O MEGAVAZAMENTO DO BRASIL

Artigo por Felipe Daragon e Equipe Syhunt. 12 de fevereiro, 2021

Agradecemos às mais de 8.000 empresas que nos contataram após o artigo inicial e solicitaram mais informações sobre como foram expostas pelo vazamento. Enquanto aguardamos as providências das autoridades brasileiras, continuamos monitorando as notícias e atualizações sobre o vazamento. Leia abaixo nossas descobertas e análise.

NOSSA ANÁLISE

Por meio de nossa análise de especialistas e participação em uma série de artigos pela mídia, ajudamos a destacar a dimensão do megavazamento que expôs dados de praticamente todos os brasileiros em janeiro de 2021.

Concluímos que entre 673 GB e 873 GB (quase 1 TB) de dados sobre empresas e indivíduos brasileiros foram roubados em 2020 e compilados em um único arquivo, provavelmente devido a vários vazamentos que ocorreram ao longo do tempo. Como resultado, detalhes importantes de um total surpreendente de 223 milhões de brasileiros e 40 milhões de empresas brasileiras foram expostos e estão sendo ativamente comercializados por cibercriminosos ou compartilhados por hackers em fóruns da Internet e na Dark Web.

Atendendo a uma solicitação do *Estadão*, analisamos o caso em parceria com o jornal. Revelamos uma série de números do vazamento, entre outros detalhes relevantes, com a intenção de informar o público e as empresas e aumentar a conscientização sobre o tamanho preocupante do vazamento

Dias depois, a publicação de um segundo artigo pelo Estadão levou o Supremo Tribunal Federal a ordenar uma investigação e o bloqueio de acesso às postagens e links do cibercriminoso. Desde então, uma investigação pelas autoridades brasileiras está em andamento. Análises adicionais da Syhunt em parceria com o jornal revelaram que 1) as fotos de rosto no arquivo do cibercriminoso eram, na realidade, copiadas do DivulgaCand e também que 2) meio milhão de números de celulares corporativos foram expostos.

O VAZAMENTO EM NÚMEROS

Seguindo a solicitação do *Estadão*, processamos o catálogo e pequenas amostras publicadas pelo cibercriminoso, simulamos exportações individuais de CSV e executamos cálculos para confirmar suas alegações, revelar uma "imagem" e com precisão **estimar** o tamanho total do vazamento e dos bancos de dados nas mãos do cibercriminoso:

223 milhões	40 milhões	104 milhões
Total de Brasileiros Expostos	Total de Empresas Expostas	Total de Veículos Expostos
37	17	1 (48 col.)
Categorias de Informação	Categorias de Informação	Categorias de Informação
650GB	200GB	23GB
Tamanho Aprox do Banco de Pessoas Descompactado	Tamanho Aprox do Banco de Empresas Descompactado	Tamanho do Banco Descompactado
3KB	4KB	200 bytes
Tamanho Estimado dos Dados Por Pessoa (Sem Foto)	Tamanho Estimado dos Dados Por Empresa	Tamanho Aprox. dos Dados Por veículo
39.645	22.983	288.167
Total de Pessoas nas Amostras	Total de Empresas em Amostras	Total de Veículos em Amostras
1.1M (aprox. 16GB)	N/A	N/A
Total de fotos de rosto	Total de fotos de rosto	Total de fotos de rosto

O VAZAMENTO EM NÚMEROS: TELEFONES

159 milhões	28 milhões
Total de brasileiros com detalhes de telefone expostos (Celular / Fixo): **159845321**	Total de empresas com detalhes de telefone expostos (Celular / Fixo): **28695845**
6.945	532,696
Total de números de celulares em amostras vazadas	Total de números de celulares corporativos em amostras vazadas

TOTAL DE NÚMEROS DE CELULARES CORPORATIVOS VAZADOS NAS AMOSTRAS - POR ESTADO

Paraná	205.640
São Paulo	202.829
Minas Gerais	19.801
Rio Grande do Sul	17.802
Rio de Janeiro	14.721
Distrito Federal	6.513
Santa Catarina	5.134
Espírito Santo	4.080
Mato Grosso	836
Tocantins	439

A ORIGEM DO VAZAMENTO

Chamamos o vazamento de BLB20 (Grande Vazamento do Brasil 2020), pois os dados do cibercriminoso estão atualizados até 2020. Muito se especulou sobre a origem do

vazamento e provavelmente saberemos mais sobre sua origem na medida em que avançar a investigação das autoridades brasileiras, pesquisadores de segurança da informação e organizações de mídia.

Parte da violação de dados pode ter sido um trabalho interno - realizado de forma deliberada e maliciosa por um funcionário de alguma empresa, uma opinião compartilhada por muitos pesquisadores de segurança. Acreditamos que os cibercriminosos ou alguma empresa de análise compilou vários vazamentos que ocorreram ao longo dos anos em um único arquivo, provavelmente armazenando tudo em um **disco SSD de no mínimo 1TB**. Concluímos e posteriormente o *Estadão* confirmou que as fotos de rosto no banco de dados eram copiadas do DivulgaCand do TSE, o que parece confirmar a compilação de dados de vários vazamentos e lugares diferentes.

O que disse o hacker? O cibercriminoso se referiu ao arquivo como um banco de dados da Serasa Experian. A Serasa Experian é uma grande empresa brasileira de pesquisa de crédito, mas a empresa afirmou que fez uma investigação interna e os dados do arquivo vazado não coincidem com os dados do banco de dados da empresa.

Outro megavazamento? Em 10 de fevereiro, surgiram relatos de um segundo mega vazamento, mas, embora venha de uma fonte confiável que alertou sobre o primeiro vazamento, devido à falta de referências, não fomos capazes de confirmar o novo vazamento. Esta análise e artigo são apenas sobre o primeiro vazamento.

AS CATEGORIAS DE INFORMAÇÕES VAZADAS

A seguir estão as categorias de informações expostas no vazamento e o **tamanho estimado** pela Syhunt de cada banco de dados individual:

🏢 DADOS DE EMPRESA / PESSOA JURÍDICA

Nome do Conjunto de Dados	Descrição	Tamanho Estimado
01 - Básico	Número do CNPJ, razão social, nome fantasia, cadastro (matriz / filial, situação), data de fundação, número de empregados, tamanho, natureza legal	8.3GB

02 - E-mail		2.9GB
03 - Telefone	Código de área, número, operadora, plano, tipo de linha (fixa, pré-paga, pós-paga), data de instalação	48.2GB
04 - Endereço	Endereço, número, bairro, cidade, estado, código postal, tipo (Residencial / Comercial), latitude e longitude	8.5GB
05 - Mosaic	Grupo de segmentação e subgrupo	1.7GB
06 - Empresarial	Nome e CPF dos sócios da empresa, participação (ações e percentual), data de entrada na empresa	45.9GB
07 - Receita Federal	Data de fundação, status do registro (ativa / baixada / inapta)	5.5GB
08 - Score de Crédito	Pontuação de risco, nível de risco (baixo / médio / alto)	2.2GB
09 - Representante Legal	CPF e nome do representante, situação cadastral (ativa / baixada / imprópria)	2.0GB
10 - Cheques Sem Fundos	Código do banco e agência, motivo (sem fundos / conta encerrada)	0.1GB
11 - Classe de Operação	Horário de funcionamento (24h, comercial 9h às 18h, almoço, noite etc.), tipo de distribuição (varejo físico, varejo online, atacado físico)	0.2GB
12 - Simples Nacional	Situação (optante / Não-optante)	4.3GB
13 - Natureza Jurídica	Empresa, empresário individual, cooperativa, agência pública, etc.	2.6GB
14 - Capital Social	Valor do capital social	1.7GB
15 - Devedores	Modalidade (principal, corresponsável), unidade responsável, cadastro, modalidade de crédito (multa, IRPJ, COFINS, CSLL etc.), valor	9.5 - 20 GB

16 - Sintegra	Número de registro do estado, data de início da atividade, status do registro	1.4GB
17 - CNAE		3.8GB
Tamanho total aprox. de todos os conjuntos de dados		**150 - 200GB**

DADOS PESSOAIS

- **01 - Básico**: nome da pessoa, CPF, sexo, data de nascimento, nome do pai, nome da mãe, estado civil (casado, solteiro, divorciado, viúvo, outros)
- **02 - E-mail**
- **03 - Telefone**: Código de área, número, operadora, plano, tipo de linha (fixa, pré-paga, pós-paga), data de instalação
- **04 - Endereço**: logradouro, número, bairro, cidade, estado, CEP, tipo (residencial / comercial), latitude e longitude domicílios: CPF do domicílio, número de pessoas, faixa de renda, endereço completo escolaridade: nível (analfabeto / fundamental / técnico / superior etc.)
- **05 - Mosaic**: grupo-alvo e subgrupo
- **06 - Emprego**: cargo, número de CBO (Classificação Brasileira de Ocupações)
- **07 - Score de Crédito**: atividade de crédito, pontuação de risco, nível de risco (baixo / médio / alto)
- **08 - Registro Geral (RG)**
- **09 - Título de Eleitor**: número de registro, zona, seção, endereço, município, estado
- **10 - Escolaridade**
- **11 - Empresarial**: nome do sócio de uma empresa, participação (quotas e percentual), razão social e nome fantasia da empresa, data de entrada na empresa
- **12 - Receita Federal**: situação (regular / suspenso / cancelado / falecido)
- **13 - Classe Social**: A1, A2, B1, B2, C1, C2, D, E
- **14 - Estado Civil**: casado, solteiro, divorciado, viúvo, outros
- **15 - Emprego**: CNPJ e razão social do empregador, número PIS / PASEP / NIT, número CTPS, tipo de vínculo (CLT, autônomo, servidor, aprendiz etc.), data de admissão, salário, horas de trabalho semanais
- **16 - Afinidade**: nível de precisão, percentil
- **17 - Modelo Analítico**: prevê chance do consumidor ter afinidade para comprar um produto ou serviço
- **18 - Poder de Aquisitivo**: nível (baixo, médio, alto), renda, salário
- **19 - Fotos de Rostos**: 1,176,157 imagens JPEG com datas entre 2012 e 2020; o nome do arquivo é o CPF da pessoa correspondente
- **20 - Servidores Públicos**: descrição do trabalho, capacidade, exercício, renda bruta, status, vínculo, afastamento (Sim / Não)
- **21 - Cheques sem Fundos**: código do banco e agência, motivo (sem fundos / conta encerrada)
- **22 - Devedores**: nome, tipo de devedor (principal, corresponsável), situação (ativo, em cobrança,

ajuizado), tipo de débito (multa, imposto de renda, PIS etc.), valor, foi judicializado? (Sim / Não)

- **23 - Bolsa Família**: valor, status do benefício (Liberado / Bloqueado), status do benefício (Ativo / Inativo), número e nome dos dependentes, NIS (Número de Identificação Social)
- **24 - Universitários**: 1,643,105 pessoas com nome da faculdade, curso, ano de entrada e ano de conclusão
- **25 - Conselhos**: 2,260,960 pessoas que prestam consultoria na esfera pública ou privada, incluindo situação, especialidade e código de ocupação
- **26 - Domicílios**: todas as pessoas que compartilham o mesmo endereço
- **27 - Vínculos**: categoriza as pessoas de acordo com o primeiro grau (mãe, pai, filho, filha, irmão, irmã, cônjuge) ou segundo grau (avô, neto, tio, sobrinho, primo, etc.)
- **28 - LinkedIn**: 5,051,553 perfis de redes sociais com número de identificação e URL de acesso
- **29 - Salário**: valor, tipo (mensal, quinzenal, semanal, etc.), horas por semana
- **30 - Renda**: valor mensal (inclui salário, aluguel, juros, etc.), classe social (baixa, média, alta), faixa de renda
- **31 - Óbitos**: data do óbito, idade, data da certidão de óbito, nome e endereço do cartório.
- **32 - IRPF (Imposto de Renda)**: nome da instituição bancária, código da agência, lote de reembolso
- **33 - INSS**: nome do segurado, número do benefício, data de início, tipo (aposentadoria, pensão, salário-maternidade, etc.)
- **34 - FGTS**: número do PIS
- **35 - CNS (Cartão Nacional de Saúde)**
- **36 - NIS (Número de Identificação Social)**
- **37 - PIS / PASEP**

Tamanho total aprox. de todos os conjuntos de dados: **500 - 650GB**

🚗 DADOS DE VEÍCULOS

- **ID**:número interno do banco de dados
- **Tipo de Pessoa**: físico ou legal
- **Data de Atualização**: varia de 1993 a 2020
- **Placa**: em formato antigo ou novo
- **Município e UF da Placa**
- **Situação do Veículo**
- **Restrições**: sem restrição, restringido por furto, penhor, alienação fiduciária, etc.
- **Número do Chassi**
- **Situação do Chassi**: Normal, Restrita
- **Número do Motor**
- **Número da Caixa de Câmbio** (se aplicável)
- **Número da Carroceria** (se aplicável)
- **Tipo de carroceria**: aberta, fechada, jipe, van, cabine dupla, motocicleta etc.
- **Tipo de documento faturado**

- **UF faturado**
- **Faturado**: contém sequência de números relacionados ao documento faturado, como fatura
- **Marca e Modelo**: existem 37 mil modelos diferentes
- **Ano Modelo**
- **Ano de Fabricação**
- **Cor do Veículo**
- **Tipo de Veículo**: bicicleta, ciclomotor, scooter, motocicleta, automóvel, ônibus, caminhão, etc.
- **Espécie de Veículo**: passageiro, carga, misto, tração, coleta etc.
- **Combustível**: gasolina, álcool, diesel, gás natural, elétrico, etc.
- **Poder**: poder em HP
- **Cilindradas**
- **Capacidade Máxima de Tração**
- **Peso Bruto Total**
- **Capacidade de Carga**
- **Número de Passageiros**
- **Número de Eixos**
- **Nacionalidade**: nacional ou importado
- **DI**: declaração de importação
- **Identidade da Importadora**
- **Tipo de Documento da Importadora**

COMO CHEGAMOS AOS NÚMEROS

Por meio de nossa colaboração com jornais e mídia, o que incluiu o *Estadão*, *Folha de São Paulo* e o *Tecnoblog*, para produzir as análises e estimativas acima, como pesquisadores e profissionais de segurança da informação de longa data, nós agimos com total lisura e responsabilidade - durante todo o processo, não buscamos entrar em contato com o cibercriminoso ou adquirir conjuntos de dados do hacker, e não obtivemos uma cópia de seu arquivo, que apenas estimamos o tamanho total. Além disso, em nenhum momento visamos lucrar financeiramente com o vazamento.

- **Tamanho estimado dos dados por pessoa (sem imagem do rosto)** e **Tamanho estimado dos dados por empresa**: com base nas amostras e no catálogo de conjuntos de dados fornecidos pelo cibercriminoso, simulamos uma exportação CSV de dados de indivíduos e empresas individuais. Concluímos, por exemplo, que os dados de negócios vazados sobre o próprio Syhunt giravam em torno de 7,33 KB de dados de texto. Depois de examinar o tamanho de várias exportações simuladas, estimamos o tamanho dos dados por pessoa e por empresa.
- **Tamanho aproximado dos dados por veículo** - o tamanho normal de cada linha do arquivo de veículos vazado.
- **Total de imagens de rostos no banco de dados** - estimado em 16-20GB: dividimos o tamanho em

bytes do arquivo de fotos de amostra (17,3 MB) por 1,334 arquivos JPEG. Em seguida, multiplicamos pelo número de fotos de rosto disponíveis no arquivo completo (1.1 milhões, ou para ser mais exato 1,176.157).

- **Tamanho estimado do banco de dados de pessoas (descompactado)**: multiplicamos o tamanho estimado dos dados por pessoa em bytes com o total de brasileiros expostos. Também adicionamos o tamanho estimado (descompactado) das fotos de rosto no banco de dados.
- **Tamanho estimado do banco de dados de empresas (descompactado)**: multiplicamos o tamanho estimado dos dados por empresa em bytes com o total de pessoas jurídicas brasileiras expostas. Também contamos as informações do catálogo por coluna de conjunto de dados e geramos as estimativas disponíveis acima.
- **Tamanho estimado do banco de dados (todos os bancos de dados descompactados)** - Quase 1 TB: a soma dos tamanhos estimados de banco de dados de Pessoas, Empresas e Veículos.

Em alguns casos específicos, utilizamos um pequeno software para contar o total de empresas afetadas.

CONCLUSÃO

Este é o maior e mais sério vazamento de dados que o Brasil já experimentou. A Syhunt recomenda esforços reais, imediatos e contínuos, por parte do governo e do setor privado, para responder vigorosamente a este vazamento, que deve incluir, entre outras coisas:

- Acelerar a resposta a este vazamento e vazamentos futuros.
- Combater a venda das informações vazadas.
- Impedir que os dados vazados sejam **explorados ativamente** por criminosos.
- Criar novos mecanismos para detectar, monitorar e denunciar vazamentos.
- Cooperação internacional com outras agências de aplicação da lei.
- Discutir e implementar contramedidas concretas com a ajuda das principais empresas e profissionais de segurança da informação.

SOBRE A SYHUNT SECURITY

Com tecnologia de análise de segurança de última geração, a Syhunt se estabeleceu como líder no campo da segurança de aplicações, fornecendo suas ferramentas de análise de segurança para diversas organizações em todo o mundo, indo de empresas pequenas e médias a empresas grandes. Os produtos da Syhunt ajudam as organizações a se defenderem contra uma ampla variedade de ataques cibernéticos sofisticados que ocorrem atualmente na camada de aplicações.

O Syhunt detecta proativamente vulnerabilidades e fraquezas que levam ao vazamento de dados ou violações - as ferramentas Syhunt se concentram nos vários ângulos e visões que podem ser usados para

avaliar o estado de segurança de uma aplicação, como sua versão live (por meio de análise dinâmica / DAST), código-fonte (SAST), log do servidor (perícia proativa) e configuração (hardening).

A Syhunt foi fundada pelo especialista em segurança Felipe Daragon, que assina este artigo, e que começou sua carreira trabalhando como consultor de segurança para organizações governamentais e empresas nos anos 90, atuando em empresas líderes de segurança da informação no Brasil e na segurança das páginas de divulgação de resultados das eleições do TRE-RJ. Os últimos 22 anos de Daragon no setor de segurança da informação foram dedicados a defender, proativamente, empresas e governo de ataques e aumentar a conscientização sobre questões urgentes de segurança e novas tendências dos ataques cibernéticos.

REFERÊNCIAS E AGRADECIMENTOS

AGRADECIMENTOS ESPECIAIS

1. Obrigado, Felipe Ventura, pelas primeiras análises detalhadas sobre o vazamento, que foram publicadas pelo Tecnoblog como parte de duas reportagens e passaram a destacar a dimensão do vazamento e nos ajudaram durante a nossa análise também. Agradeço a Renato Kopke por me enviar os links dos artigos do Tecnoblog.
2. Obrigado, Paulo R. Santos (Jump2) e Mario C. Fialho, por participar das análises junto com a Syhunt e os jornais.
3. Obrigado, Roberto F. Marc (Syhunt) por revisar as centenas de cálculos matemáticos e confirmar as estimativas da nossa análise.

REFERÊNCIAS

1. Megavazamento de dados de janeiro expôs mais de 500 mil celulares corporativos, Gizmodo. 11 de fevereiro de 2021
2. Megavazamento de janeiro fez meio milhão de celulares corporativos circularem na internet, Estadão, 10 de fevereiro de 2021
3. Fotos de megavazamento são de políticos que se candidataram entre 2012 e 2020, Canaltech, 5 de fevereiro de 2021
4. Fotos em megavazamento de dados são de candidatos nas eleições entre 2012 e 2020, Estadão, 4 de fevereiro de 2021
5. PF investiga venda de dados de Bolsonaro e de ministros do STF, CNN, 3 de fevereiro de 2021
6. Após megavazamento, dados de ministros do Supremo são postos à venda Conjur. 2 de fevereiro de 2021
7. Dados vazados podem render R$ 80,8 milhões ao criminoso Folha de São Paulo. 2 de fevereiro de 2021
8. Dados de Bolsonaro e ministros do STF estão à venda na internet após megavazamento Estadão, 1 de fevereiro de 2021
9. Após vazamento, dados de 40 mil pessoas já circulam na internet. CNN (Via Estadão), 29 de janeiro de 2021
10. Após megavazamento, dados de 40 mil brasileiros já circulam na internet, Estadão, 28 de janeiro de 2021
11. O que há no vazamento que afetou 40 milhões de CNPJs, Tecnoblog, 22 de janeiro de 2021
12. Vazamento que expôs 220 milhões de brasileiros é pior do que se pensava, Tecnoblog, 22 de janeiro de 2021