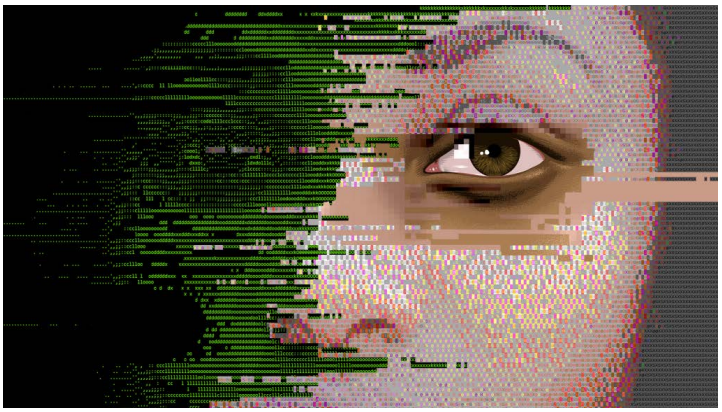
A re-identificação de pacotes de dados é um dos tópicos em alta na ciência da computação hoje em dia, com pesquisadores criando ferramentas automáticas que conseguem juntar peças de diferentes pacotes de dados para identificar a quem eles pertencem. Por exemplo, você pode fundir dados anônimos e perecíveis de uma autoridade de saúde, como médico consultado, remédios receitados, data e horário, com um pacote de dados roubados de uma empresa de táxis que inclua viagens a um hospital em particular. Com isso, você consegue descobrir coisas como quem está tomando remédios para depressão, antirretrovirais ou fazendo tratamento para o câncer.

Muitos provedores de proteção de dados prometem que seus sistemas irão injetar ruídos nos pacotes de arquivo para evitar a re-identificação de dados, mas essas promessas raramente passam no teste dos pesquisadores.

Já faz anos que o primeiro trabalho significativo de re-identificação teórica foi feito e as coisas continuam piorando para aqueles que insistem que manter os dados anônimos ainda é possível.

Mas os métodos de re-identificação nos dizem muito sobre como os criminosos digitais operam e sobre sua incrível desenvoltura e autocontrole.

> 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.



Nomes de usuário e senhas podem ser reciclados e usados para invadir câmeras de segurança de plataformas como Ring e Nest, pedir comida em aplicativos ou rastrear e mobilizar frotas inteiras de veículos corporativos. Dados de usuário vazados podem ser utilizados para sobrecarregar procedimentos regulatórios com comentários falsos, mas plausíveis, e até para criar dezenas de contas de Twitter falsas.

Criminosos operam combinando e recombinando pacotes de dados, usando o vazamento de uma empresa em combinação com uma fonte de dados públicos e, ainda, dados anônimos de uma terceira empresa para causar estragos em proporções gigantes. Eles podem até conseguir fragmentos suficientes para obter uma escritura duplicada da sua casa e assim vendê-la para outra pessoa enquanto você estiver de férias.

Não é possível apontar para um dado específico e dizer: "esse é o dado que vai te fazer perder sua casa" ou "esse é o dado que vai permitir dar acesso aos ladrões para sua limpar sua aposentadoria".

Também não é factível chegar numa fábrica, ver uma fumaça estranha saindo de uma chaminé e dizer: "É essa aí! Isso é o que vai causar câncer naquela mulher, mãe de três filhos, que mora a 10 quilômetros daqui". Mas isso não impede que empresas que poluam o ar ou a água sejam responsabilizadas por suas irregularidades.

Danos causados por vazamentos são imprevisíveis e difíceis de determinar.

> Não temos como ter certeza sobre quais dados podem causar cada problema, mas sabemos que esses danos são inevitáveis e que aumentam conforme o tamanho do vazamento.

Até o momento, as compensações para quem é afetado por vazamentos de dados têm sido extremamente limitadas, mas isso está melhorando. O vazamento ocorrido na Home Depot, em 2014, teve um custo de apenas US$ 0,34 em indenizações por cada consumidor afetado. Mas isso foi seis anos atrás. Clientes do Yahoo! que tiveram seus dados vazados recentemente devem receber uma compensação próxima dos US$ 100 cada. O Facebook é outro que acaba de receber uma multa de incríveis US$ 5 bilhões. E essa festa está só começando.

Os danos causados por vazamentos são cumulativos: como lixo tóxico na natureza, vazamentos geram um acúmulo de informações à deriva e se tornam praticamente imortais na capacidade de gerar problemas. Enquanto o público – e a Lei – começam a perceber esses efeitos, estamos no caminho de ver compensações cada vez maiores para aqueles que veem seus dados privados serem vazados de maneira definitiva para o mundo.

É importante lembrar que esse tipo de vazamentos afeta a todos: civis e militares, pobres e ricos, incluindo políticos, legisladores e membros do judiciário. Inevitavelmente, teremos precedentes suficientes para ver as compensações por vazamentos de dados se aproximarem cada vez mais às de outros problemas, como os crimes ambientais.

Pena que quando isso acontecer já será tarde demais. Os dados que sua empresa armazena hoje em dia muito provavelmente já foram retirados da sua rede sem você nem perceber, até que um de seus consumidores descubra do pior jeito possível que você comprometeu sua privacidade e decida ir atrás de seus direitos na justiça.

Sua seguradora também não vai te ajudar ou criar novas apólices para sua empresa ou proteções contra erros e omissões para seu conselho, porque você está armazenando um material frágil e passível de vazamentos. E essa ajuda se tornará ainda menos provável quando as penalidades por perder o controle sobre esses dados começarem a se transformar em perdas financeiras.

Talvez você ainda consiga justificar todo esse risco se os lucros obtidos com esses dados forem da mesma magnitude. Mas, como têm descoberto os pesquisadores, os benefícios da obtenção e armazenamento de dados costumam ser muito supervalorizados, até porque a eficácia dos anúncios segregados baseados no comportamento dos usuários é quase idêntica àqueles baseados no conteúdo das páginas onde os anúncios aparecem, que no caso não precisam usar dados dos usuários.

Mas se você é uma agência de publicidade data-driven ou uma das gigantes da tecnologia, como Facebook ou Google, toda essa mística que existe sobre a capacidade de transformar dados em conversões permite que você venda seu produto como um serviço premium, enquanto intimida possíveis competidores que nem se atrevem a entrar no mercado por medo de não conseguir colher a mesma quantidade de dados de quem já domina o setor.

Aqueles que afirmam que os dados são o "novo petróleo" são os mesmos que os vendem. E as afirmações feitas por eles – de que dados como esses permitirão que você faça coisas incríveis – ainda são apenas um truque de vendedor, e não algo comprovado.

Os dados nunca foram o novo petróleo e sim os novos dejetos tóxicos: potentes e imortais, mas impossíveis de conter. Se fosse você, eu não tentaria extrair mais deles e sim me livrar dessa quantidade enorme de informação acumulada sem critério, que pode se tornar um problema.

Minimizar seu acúmulo de dados não é somente uma boa prática, mas também uma boa decisão de negócios. Colete somente aquilo que você realmente precisa e guarde esses dados pelo menor tempo possível. Se sua política de privacidade couber num guardanapo, é porque você está coletando e processando uma quantidade mínima e específica de dados, e apagando-os logo em seguida. Isso quer dizer que você está no caminho certo.

Este artigo reflete somente as opiniões de seu autor e não necessariamente as opiniões e políticas da Kaspersky.

kaspersky