Vazamentos de Dados: Histórico, Impacto Socioeconômico e as Novas Leis de Proteção de Dados

Rodrigo Bisso[1,3], **Diego Kreutz**[1,2,3], **Gustavo Rodrigues**[1,3], **Giulliano Paz**[1,2,3]

[1]Laboratório de Estudos Avançados (LEA)
[2] Mestrado Profissional em Engenharia de Software (MPES)
[3] Universidade Federal do Pampa (UNIPAMPA)
{bisso,giulliano94,crod}@gmail.com,kreutz@unipampa.edu.br

Resumo. *Notícias e relatórios de segurança sobre vazamentos de dados sensíveis têm surgido com uma frequência cada vez maior. Muitos desses vazamentos, cujo volume e criticidade são altos, vêm afetando empresas e governos de forma significativa. Este cenário, cuja perspectiva atual é piorar, tem chamado tanta atenção dos governos e empresas que segurança da informação virou uma das maiores prioridades de estado em muitos países. Este trabalho tem por objetivo apresentar um histórico e o impacto socioeconômico de alguns dos vazamentos de dados mais significativos dos últimos anos. Adicionalmente, identificar e discutir as recentes leis de proteção de dados, criadas com o intuito de ajudar a combater o problema de falta de formação e investimentos em segurança da informação, a principal causa desses vazamentos de dados.*

1. Por que Leis Rígidas de Proteção de Dados?

Notícias e relatórios sobre vazamentos de dados têm surgido com uma frequência cada vez maior. As estatísticas mostram um volume enorme de dados vazados nos últimos anos. Além do número e da frequência dos vazamentos, outro aspecto estarrecedor é o impacto dos dados vazados, como o re-projeto de dezenas (e até centenas) de sistemas, como ocorreu recentemente nos EUA devido ao caso de vazamento da Equifax [Ng 2019].

Os incidentes de segurança têm afetado diferentes áreas e setores da sociedade. Recentemente, dados de mais de 60 universidades e colégios dos EUA foram comprometidos devido a um conjunto de vulnerabilidades existentes num único sistema [Muncaster 2019]. No setor alimentício, o aplicativo EatStreet foi alvo de um vazamento que comprometeu dados de pagamento e cartões de crédito [Gatlan 2019a]. Segundo informações do site oficial do aplicativo, há parcerias firmadas com mais de 15 mil restaurantes em 1100 cidades.

No setor bancário, em 2018, o Banco Inter vazou dados de 20 mil clientes segundo investigação do MP [Higa 2018]. Quase um ano depois, outra falha de sistema deixa dados de mais de 1,4 milhões de clientes expostos para acesso na Internet [Payao 2019]. Considerando que instituições bancárias, por via de regra, se preocupam mais e investem muito mais em segurança da informação, este é um cenário bastante crítico.

Na verdade, o cenário é tão crítico e preocupante que governos começaram a criar leis a fim de definir os direitos de privacidade dos dados dos usuários e penalidades explícitas para os casos onde as regulamentações não forem cumpridas. A exemplo disso, a União Européia e o Brasil propuseram a *General Data Protection Regulation* 2016/679

(GDPR)[1] e a Lei Geral de Proteção de Dados Pessoais (LGPD)[2], respectivamente. Estas leis aplicam multas severas às empresas que fazem mal uso dos dados de seus usuários.

O principal objetivo deste trabalho é apresentar dados e desafios do cenário atual. As principais contribuições do trabalho podem ser resumidas em: (a) um levantamento histórico de alguns dos principais vazamentos de dados de 2014 a 2019; (b) uma análise do impacto sócioeconômico dos vazamentos de dados; (c) uma síntese das principais leis de proteção de dados; (d) uma discussão sobre avanços tecnológicos recentes e desafios de pesquisa e desenvolvimento; e (e) informar e conscientizar empresas e profissionais da área de tecnologia sobre a extrema importância do assunto. Devido a limitação de espaço, uma versão estendida do paper está disponível online [Machado et al. 2019].

2. Vazamentos de Dados

No primeiro semestre de 2019, ocorreram vários vazamentos de grandes volumes de dados sensíveis [Turner 2019]. Um exemplo é o vazamento de dados de 2,4 milhões de usuários de uma empresa de gerenciamento de senhas, incluindo nomes de usuários, emails, dicas de senhas, endereços de IP e senhas cifradas. Outros dois vazamentos marcantes de 2019 são os 540 milhões de registros de usuários do Facebook e as 773 milhões de senhas e dados de usuários da *Collection #1* [Hern 2019, Turner 2019].

Nos anos anteriores não foi diferente. A Tabela 1 apresenta um resumo dos 5 maiores vazamentos de dados ocorridos em cada ano, de 2014 a julho de 2019. Em 2018, o maior vazamento de dados comprometeu mais de 1 bilhão e 100 milhões de registros de dados pessoais de cidadãos da Índia [Leskin 2018]. Dadas as proporções, algo similar pode ser observado em 2017, 2015 e 2014. A única excessão foi 2016, cujo maior vazamento registrado foi de 5 milhões de registros de usuários. É interessante observar também que apenas em 2014 e 2016 houveram incidentes com menos de 1 milhão de registros entre os 5 maiores vazamentos de dados.

	2019	**2018**	**2017**	**2016**	**2015**	**2014**
#1	773M	1,1B	145,5M	5M	78,8M	145M
#2	200M	500M	5,5M	2,2M	25M	2,6M
#3	24M	340M	2,2M	1,5M	15M	1,3M
#4	12M	150M	1,8M	950m	11M	774m
#5	7,7M	100M	1,6M	320m	10M	550m

Tabela 1. Vazamentos de dados sensíveis (B = bilhão, M = milhão, m = mil)

Outro aspecto a ressaltar é o fato de os vazamentos de dados atingirem os mais diversos ramos e setores da sociedade. Recentemente, dois grandes vazamentos de dados, ambos envolvendo empresas de assistência médica, chamaram a atenção. No primeiro caso, a empresa Quest Diagnostics teve dados de 12 milhões de clientes vazados [McKay 2019]. Em outro caso similar, a empresa LabCorp teve dados de 7,7 milhões de clientes vazados [Lam 2019]. Os dados vazados incluem nomes, endereços, data de nascimento, informações de pagamento e dados de seguro social. Além de empresas laboratoriais, hospitais também têm sido vítimas de *hackers* [Riley 2019]. Em um caso

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[2] http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm

recente, dados do setor de pesquisa do *Massachusetts General Hospital* foram roubados, envolvendo aproximadamente 10 mil pacientes.

No ramo alimentício, um caso que chamou bastante atenção foi o da franquia Ceckers and Rally's, no qual 103 pontos de venda foram vítimas de um *malware* que roubava dados de cartões de crédito dos clientes. O detalhe estarrecedor: o *malware* estava ativo há três anos sem ter sido descoberto [Khandelwal 2019]. O número de pontos de venda, verificados e comprovadamente afetados até o momento, representa 15% do total de lojas.

Na área governamental, o caso recente da Bulgária ganhou as manchetes. Segundo relatórios, um hacker roubou dados de mais de 5 milhões de cidadãos, sendo que o país possui uma população aproximada 7 milhões de pessoas [Cimpanu 2019a]. Os dados (i.e. nomes, endereços, informações sobre renda) foram extraídos de aproximadamente 110 bases de dados governamentais, totalizando 21 GB. Este é mais um exemplo da dimensão e sofisticação dos ataques atuais.

No ramo financeiro, dois casos graves, envolvendo as empresas Mastercard e Capital One, merecem destaque. No caso da Mastercard, dezenas de milhares de clientes do programa de fidelidade *Priceless Specials* da Alemanha e da Bélgica tiveram seus dados distribuidos na Internet [Gatlan 2019b]. Já no caso da Capital One, mais de 140 milhões de cidadãos americanos e 6 milhões de cidadãos canadenses tiveram seus dados vazados.

O interessante é observar que nem as empresas de alta tecnologia, que supostamente possuem times especializados e preparados para proteger suas infra-estruturas, escapam das manchetes de incidentes de segurança. Recentemente, foram descobertos dados (i.e. nomes, senhas, comentários e outras informações) de 540 milhões de usuários do Facebook armazenados em servidores da Amazon [Cimpanu 2019b]. Os dados, descobertos por pesquisadores especializados em vazamentos de dados, estavam em um servidor que pertencia à empresa mexicana Cultura Colectiva.

3. Impacto Socioeconômico

Em 2018, o custo envolvendo vazamentos de dados, apenas nos EUA, somaram 654 bilhões de dólares e expuseram 2,4 bilhões de dados de usuários [Security 2018]. Segundo o relatório, os tipos de dados mais vazados são data de nascimento e número de seguro social (21,6%) e nome e endereço (20%). Com relação aos tipos de ataques, os que aparecem em maiores percentuais são de acesso não autorizado (34,2%) e *malware* (17,3%).

A Tabela 2 apresenta alguns casos onde foram aplicadas (ou estão em processo de aplicação) multas devido a vazamentos de dados. No caso dos EUA, como ainda não existe uma regulamentação federal específica, fica a cargo da *Federal Trade Commission* (FTC), responsável pelos direitos do consumidor, a definição e aplicação das multas.

O Facebook está sendo multado em US$5 bilhões pelo governo americano. O principal motivo é o caso envolvendo a empresa Cambridge Analytica, no qual haviam especulações sobre a relação de vendas e o uso indevido de dados na campanha eleitoral de Donald Trump, em 2016. O caso veio a público pela primeira vez em dezembro de 2015 e chegou a corte americana no início de 2018.

A companhia hoteleira Marriott está sendo multada em £100 milhões de libras

Valor	Empresa	País	Ano
US$ 5 B	Facebook	EUA	2019
£ 183,39 M	British Airways	Reino Unido	2019
US$ 148 M	Uber	EUA	2018
US$ 85 M	Yahoo	EUA / Israel	2018
€ 50 M	Google	França	2019
US$ 22,5 M	Google	EUA	2012
US$ 10 M	Blue Cross Blue Shield	EUA	2019
US$ 3,8 M	AMCA	EUA	2019
R$ 1.5 M	Banco Inter	Brasil	2018
€ 600 m	Uber	Holanda	2018
£ 385 m	Uber	Reino Unido	2018

Tabela 2. Penalidades aplicadas a empresas (B = bilhão, M = milhão, m = mil)

pelo vazamento de dados de 339 milhões de clientes. Entre os dados vazados estão números de cartão de crédito e dados de passaporte. A falha é decorrente de um sistema adotado pela empresa após a compra de outra rede hoteleira, a Starwood, que já havia sido notificada de problemas de segurança em seus sistemas em 2014.

Os vazamentos envolvendo as empresas Quest Diagnostics e LabCorp, que utilizavam a empresa AMCA para realizar os pagamentos, afetaram mais de 20 milhões de clientes. O impacto foi tamanho que AMCA pediu proteção contra falência depois de ter sido condenada a pagar mais de 3,8 milhões de dólares em multas.

Até pouco tempo, pouco se dava atenção à privacidade e a segurança de dados. Entretanto, os dados e exemplos de penalidades mostram que o cenário mudou drasticamente nos últimos anos, isto é, estes assuntos tornaram-se uma prioridade de estado. Diferentes países já possuem leis estabelecidas de proteção de dados. Os que ainda não possuem, como os EUA, estão aplicando penalidades caso-a-caso e caminhando para criar as suas leis de proteção de dados. Empresas como a Google e a Uber já foram multadas em diferentes países. Esta é uma tendência que veio para ficar e vai atingir todas as empresas, de todos os portes. Entretanto, a maioria absoluta das empresas ainda não está preparada para este novo cenário. É preciso informar e conscientizar empresas e profissionais da área sobre a importância e necessidade de atenção do assunto.

4. Leis de Proteção de Dados

Com o objetivo de mudar este cenário, indiscutivelmente crítico, governos têm tomado medidas para que as empresas aumentem os investimentos e a preocupação com a segurança dos dados dos usuários. A União Europeia (EU) criou, em 2016, uma nova regulamentação para a proteção de dados pessoais, a *General Data Protection Regulation* 2016/679 (GDPR). A GDPR é um marco legal para a proteção e privacidade de dados de todos os cidadãos da EU e do Espaço Econômico Europeu (EEE), tornando a proteção de dados pessoais um direito fundamental, assim como a liberdade. Inspirada na GDPR, em 2018, foi sancionada a Lei Geral de Proteção de Dados Pessoais (LGPD), nº 13.709, a qual entrará em vigor em agosto de 2020. Tanto a GDPR quanto a LGPD visam proteger e fortalecer a privacidade, dando um maior controle aos cidadãos sobre seus dados pessoais e determinando como devem acontecer a coleta e o tratamento desses dados por terceiros.

Dados pessoais, segundo ambas as leis, são informações que possam identificar,

direta ou indiretamente, uma pessoa natural, como CPF, RG e nome completo. Além disso, dados não pessoais como profissão, localização e endereço IP podem se tornar dados pessoais, se utilizados em conjunto com outros dados, para identificar uma pessoa natural. Dados pessoais sensíveis são informações que podem violar a intimidade, honra e imagem das pessoas naturais, como origem racial e étnica, convicções religiosas, políticas e filosóficas, dados genéticos e biométricos e dados referentes à saúde e vida sexual.

Ambas as leis determinam que deverão responder às regulamentações toda e qualquer empresa, pública ou privada, ou pessoa, física ou jurídica, que: armazene ou trate dados pessoais em seu território; a coleta e tratamento de dados tenha como objetivo oferecer ou fornecer serviços em seu território; e colete e manipule dados de seus cidadãos, independente da nacionalidade ou localização da empresas e dados. As multas por violação das regulamentações podem chegar a €20 milhões ou 4% do faturamento anual da pessoa jurídica envolvida, no caso da GDPR, e 2% do faturamento anual da pessoa jurídica ou R$50 milhões, pela LGPD.

Tanto a GDPR quanto a LGPD estipulam que a coleta e o tratamento de dados pessoais, sensíveis ou não, se darão apenas mediante autorização explícita do titular dos dados, ou seja, a quem os dados referem-se. Os termos de uso deverão ser sucintos e explícitos, informando com qual finalidade, por quanto tempo e quais empresas e serviços terão acesso aos dados. A autorização de utilização de dados poderá ser cancelada facilmente e a qualquer momento, assim como a modificação e deleção dos dados pessoais. Em suma, os dados pessoais pertencem única e exclusivamente aos seus titulares, cabendo a estes a decisão de utilização, deleção e comercialização.

5. Discussão

No caso do grupo Disjardins, discutido anteriormente, o vazamento ocorreu por conta de um funcionário mal intencionado. Este é, de fato, um problema bastante preocupante. Relatórios recentes apontam que mais de 90% das empresas tem medo de usuários internos maliciosos [Cybersecurity Insiders 2018]. Não é para menos, pois relatórios de segurança mostram que aproximadamente 50% dos incidentes de segurança são causados por funcionários ou ex-funcionários das empresas. Entretanto, ainda há poucas alternativas tecnológicas para evitar incidentes internos.

Algumas das alternativas existentes, cujo principal objetivo é reduzir a possibilidade ou o impacto de vazamentos de dados, são os bancos de dados cifrados e tecnologias como Intel SGX. Bancos de dados cifrados, como o CryptDB [Popa et al. 2011], foram criados para impedir que administradores do sistema e do banco de dados tenham acesso aos dados em texto plano. Entretanto, a utilização de bancos de dados como o CryptDB [Popa et al. 2011], que utilizam criptografia homomórfica, ainda é tecnicamente inviável devido ao alto custo computacional envolvido.

Tecnologias como Intel SGX surgiram para parcialmente resolver o problema de desempenho imposto por soluções como a CryptDB. Intel SGX permite criar uma região de memória isolada, onde nem o sistema operacional tem acesso aos dados. A partir da SGX, recentemente, começaram a surgir soluções de armazenamento e processamento de dados seguro como a EnclaveDB [Priebe et al. 2018]. Estas soluções limitam a superfície de ataque de agentes maliciosos internos. Entretanto, apesar de representarem uma evolução significativa em termos de arquitetura e desempenho, mantendo a

segurança, ainda há desafios pela frente até tornarem-se soluções de produção, como as limitações em termos de operações de I/O e memória interna (i.e. 80MB para dados).

6. Conclusão

Resumidamente, este trabalho reúne informações (i.e., dados e consequências) sobre alguns dos maiores casos de vazamento de dados dos últimos anos. Os dados apresentados deixam clara a importância da necessidade de conscientização das instituições e empresas, públicas e privadas, com a responsabilidade sobre os dados de seus clientes. Como a sofisticação dos ataques não para de aumentar e a tecnologia está em constante evolução, investir em capacitação de pessoas, tecnologia e pesquisa é um caminho necessário, sem volta, a fim de evitar as penalidades de leis de proteção de dados como a LGPD, lei nº 13.709. Mais dados e discussões podem ser encontradas na versão estendida, disponível online [Machado et al. 2019].

Referências

Cimpanu, C. (2019a). Hacker steals data of millions of Bulgarians, emails it to local media. http://bit.do/e25Qd.

Cimpanu, C. (2019b). Over 540 million Facebook records found on exposed AWS servers. http://bit.do/e25QX.

Cybersecurity Insiders (2018). Insider threat - 2018 report. http://bit.do/e25Rf.

Gatlan, S. (2019a). Hacker Steals Customer Payment Info in EatStreet Data Breach. http://bit.do/e25PW.

Gatlan, S. (2019b). Mastercard reports data breach to german and belgian dpas. http://bit.do/e5Jwq.

Hern, A. (2019). Largest collection ever of breached data found. https://bit.ly/2Hf3E7V.

Higa, P. (2018). Banco Inter vazou dados de quase 20 mil clientes, diz investigação do MP. https://bit.ly/2O39mZd.

Khandelwal, S. (2019). Hackers Stole Customers' Credit Cards from 103 Checkers and Rally's Restaurants. http://bit.do/e25P6.

Lam, K. (2019). LabCorp says 7.7 million customers may have been affected by data breach. http://bit.do/e25Pd.

Leskin, P. (2018). The 21 scariest data breaches of 2018. http://bit.do/e25M2.

Machado, R. B., Kreutz, D., Paz, G., and Rodrigues, G. (2019). Vazamentos de Dados: Histórico, Impacto Socioeconômico e as Novas Leis de Proteção de Dados. In *4o WRSeg*. https://arxiv.kreutz.xyz/wrseg2019_data_leaks_ev1.pdf.

McKay, T. (2019). Lab Testing Giant Quest Diagnostics Says Data Breach May Have Hit Nearly 12 Million Patients. http://bit.do/e25Ps.

Muncaster, P. (2019). Over 60 US Colleges Compromised by ERP Exploit. https://bit.ly/2SC8zlm.

Ng, A. (2019). Thanks to Equifax breach, 4 US agencies don't properly verify your data. http://bit.do/e25LY.

Payao, F. (2019). Dados de 1,4 milhão de clientes do Banco Inter estavam expostos para acesso. https://bit.ly/2LKotJR.

Popa, R. A., Redfield, C., Zeldovich, N., and Balakrishnan, H. (2011). CryptDB: protecting confidentiality with encrypted query processing. In *31rst ACM SOSP*.

Priebe, C., Vaswani, K., and Costa, M. (2018). Enclavedb: A secure database using sgx. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 264–278. IEEE.

Riley, D. (2019). Massachusetts general hospital data breach latest failure to protect patient data. http://bit.do/e5KZP.

Security, H. N. (2018). 2018 in numbers: Data breaches cost $654 billion, expose 2.8 billion data records in the U.S. http://bit.do/e25NV.

Turner, S. (2019). 2019 Data Breachers - The Worst So Far. http://bit.do/e25MP.