Big Data e Guerra Cibernética: Os Impactos de Vazamento de Dados em Estruturas Governamentais

Big Data and Cyber War: The Impacts of Data Leakage on Government Structures

Luís Fernando Gomes Maquieira[1]
Istivson Leandro Sousa Ribeiro[2]
Georges Lucas Silva Queiroz de Albuquerque[3]

Na Era Digital, o conceito de *Big Data* revolucionou o modo como os dados são coletados, analisados e utilizados, tornando a informação, um recurso vital e estratégico, crucial para a soberania de um Estado. À medida que cresce a dependência de governos às tecnologias de *Big Data* para tomada de decisões que abrangem de politicas públicas a estratégias de defesa nacional, os ataques que visam o vazamento de dados e informação ganham o foco na Guerra Cibernética. A interconectividade e o armazenamento de grandes volumes de dados sensíveis em redes distribuídas tornam Nações e suas infraestruturas críticas, alvos primários para hackers e agentes mal-intencionados. Estes atores conduzem ataques bem coordenados e altamente sofisticados que buscam explorar vulnerabilidades em sistemas de segurança que não acompanham a velocidade e complexidade do ritmo de coleta de dados em constante expansão que o Big Data trouxe ao século vinte e um.

Neste artigo, pretende-se fazer uma rápida introdução ao conceito de Big Data e suas tecnologias e explorar os impactos sofridos por Governos ao lidar com o vazamento de dados sofridos por ataques cibernéticos citando casos recentes de ataque como por exemplo o vazamento de dados do *FBI* em 2015, *Australian Defense Force* em 2016, do Pentágono em 2018 e do Ministério da Saúde do Brasil em 2021. Além disso, o artigo se propõe a abordar as perspectivas futuras quanto a exploração das ferramentas de inteligência artificial e *machine learning*, que nasceram como solução para manipular grandes volumes de dados, podem auxiliar na construção de ataques cibernéticos voltados para o vazamento de informações.

No início do século vinte e um, a Internet das Coisas, a evolução dos meios de transmissão da internet, a democratização do acesso e a constate evolução tecnológica proporcionaram uma produção em massa de dados das mais variadas formas e dificuldades de processamento. Apesar de ter se tornado popular nos anos dois mil, em 1990 o pesquisador e cientista da computação John R. Marshey já estudava o assunto e cunhou o termo *Big Data* para se referir a grandes volumes de dados que crescem em variedade e velocidade. Na esteira da constante evolução tecnológica, muitas estruturas críticas de governo adotaram a digitalização para manter a escalabilidade e funcionalidade dos seus meios de produção e prestação de serviço. Sendo assim, Bancos, Hospitais, Industrias, Usinas de Energia, Escolas, Sistemas de Transporte e Sistemas de Telecomunicações adentram ao espaço cibernético de uma nação e, inevitavelmente, o torna um intenso campo de batalha na busca por dados sensíveis e comprometer estruturas de governo, ainda que não haja uma declaração oficial de guerra entre Nações, já que o espaço cibernético facilita o anonimato e acessibilidade independente de posição geográfica.

1 Aluno do Curso de Guerra Cibernética no Centro de Instrução de Guerra Eletrônica e Cibernética (fernando.maquieira@eb.mil.br)
2 Aluno do Curso de Guerra Cibernética no Centro de Instrução de Guerra Eletrônica e Cibernética (istivson.ribeiro@eb.mil.br)
3 Aluno do Curso de Guerra Cibernética no Centro de Instrução de Guerra Eletrônica e Cibernética (georges.albuquerque@eb.mil.br)

As tecnologias que surgem para lhe dar com o *Big Data* se desenvolvem e três grandes eixos, Armazenamento, Análise e Processamento e a Visualização. No campo de atuação do armazenamento, outros dois importantes conceitos surgem para fazer referência ao tipo de dados que se está armazenando *Data Lake* e *Data Warehouse*. O *Data Lake* faz referência a um sistema onde os dados são armazenados de forma centralizada, independente do seu tipo, podendo ser estruturados, semi estruturados ou não relacionais. Neste ambiente, ficam os dados que ainda não foram requisitados para análise, evitando com que se prolifere na organização vários sítios de armazenamento de dados que pode culminar com o esquecimento destes dados. Paralelamente, o *Data Warehouse* é um estrutura otimizada e altamente organizada de armazenamento onde os dados são processados para facilitar a sua análise, desenvolvimento de modelos de *Machine Learning, Deep Learning* e Inteligência Artificial. Atualmente no mercado, as tecnologias em uso para a construção de *Data Lake* e *Data WareHouse* são serviços de armazenamento em nuvem, sistemas de arquivos distribuídos e bancos de dados de linguagem não estruturada (*NoSQL*).

A etapa mais complexa de um sistema de *Big Data* é a Análise e Processamento dos dados. Neste ambiente, os dados armazenados são tratados para alimentar os modelos matemáticos que buscam determinar algum comportamento e construir modelos de inteligência artificial e *deep learning* para lidar com análises de dados ou automação de tarefas complexas. No contexto de organizações governamentais, estes dados são utilizados em construção de modelos de análise que auxiliam em tomadas de decisão de Estado como politica econômica e saúde e automação de infraestruturas críticas. Além disso, modelos de *deep learning* auxiliam países no progresso de pesquisa em diferentes áreas científicas. No mundo moderno é de vital importância um robusto sistema de *Big Data*, pois permite um acompanhamento aprofundado dos mais diversos setores de uma Nação.

Por fim, o eixo da visualização de dados trata da apresentação e disposição dos dados de forma que permita autoridades diversas a avaliar determinado cenário e tomar uma decisão embasada em dados matemáticos, diminuindo danos de uma decisão equivocada. A construção da informação propriamente dita se dá na fase de visualização, pois nela se busca em como construir conhecimento a partir do que foi produzido nos modelos anteriores e como será apresentado esse conhecimento ao público final.

Em uma guerra cibernética, o principal eixo foco de ações maliciosas é o armazenamento de dados. O comprometimento de uma robusta base de dados prejudica toda a cadeia do *Big Data*. Os ataques buscam comprometer três dos quatro pilares da segurança da informação diretamente relacionados ao Armazenamento de Dados, a disponibilidade, integridade e a confidencialidade. Quando um banco de dados robusto é comprometido de tal forma, sistemas inteiros podem colapsar, decisões erradas podem ser tomadas se a base de dados não for íntegra, sistemas baseados em automação podem parar de funcionar ou produzir o resultado errado, opinião pública pode ser facilmente moldado com divulgação de dados confidenciais que sejam interpretados de forma errada. Toda uma sociedade pode ruim de dentro para fora em ataques coordenados a estruturas de *Big Data* governamentais.

A história recente nos mostra que diversas situações onde Governos foram comprometidos com base em ataques cibernéticos. Em 2015 cerca de 20 mil agentes do FBI e 9 mil funcionários do Departamento de Segurança Interna dos Estados Unidos

tiveram seus dados pessoais expostos. Esta exposição colocou em risco a integridade dos funcionários, a segurança nacional comprometendo investigações e operações em andamento e a reputação das instituições de governo *Federal Bureau of Investigation* (FBI) e *Departament of Homeland Security* (DHS). O ataque se deu através técnicas de Engenharia Social, Extração e Divulgação de Dados através de um acesso não autorizado a um extenso banco de dados corporativo contendo informações classificadas. Operações de contraterrorismo, contrainteligência, segurança nacional e investigações criminais foram gravemente prejudicadas. No ano seguinte, em 2016, foram vazados dados de projetos militares da *Australian Defense Force* onde novamente, dados classificados relativos a projetos militares foram expostos na internet colocando em risco a soberania da Austrália. Dois anos mais tarde, em 2018, ocorreu um incidente de vazamento de dados envolvendo o Departamento de Defesa dos Estados Unidos, mais especificamente informações sobre programas militares secretos e sensíveis. Esse incidente foi revelado pelo jornal *The Washington Post* em março de 2018. O vazamento envolveu informações altamente sensíveis sobre programas militares secretos do Pentágono. Isso incluiu detalhes sobre o desenvolvimento e operações de armas avançadas, como mísseis hipersônicos, sistemas de defesa antimíssil e outros projetos classificados. Segundo o relatório do *The Washington Post*, os dados foram obtidos por meio de uma violação de segurança em uma empresa terceirizada que trabalhava com o Departamento de Defesa dos EUA. Os hackers conseguiram acessar essas informações confidenciais por meio de uma brecha nos sistemas da empresa. O vazamento dessas informações sensíveis representou uma ameaça significativa à segurança nacional dos Estados Unidos. Os detalhes sobre os programas militares secretos poderiam potencialmente comprometer a vantagem estratégica e tecnológica dos EUA em relação a outras nações

No Brasil, o Ministério da Saúde do Brasil foi alvo de um incidente de vazamento de dados que envolveu informações pessoais sensíveis de milhões de brasileiros. Este incidente ficou conhecido como o vazamento da base de dados do Sistema Único de Saúde (SUS), uma das maiores violações de dados já registradas no país. O vazamento envolveu a exposição de informações pessoais de aproximadamente 243 milhões de brasileiros, incluindo dados como nome completo, CPF, endereço, histórico de saúde, entre outros detalhes sensíveis. O vazamento gerou indignação e preocupação entre os brasileiros em relação à proteção de dados pessoais pelo governo e instituições públicas. Isso destacou a necessidade urgente de aprimorar a segurança cibernética e a proteção de dados em todo o setor público. Além disso, o vazamento provocou discussões profundas na sociedade quanto ao sistema de saúde e seu desempenho durante a Pandemia do Corona Virus e agravou a polarização política que o país vem sofrendo desde 2018.

A interconectividade e o armazenamento de grandes volumes de dados sensíveis em redes distribuídas tornam nações e suas infraestruturas críticas alvos primários para hackers e agentes mal-intencionados. Esses atores exploram vulnerabilidades em sistemas de segurança que não acompanham a velocidade e a complexidade do ritmo de coleta de dados em constante expansão que o Big Data trouxe ao século XXI. Diante dos desafios apresentados, é fundamental que os governos adotem medidas robustas para proteger seus sistemas de Big Data de ataques cibernéticos como implementar medidas de segurança rigorosas, como criptografia de dados, autenticação multifator e *firewalls* de última geração, para proteger os sistemas de Big Data contra acessos não autorizados, capacitar funcionários públicos em práticas de segurança cibernética para

identificar e evitar ataques de *phishing*, engenharia social e outras técnicas de *malware* e colaborar com outras nações e organizações internacionais para compartilhar inteligência sobre ameaças cibernéticas e desenvolver estratégias conjuntas de defesa cibernética. A proteção dos sistemas de Big Data é crucial para garantir a segurança nacional e a confiança dos cidadãos. Ao investir em medidas de Segurança Cibernética robustas e promover a conscientização sobre as ameaças cibernéticas, os governos podem fortalecer sua capacidade de se defender contra ataques cada vez mais sofisticados e proteger a informação sensível que é vital para o funcionamento da sociedade moderna.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] MARR, Bernard. ***Big Data: Como Extrair Volume, Variedade, Velocidade e Valor da Avalanche de Informações Cotidianas***. São Paulo: Editora Sextante, 2015.

[2] GRUS, Joel. *Data Science do Zero*. Sebastopol: O'Reilly Media, 2013

[3] UOL. **Dados pessoais de 24 milhões de usuários do SUS são vazados na internet**. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2019/04/11/dados-pessoais-de-24-milhoes-de-usuarios-do-sus-sao-vazados-na-internet.htm. Acesso em: 01 jul. 2024.

[4] RISKLOGIC. **How the Defense Force was hacked**. Disponível em: https://risklogic.com.au/how-the-defense-force-was-hacked/. Acesso em: 04 jul. 2024.

[5] CNN BRASIL. **Ex-funcionário da CIA é condenado por vazamento de dados ao WikiLeaks**. Disponível em: https://www.cnnbrasil.com.br/internacional/ex-funcionario-da-cia-e-condenado-por-vazamento-de-dados-ao-wikileaks/. Acesso em: 09 jul. 2024