UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vinicius Dias Martins Rocha

Construção de um conjunto de dados de tráfego DNS para apoio à detecção de exfiltração de dados

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vinicius Dias Martins Rocha

Construção de um conjunto de dados de tráfego DNS para apoio à detecção de exfiltração de dados

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Rodrigo Sanches Miani

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2026

Vinicius Dias Martins Rocha

Construção de um conjunto de dados de tráfego DNS para apoio à detecção de exfiltração de dados

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 27 de março de 2026:

Rodrigo Sanches Miani
Orientador

Cristiano Lopes Martins Borges

Paulo Henrique Ribeiro Gabriel

Uberlândia, Brasil
2026

Agradecimentos

Agradeço, primeiramente, aos meus pais, Nice Iara e Flávio André, e à minha tia Sandra Luisa, por me mostrarem que a educação é uma ferramenta importantíssima e a chave para o sucesso pessoal e profissional. Sou grato também pelo apoio constante, mesmo à distância, independentemente de todas as dificuldades, e por serem minha inspiração dia após dia.

Aos professores da FACOM, com os quais tive o imenso prazer de compartilhar bons momentos, por compartilharem seus conhecimentos e por me prepararem para a vida além da faculdade. Em especial, ao meu orientador, Prof. Dr. Rodrigo Miani, pelo suporte constante, pela maneira como conduziu meu trabalho nessa reta final da graduação e por despertar em mim uma paixão pelo ambiente acadêmico.

Aos meus amigos de longa data, João Pedro Teixeira, Muryllo Henrique e Iury Gregório, que se mostraram presentes durante todo o processo — desde a preparação para os vestibulares até a conclusão deste ciclo — ajudando-me a adaptar-me à nova cidade e a superar as dificuldades, sem deixar de apreciar cada momento único.

Aos meus amigos de faculdade, Rafael Godoi, Alex Campos, Pablo Henrique, e a todos os demais colegas que, de forma direta ou indireta, contribuíram para esta jornada; que estiveram presentes comigo durante todo o período de graduação e me deram suporte nas matérias, nos estudos e na forma de lidar com os diversos contratempos que a vida tem a oferecer.

Resumo

A exfiltração de dados por meio do Sistema de Nomes de Domínio, do inglês Domain Name System (DNS) representa uma ameaça crítica à segurança da informação, explorando a permissividade deste serviço em ambientes corporativos para criar canais encobertos de comunicação. Em termos fundamentais, o DNS atua como o diretório da internet, traduzindo nomes de domínios legíveis em endereços IP. A exfiltração, por sua vez, consiste na transferência não autorizada de informações para fora de uma rede; quando realizada via DNS, o atacante fragmenta e oculta dados sigilosos dentro de consultas que mimetizam o tráfego legítimo, aproveitando que esse serviço raramente é inspecionado por firewalls convencionais.

Este trabalho apresenta a construção e validação de um conjunto de dados de tráfego DNS rotulado, focado nos protocolos Do53 e DoH, visando suprir a carência de bases de dados que reflitam o comportamento real das redes modernas. A metodologia abrangeu o processamento de capturas brutas no formato PCAPNG e a extração de fluxos estatísticos bidirecionais através do framework NFStream, integrando uma camada de engenharia de atributos para a criação de métricas sintéticas e comportamentais. Os resultados obtidos via Análise Exploratória de Dados (EDA) revelaram que métricas sintéticas isoladas, como a entropia e o comprimento das consultas, enfrentam desafios de distinção devido ao ruído gerado pela telemetria legítima, que frequentemente mimetiza padrões de ataque. Em contrapartida, os atributos volumétricos e temporais demonstraram uma capacidade superior de separação entre as classes, apresentando correlações positivas significativas com a atividade maliciosa. O conjunto de dados final, consolidado com aproximadamente 110.000 registros e 20 atributos de tráfego DNS tradicional (porta 53), foi disponibilizado publicamente no repositório Mendeley Data. Este recurso de alta fidelidade visa subsidiar o treinamento e a validação de futuros sistemas de detecção de intrusão (IDS) baseados em aprendizagem de máquina.

Palavras-chave: Domain Name System, Exfiltração de Dados, Conjunto de Dados, Tráfego DNS, Segurança da Informação, Análise Exploratória de Dados.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

DNS	Domain Name System (Sistema de Nomes de Domínio)
Do53	DNS tradicional
DoH	DNS over HTTPS (DNS sobre HTTPS)
IP	Internet Protocol (Protocolo de Internet)
TCP	Transmission Control Protocol (Protocolo de Controle de Transmissão)
UDP	User Datagram Protocol (Protocolo de Datagrama de Usuário)
TLS	Transport Layer Security (Segurança da Camada de Transporte)
PCAP	Packet Capture (Captura de Pacotes)
PCAPNG	PCAP Next Generation (Captura de Pacotes de Próxima Geração)
IDS	Intrusion Detection Systems (Sistemas de Detecção de Intrusão)
EDA	Exploratory Data Analysis (Análise Exploratória de Dados)

Sumário

1 Introdução

A exfiltração de dados é uma técnica amplamente utilizada por atacantes para transferir informações sigilosas para servidores externos sem levantar suspeitas. Com os avanços na computação, essas técnicas evoluíram, tornando sua detecção um desafio muito complexo nos dias atuais. Nesse contexto, a exfiltração via DNS destaca-se por sua discrição e eficiência. Como o DNS é um serviço pilar da internet, muitas ferramentas de segurança convencionais permitem seu livre tráfego, criando uma brecha explorada por mais de 91% dos ataques cibernéticos em alguma etapa de sua operação (ŽIžA; TADIć; VULETIć, 2023).

O impacto financeiro dessas brechas é severo. De acordo com o relatório da IBM Security (2023), o custo médio global de uma violação de dados em 2023 atingiu 4,45 milhões de dólares. Esse cenário impulsiona a necessidade de soluções baseadas em Aprendizado de Máquina, que se mostram promissoras na identificação de padrões maliciosos onde abordagens tradicionais falham.

A crescente adoção do protocolo DNS sobre HTTPS, do inglês DNS over HTTPS (DoH) intensifica as dificuldades de monitoramento, uma vez que as consultas DNS são encapsuladas em conexões TLS (Segurança da Camada de Transporte) na porta 443, confundindo-se com o tráfego de navegação comum. Essa convergência dificulta a diferenciação entre atividades legítimas e canais maliciosos de exfiltração. Paralelamente, o Do53 permanece como um vetor crítico para exfiltração de dados, favorecido por sua ampla utilização.

Para que modelos de detecção sejam eficazes, é necessário que o tráfego de rede seja traduzido em características quantificáveis que capturem o comportamento das comunicações. No entanto, existe uma carência de conjuntos de dados de alta qualidade que unam o tráfego tradicional (Do53) e o criptografado (DoH) em cenários realísticos. Pesquisas recentes sobre exfiltração via DNS abrangem tanto o tráfego convencional quanto o cifrado. Mahdavifar et al. (2021) focam no Do53 utilizando 30 atributos sintáticos e de resolução, enquanto Zhan et al. (2022) analisam o DoH através de 32 métricas temporais e volumétricas de fluxos TLS. Paralelamente, bases de dados públicas como o **CIC-Bell-DNS-EXF-2021** (LASHKARI et al., 2021) destacam-se pela captura de fluxos estatísticos em cenários de exfiltração via Do53, enquanto o **CIRA-CIC-DoHBrw-2020** (MONTAZERI-GHAFOORI et al., 2020) foca na caracterização de fluxos criptografados DoH através de metadados TLS.

A construção deste trabalho justifica-se pela escassez de repositórios que integrem tráfego Do53 e DoH sob um ambiente controlado e replicável. Identificou-se que as bases

de dados correlatas carecem de documentação detalhada sobre a origem dos dados e apresentam heterogeneidade no processamento de fluxos, o que impede análises comparativas robustas. Diante desse cenário, este trabalho propõe o desenvolvimento de um novo conjunto de dados que sane essas limitações, oferecendo rastreabilidade metodológica e um processo de extração padronizado.

O trabalho busca consolidar uma base de dados rotulada e documentada para que possa ser utilizada pela comunidade acadêmica. Além disso também foi realizada uma Análise Exploratória de Dados para investigar quais atributos possuem maior potencial discriminatório entre o tráfego benigno e as tentativas de exfiltração.

Portanto, este trabalho apresenta as seguintes contribuições:

- Contrução de um conjunto de dados que contém tráfego DNS tradicional (porta 53) em ambos os cenários, tanto de exfiltração quanto de casos benignos;
- Processo de extração e de rotulação dos tráfegos maliciosos e benignos.
- Uma análise exploratória do conjunto de dados contruído.
- Disponibilização pública do conjunto de dados para toda a comunidade científica;

1.1 Objetivos

Nesta seção, apresentam-se as metas norteadoras deste trabalho, divididas entre o propósito amplo da pesquisa e as etapas técnicas fundamentais para a sua viabilização. Tais objetivos delimitam o escopo da construção do conjunto de dados e os critérios utilizados para sua validação estatística.

1.1.1 Objetivo Geral

Este trabalho tem como objetivo geral desenvolver e validar um conjunto de dados estruturado para a detecção de exfiltração de dados via tráfego DNS, fundamentado em fluxos estatísticos e comprovado por meio de análise exploratória.

1.1.2 Objetivos Específicos

- Processar capturas de tráfego bruto no formato PCAPNG em fluxos bidirecionais utilizando o *framework* NFStream;
- Implementar uma camada de engenharia de atributos incluindo métricas de Entropia de Shannon, comprimento de consulta e densidade numérica;

- Automatizar a rotulagem e unificação de tráfego benigno e malicioso em uma estrutura tabular unificada;
- Realizar uma Análise Exploratória de Dados para investigar a separabilidade das classes e o comportamento dos atributos sintáticos frente ao tráfego de rede real;

Para nortear a investigação e validar a utilidade do conjunto de dados proposto, este trabalho busca responder às seguintes perguntas de pesquisa:

- Quais são as características do tráfego DNS que apresentam maior potencial para identificar atividades de exfiltração de dados?
- Quais atributos extraídos possuem maior correlação com a classe alvo?

1.2 Organização do Trabalho

Este trabalho está organizado em seis capítulos principais. No Capítulo 2, apresenta-se a fundamentação teórica necessária para a compreensão do estudo, abordando o funcionamento do protocolo DNS e suas variantes cifradas, as técnicas de exfiltração de dados e o papel dos Sistemas de Detecção de Intrusão, além de descrever as ferramentas e formatos tecnológicos como o PCAPNG e o *framework* NFStream. Na sequência, o Capítulo 3 é dedicado à revisão de trabalhos relacionados, onde são discutidas pesquisas contemporâneas sobre detecção de ameaças e a lacuna existente na oferta de conjuntos de dados realísticos.

O Capítulo 4 detalha o desenvolvimento e a metodologia aplicada, expondo as etapas de obtenção do tráfego bruto, os processos de extração de fluxos e a engenharia de atributos, culminando na rotulagem e unificação da base de dados final. Os resultados e a validação científica do recurso gerado são apresentados no Capítulo 5, por meio de uma Análise Exploratória de Dados fundamentada em métricas estatísticas, matrizes de correlação e diagramas de caixa. Por fim, o Capítulo 6 sintetiza as conclusões alcançadas, recapitulando o cumprimento dos objetivos propostos e sugerindo direcionamentos para investigações futuras no campo da segurança de redes.

2 Revisão Bibliográfica

Este capítulo apresenta a fundamentação teórica necessária para o embasamento dos conceitos explorados ao longo do trabalho. O objetivo é fornecer uma visão detalhada sobre a infraestrutura do protocolo DNS e suas variantes modernas, além de contextualizar as ameaças de exfiltração de dados e os mecanismos de detecção de intrusão. Para tanto, as seções subsequentes abordam as especificidades dos protocolos Do53 e DoH, as técnicas de exfiltração e os sistemas de monitoramento, finalizando com a descrição das ferramentas e formatos tecnológicos utilizados na construção do conjunto de dados.

2.1 Domain Name System

O Domain Name System (DNS) é um dos principais pilares da internet, atuando como um sistema de resolução de nomes que traduz endereços web (www.exemplo.com) em endereços IP (Internet Protocol) utilizáveis por computadores. Criado nos anos 1980, o DNS permitiu uma escalabilidade maior para o crescimento mundial da internet.

Nos dias atuais, o DNS não é apenas um serviço essencial para a navegação na web, mas também se tornou um alvo frequente de atacantes que buscam realizar exfiltrações de dados por meio de técnicas como DNS tunneling, spoofing, entre outros ataques. Além disso, o uso malicioso de domínios, onde cibercriminosos registram e utilizam domínios temporários para atividades como phishing, distribuição de malware e ataques de comando e controle, tem se tornado cada vez mais comum. Pesquisas como do modelo híbrido de Mahdavifar et al. (2021), entre outras, estão explorando o uso de machine learning para análise e detecção de atividades maliciosas via DNS.

Devido ao seu uso em grande escala e ao papel fundamental que desempenha na infraestrutura da internet moderna, o tráfego DNS frequentemente não recebe a devida atenção nas medidas de segurança. Muitas vezes, esse tráfego é liberado pela rede sem uma verificação adequada pelas ferramentas de segurança convencionais. Embora soluções como DNS sobre HTTPS (DoH) e DNS sobre TLS (DoT) tenham sido introduzidas para criptografar as requisições DNS, garantindo maior privacidade e proteção contra interceptações, porém acaba dificultando também a detecção de exfiltração de dados. Sobre detecção de DoH temos mais informações no trabalho de Zhan et al. (2022).

2.1.1 O Protocolo Do53

O Do53 é o modelo tradicional de resolução de nomes, fundamentado nas especificações das RFCs 1034 (MOCKAPETRIS, 1987a) e 1035 (MOCKAPETRIS, 1987b).

Ele foi projetado para traduzir domínios em endereços IP de forma rápida, geralmente operando na porta 53 usando o Protocolo de Datagrama do Usuário (UDP). Essa escolha pelo UDP é por uma questão de velocidade, já que ele oferece menor latência, o que é fundamental para a navegação na internet.

Por outro lado, um dos principais problemas do Do53 é que ele transmite as consultas em texto simples (plain text). Como não há criptografia nativa, qualquer nó na rede — desde o provedor de internet (ISP) até um atacante em uma rede Wi-Fi pública — consegue monitorar quais sites estão sendo acessados ou até mesmo manipular as respostas para redirecionar o tráfego.

2.1.2 O Protocolo DoH

Para melhorar a privacidade nessas consultas, surgiu em 2018 o padrão DoH (DNS over HTTPS), definido pela RFC 8484 (HOFFMAN; MCMANUS, 2018). A grande diferença é que o DoH embute as requisições DNS dentro de fluxos HTTPS comuns na porta 443, usando a criptografia TLS para proteger os dados e garantir mais segurança na comunicação.

Essa abordagem deixa as consultas DNS indistinguíveis do tráfego comum de navegação. Isso impede que terceiros visualizem o histórico de acesso do usuário. Além disso, o uso de certificados digitais garante que a resposta veio realmente do servidor DNS esperado. Vale ressaltar que o DoH desloca o controle da resolução de nomes do sistema operacional para o navegador, criando uma camada extra de proteção contra o monitoramento e a censura na rede.

2.2 Exfiltração de Dados via DNS

A exfiltração de dados é um dos diversos tipos de ataques cibernéticos nos quais informações sensíveis ou confidenciais são transferidas de um sistema ou rede sem autorização. Esse processo pode ser realizado por agentes internos ou externos, comprometendo a integridade e a segurança dos dados da organização (Fortinet, 2024). Com a evolução das defesas, atacantes passaram a utilizar métodos sofisticados para evitar detecção, destacando-se o uso de canais encobertos.

Historicamente, o termo evoluiu junto com os ataques cibernéticos, e diversos casos notáveis ao longo dos anos incluem vazamentos de dados de grandes corporações e instituições governamentais.

Como o tráfego DNS é vital para o funcionamento da internet, ele costuma ser pouco monitorado por firewalls tradicionais, o que abre brechas para o tunelamento. Essa técnica utiliza de consultas DNS comuns para transportar pedaços de dados codificados.

Assim, o atacante consegue transferir informações para fora da rede utilizando um canal que as ferramentas de segurança padrão consideram seguro.

Para ilustrar a dinâmica operacional dessa ameaça, a Figura 1 apresenta um cenário típico de ataque, detalhando o percurso da informação desde a sua captura na máquina comprometida até a reconstrução final pelo agente malicioso.

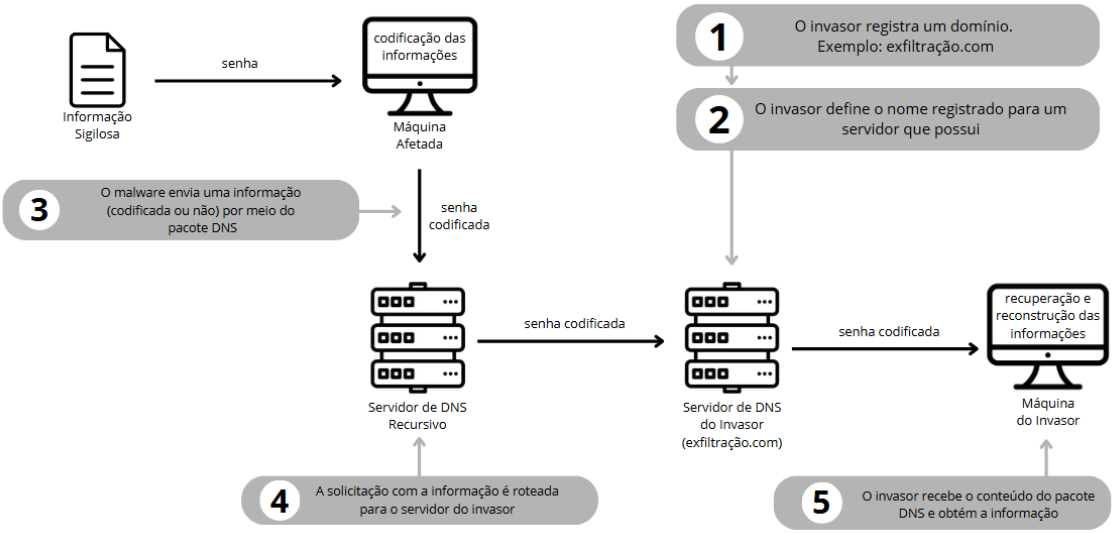

Figura 1 – Exemplo de um cenário típico de Exfiltração via DNS

Fonte: Do Autor

Devido à crescente sofisticação desses métodos, o combate a esses ataques tornou-se um desafio ainda maior com a popularização de protocolos como o DoH (DNS over HTTPS), que criptografa as consultas e impede a inspeção direta do conteúdo. Nesse cenário, o monitoramento estatístico de fluxos e a análise de tráfego tornam-se essenciais para identificar comportamentos maliciosos através da rede, permitindo a detecção de anomalias mesmo em canais cifrados (MAHDAVIFAR et al., 2021).

2.3 Sistemas de Detecção de Intrusão

Os Sistemas de Detecção de Intrusão (IDS) são ferramentas de segurança que monitora e identifica atividades suspeitas ou maliciosas em redes e sistemas computacionais. Eles operam analisando tráfego de rede, registros de eventos e comportamentos de certos usuários para detectar potenciais ameaças (IBM, 2025). Normalmente, os IDS não param os ataques por conta própria, geralmente são integrados em sistemas de segurança.

Existem dois principais tipos de IDS: baseados em assinaturas, que identificam ameaças conhecidas comparando as informações com uma base de dados de ataques pre-

viamente registrados, e o segundo tipo baseado em anomalias, que geralmente utilizam inteligência artificial para identificar inconsistências que possam indicar novas ameaças, para isso, temos algoritmos baseados em Random Forest, SVM e Redes Neurais.

Entretanto, confiar exclusivamente na verificação de TLS Fingerprinting pode ser insuficiente, pois esse método depende da comparação com uma base de dados pré-existente, tornando-se ineficaz contra novas ameaças que ainda não foram registradas. Nesse cenário, existe uma carência de conjuntos de dados de alta qualidade que contenham tráfego DNS realístico, especialmente em canais criptografados. O desenvolvimento de um novo conjunto de dados se torna, portanto, essencial, pois ele fornece a base necessária para o desenvolvimento futuro de métodos de aprendizado de máquina, aumentando a eficácia potencial dos Sistemas de Detecção de Intrusão.

2.4 Formato PCAPNG

O formato de arquivo *Packet Capture Next Generation* (PCAPNG) é uma evolução do modelo padrão de captura de pacotes. Esse novo modelo acaba superando grandes limitações do antigo formato PCAP. Conforme detalhado na especificação oficial mantida por especialistas como Erik Hjelmvik (TUEXEN et al., 2024), o PCAPNG utiliza uma arquitetura baseada em blocos, o que confere ao arquivo uma capacidade superior de armazenamento de metadados, informação essa muito valiosa para nosso conjunto de dados.

Diferente do formato PCAP, o PCAPNG permite que uma única captura contenha pacotes de múltiplas interfaces de rede e inclua comentários diretamente no arquivo. A estrutura modular é composta por blocos fundamentais:

- **SHB (*Section Header Block*):** Contém informações globais do arquivo e do sistema.
- **IDB (*Interface Description Block*):** Descreve as características das interfaces onde o tráfego foi coletado.
- **EPB (*Enhanced Packet Block*):** Armazena os pacotes capturados (como Do53 e DoH) com alta precisão de carimbo de tempo.

Com essas evoluções em mente, este trabalho aproveita os benefícios oferecidos pelo formato PCAPNG, visto que, para a construção de um conjunto de dados robusto e confiável, quanto mais informações detalhadas estiverem disponíveis, melhor será a sua consistência.

2.5 NFStream

Para realizar a análise de tráfego de rede escolhemos utilizar o *framework* de código aberto chamado NFStream, o qual foi projetado para extrair métricas detalhadas e estatísticas de fluxos em tempo real ou a partir de alguma captura pré-existente. Esse *framework* se destaca por sua alta performance e flexibilidade, permitindo uma análise rápida mesmo em redes de alta velocidade (BENEDICTIS, 2024).

Diferente de ferramentas tradicionais que se limitam a métricas básicas, o NFStream integra funcionalidades avançadas como por exemplo o *Deep Packet Inspection* (DPI), que é basicamente uma identificação de protocolos da camada de aplicação através do motor nDPI integrado. Outra função interessante que podemos citar é a extensibilidade via Plugins, que permite a criação de novos extratores de características utilizando Python.

O *framework* facilita também a exportação dos dados diretamente para estruturas como o formato CSV, tornando-o ideal para a construção de conjunto de dados voltados ao treinamento de modelos de inteligência artificial.

Para este trabalho, o NFStream atua como a ponte entre o dado bruto capturado no formato PCAPNG e a estrutura de dados organizada. Isso permite que as consultas Do53 e DoH sejam caracterizadas não apenas por seus endereços IP, mas também por padrões comportamentais e temporais, contribuindo bastante para a análise de segurança e classificação de tráfego.

2.6 Análise de Fluxo vs. Captura de Pacotes

Para compreender o monitoramento de tráfego de rede, foi necessário identificar duas abordagens principais: a captura de pacotes (Packet Capture - PCAP) e a análise de fluxo (Flow Analysis). A captura de pacotes consiste no registro integral de cada dado que trafega pela interface de rede, isso faz com que acabe oferecendo o nível máximo de granularidade ao permitir a inspeção completa de cabeçalhos e do conteúdo útil de cada pacote. Embora essa técnica seja muito importante para depuração detalhada de protocolos, ela acaba precisando de um alto consumo de recursos de armazenamento e processamento.

Em contrapartida, a análise de fluxo trabalha com fluxos que são definidos por um conjunto de características comuns, como a tupla feita pelos endereços IP de origem e destino, portas e protocolo. Em vez de armazenar o conteúdo bruto dos pacotes de uma mesma transação, essa abordagem registra apenas as estatísticas da conversação, como duração, volume total de bytes e contagem de pacotes (HOFSTEDE et al., 2014). Essa metodologia oferece alta escalabilidade e facilita a aplicação de algoritmos de Machine

Learning.

Tendo em vista esse trabalho, o framework NFStream desempenha um papel muito importante ao realizar a transição entre essas abordagens. Ele processa os arquivos brutos no formato PCAPNG e os converte em fluxos estatísticos, possibilitando que o comportamento do tráfego seja analisado de forma agregada. Para a detecção de protocolos como o DoH, essa estratégia é particularmente superior, pois permite que os modelos de inteligência artificial aprendam padrões comportamentais e temporais do tráfego em vez de depender da inspeção de um conteúdo que já se encontra protegido por criptografia HTTPS.

2.7 Conjunto de dados

O conjunto de dados é a base fundamental para o desenvolvimento de modelos de aprendizado de máquina, sendo definido como uma coleção de informações que representam ou estão relacionadas ao problema a ser resolvido (IBM, 2024b). No contexto desse trabalho, a importância de um conjunto de dados bem estruturado reside na sua capacidade de organizar variáveis complexas, como o tamanho dos pacotes e o tempo de chegada, dessa forma o algoritmo acaba conseguindo extrair correlações estatísticas. Além da estrutura, a densidade e a representatividade dos dados é um fator determinante conter uma quantidade suficiente de exemplos diversos para cobrir as variadas nuances do tráfego real, evitando que o modelo sofra de **overfitting** (quando o algoritmo apenas decora os dados de treino) ou **underfitting** (quando não consegue aprender o padrão).

Os benefícios de um conjunto de dados bem construído vão além da fase de treinamento, impactando diretamente na confiabilidade e na reprodutibilidade da pesquisa científica. Um conjunto de dados robusto permite que o modelo alcance maior generalização, tornando-o capaz de identificar protocolos criptografados com mais eficácia, mesmo em cenários adversos. A estruturação correta dos dados facilita a manutenção e a expansão do estudo, permitindo que novas características sejam adicionadas sem a necessidade de reestruturar todo o processo de análise, garantindo que o conjunto de dados seja eficiente e escalável.

2.8 Formato CSV

Um formato bastante conhecido e usado é o Comma-Separated Values (CSV), esse é um padrão de armazenamento de dados amplamente utilizado para a representação de conjuntos de dados tabulares de forma simples e eficiente. Definido tecnicamente pela RFC 4180 (SHAFRANOVICH, 2005), o CSV organiza as informações em texto plano, onde cada linha do arquivo acaba representando um registro e os campos são delimitados

por um caractere específico, usualmente a vírgula. No contexto da análise de tráfego de rede e ciência de dados, a relevância deste formato reside na sua independência de plataforma e na facilidade de integração com diversas ferramentas de análise.

Ao converter fluxos de rede complexos em registros CSV, viabiliza-se a estruturação do modelo proposto, pois servem como entrada direta para algoritmos de aprendizado de máquina, permitindo que o conjunto de dados proposto com suas respectivas características extraídas dos protocolos Do53 e DoH sejam manipuladas por bibliotecas de análise estatística.

2.9 Rotulagem de Dados

Como nosso objetivo aqui é a construção de um conjunto de dados bem estabelecido e elaborado, é fundamental garantir que o processo de rotulagem de dados é eficaz e muito eficiente. A rotulagem consiste na atribuição de uma etiqueta ou classe específica a cada registro do conjunto de dados, servindo como se fosse uma verdade fundamental, a qual o algoritmo utilizará para aprender padrões e realizar predições. No contexto da segurança de rede e detecção de protocolos, uma rotulagem precisa é o que diferencia um sistema robusto de um modelo propenso a falsos positivos. Se os tráfegos de Do53 e DoH forem incorretamente rotulados, o modelo incorporará ruídos estatísticos que comprometem sua capacidade de assertividade em cenários reais.

Por isso a construção de um conjunto de dados confiável exige que a rotulagem seja realizada de forma determinística, utilizando fontes de dados onde a identidade do tráfego seja conhecida. Para este trabalho, a precisão na marcação dos fluxos é vital. Uma rotulagem bem estruturada não apenas facilita o treinamento, mas é o requisito primordial para que as métricas de avaliação, como precisão e sensibilidade, retratem os resultados reais da solução proposta (IBM, 2024a).

2.10 Análise Exploratória de Dados

A Análise Exploratória de Dados é uma abordagem metodológica que visa examinar conjuntos de dados para sintetizar suas características principais, frequentemente utilizando métodos de visualização. Diferente da estatística confirmatória, que busca testar hipóteses pré-definidas, a análise exploratória atua como um processo investigativo fundamental para compreender a estrutura, identificar anomalias e descobrir padrões subjacentes antes do conjunto ser de fato aplicado em modelos de aprendizado de máquina (TUKEY, 1977).

Segundo John Tukey, precursor desta filosofia, essa abordagem assemelha-se ao trabalho de investigação, buscar evidências nos dados que orientem a escolha das fer-

ramentas de análise mais adequadas. No contexto de segurança de redes, a exploração dos dados permite validar se os atributos extraídos apresentam variância suficiente para distinguir o tráfego legítimo do malicioso. Através de técnicas é possível identificar quais variáveis são mais influentes para o problema estudado, garantindo que o conjunto de dados final possua relevância informativa.

3 Trabalhos Relacionados

Este capítulo apresenta uma revisão da literatura sobre a detecção de exfiltração de dados via DNS, estabelecendo uma base comparativa para a proposta deste estudo. O objetivo é analisar pesquisas contemporâneas que exploram diferentes abordagens de modelagem, conjuntos de dados e técnicas de extração de atributos em cenários de tráfego Do53 e DoH. Para tanto, as discussões a seguir detalham estudos focados em aprendizagem de máquina, as principais bases de dados públicas utilizadas pela comunidade científica e abordagens diferentes para esse tema.

Diversas pesquisas contemporâneas dedicam-se à identificação de exfiltração de dados através do protocolo DNS, abordando tanto o modelo tradicional quanto suas versões cifradas. No âmbito do tráfego convencional (Do53), o trabalho de Mahdavifar et al. (2021), intitulado *Lightweight hybrid detection of data exfiltration using DNS based on machine learning*, estabelece uma metodologia de detecção fundamentada em aprendizado de máquina que prioriza a análise comportamental das resoluções e metadados de domínio. Os autores adotaram uma estratégia de extração de atributos derivados tanto do nome de domínio quanto de metadados de resolução, totalizando 30 variáveis distintas. Entre os algoritmos testados, destacam-se modelos capazes de lidar com a complexidade estatística do tráfego de rede para reduzir falsos positivos em sistemas de detecção de intrusão (IDS). Assim como a referida pesquisa, este trabalho realiza um pré-processamento rigoroso para garantir a relevância das variáveis; contudo, enquanto Mahdavifar et al. utilizam uma base focada na estrutura sintática das consultas, a presente investigação expande a análise para o comportamento de fluxos estatísticos, buscando englobar também o protocolo DoH.

Sob outra perspectiva, voltando-se para a privacidade e os desafios da criptografia, Zhan et al. (2022) concentram seus esforços no protocolo DoH em seu estudo *Detecting DNS over HTTPS based data exfiltration*. Diferente de abordagens que tratam o DNS tradicional, os autores analisaram fluxos criptografados onde a inspeção de conteúdo é inviabilizada pelo TLS. Nesse caso, os autores utilizam metadados de fluxo para priorizar estatísticas de tamanho e tempo dos pacotes. O conjunto de dados empregado por Zhan et al. combina tráfego benigno e malicioso, permitindo a extração de métricas que tornam o DoH indistinguível da navegação comum para ferramentas convencionais. A pesquisa reforça a necessidade de estruturar características que capturem a natureza temporal das consultas, premissa que converge com este trabalho, embora este busque uma base de dados mais focada na diversidade de cenários de exfiltração.

Complementarmente aos estudos de modelagem, o cenário de conjuntos de dados públicos oferece referências como o CIC-Bell-DNS-EXF-2021 (LASHKARI et al., 2021),

desenvolvido pelo Canadian Institute for Cybersecurity. A principal característica desta base é o seu foco na análise baseada em fluxos em detrimento da inspeção isolada de pacotes, permitindo capturar o comportamento estatístico da comunicação, incluindo métricas de volume e duração. A semelhança metodológica com este TCC é evidente, uma vez que ambos utilizam a extração via frameworks de análise de fluxo — como o NFStream aqui empregado — para converter dados brutos em variáveis comportamentais. No entanto, enquanto o CIC-Bell foca primariamente no cenário corporativo de Do53, este estudo estende a análise para o protocolo DoH, unindo a robustez estatística à diversidade de protocolos das redes modernas.

Além das bases mencionadas, destaca-se o *DNS Exfiltration Dataset*, um repositório técnico focado na análise de consultas DNS que utiliza algoritmos de detecção baseados em assinaturas e comportamentos de subdomínios. Embora apresente um volume massivo de dados, essa base concentra-se primordialmente no tráfego Do53. A fim de proporcionar uma visão consolidada do cenário atual, a Tabela 1 apresenta uma análise comparativa entre as bases de dados anteriormente citadas e o conjunto proposto nesta pesquisa, evidenciando os tipos de tráfego e volume de amostras.

Tabela 1 – Comparativo entre Conjuntos de Dados de Tráfego DNS Públicos e o Conjunto Proposto.

Conjunto de dados	**Do53**	**DoH**	**Benigno**	**Exfiltração**	**Eng. Sintática**	**Proporção (B:M)**	**Total de Instâncias**
DNS Exfiltration Dataset	V	X	V	V	X	≈ 1:1,2	13.000.000
CIC-Bell-DNS-EXF-2021	V	X	V	V	Limitado	≈ 1:1	1.019.318
CIRA-CIC-DoHBrw-2020	X	V	V	V	X	≈ 1:1	1.167.050
UFU-Exfiltração-DNS (proposto)	**V**	**X**	**V**	**V**	**V**	**≈ 10:1**	**109.359**

Fonte: Do autor.

Em síntese, a análise destes trabalhos reitera que a eficácia da detecção de intrusão é dependente da qualidade e diversidade das bases de dados utilizadas. Enquanto as pesquisas citadas focam na aplicação de algoritmos em protocolos isolados, identifica-se uma lacuna na oferta de conjuntos de dados que integrem, de forma coesa, múltiplos cenários de exfiltração em Do53 e DoH simultaneamente. Portanto, o foco desta pesquisa não reside na predição algorítmica, mas na entrega de um recurso informacional denso. Ao estruturar um conjunto de dados via NFStream que abranja padrões comportamentais e temporais, busca-se fornecer o suporte necessário para que futuras investigações possam evoluir com dados de maior fidelidade e abrangência.

4 Desenvolvimento

O presente capítulo descreve o processo de construção e estruturação do conjunto de dados voltado à identificação de exfiltração de dados em tráfego DNS.

Para detalhar essa trajetória, o conteúdo está organizado nas seguintes seções:

- **Visão Geral:** apresenta o fluxo metodológico das cinco etapas que compõem o ciclo de vida deste trabalho;
- **Obtenção e Caracterização dos Dados:** detalha a origem das capturas brutas PCAPNG e os cenários de rede simulados na infraestrutura de nuvem;
- **Extração de Fluxos via NFStream:** descreve a utilização do *framework* para a conversão de tráfego binário em dados estruturados, destacando a camada de engenharia de atributos e as limitações identificadas no tráfego DoH;
- **Rotulagem de Dados:** explica o processo de atribuição determinística de classes (*labels*) e os filtros aplicados para garantir a integridade da base;
- **Caracterização e Disponibilização do Conjunto de Dados:** expõe as propriedades técnicas finais do *dataset*, incluindo a distribuição de classes e as diretrizes para acesso e reprodutibilidade através do GitHub e Mendeley Data.

4.1 Visão Geral

Este trabalho apresenta uma metodologia voltada ao processamento de dados brutos de rede para a construção e validação de um conjunto de dados focado em tráfego DNS. O objetivo é transformar capturas de pacotes em uma base de dados rotulada, robusta e representativa. O fluxo de trabalho adotado está estruturado em cinco etapas principais, conforme ilustrado na Figura 2: obtenção das capturas brutas, extração de fluxos, rotulagem, disponibilização do conjunto de dados e análise exploratória estatística.

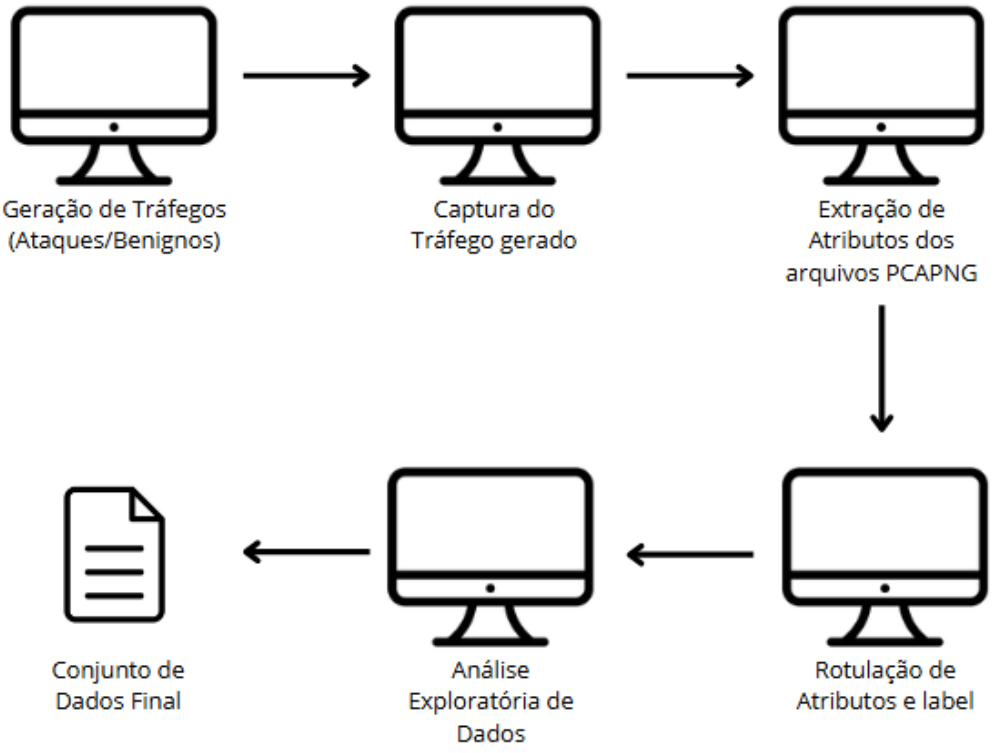

Figura 2 – Fluxo das Etapas do processo de construção do Conjunto de Dados

Fonte: Do Autor

A etapa inicial consiste na utilização de arquivos no formato PCAPNG contendo registros de tráfego DNS. Essas capturas simulam ambientes de rede reais e abrangem tanto consultas convencionais quanto cenários de exfiltração de dados, garantindo a fidelidade necessária para a análise subsequente.

Na segunda etapa, é realizada a extração de fluxos estatísticos a partir desses arquivos utilizando o NFStream. Esta ferramenta processa as capturas para agrupar pacotes individuais em fluxos bidirecionais, calculando automaticamente métricas temporais e de volume. O uso do NFStream permite converter o dado binário bruto em um conjunto de características quantificáveis e estruturadas.

Em seguida, os fluxos extraídos são submetidos ao processo de rotulagem e filtragem. Esta etapa é realizada por meio de scripts automatizados que utilizam o contexto de origem das capturas para atribuir etiquetas de forma determinística e para filtrar instâncias que não pertencem ao escopo do trabalho. Cada registro é classificado em uma de duas categorias: “benigno” ou “exfiltração”.

Na quarta etapa, o conjunto de dados resultante é estruturado no formato CSV e disponibilizado publicamente no repositório Mendeley Data. O conjunto de dados é acom-

panhado de documentação técnica detalhada, descrevendo as características extraídas e o processo de organização, garantindo a transparência e a reprodutibilidade da pesquisa.

Por fim, a metodologia encerra-se com a Análise Exploratória de Dados. Esta fase consiste na aplicação de métodos estatísticos descritivos e técnicas de visualização, como matrizes de correlação e diagramas de caixa, para investigar a separabilidade das classes. O objetivo desta etapa final é validar a eficácia dos atributos extraídos e identificar padrões de comportamento anômalo, fornecendo uma base sólida para a interpretação dos resultados e futuras aplicações em modelos de aprendizagem de máquina.

4.2 Obtenção e Caracterização dos Dados

A estruturação deste conjunto de dados iniciou-se com o processamento de capturas brutas no formato PCAPNG. É importante ressaltar que a geração desse tráfego não integrou o escopo de execução deste trabalho; os arquivos foram fornecidos previamente, fazendo parte do projeto de doutorado do aluno Cristiano Borges, do PPGCO/FACOM. Tais capturas serviram como material fundamental para as etapas subsequentes de análise e extração de atributos, sendo que o artigo completo detalhando o rigoroso processo de coleta desses dados será submetido ao Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2026).

A geração desses dados brutos envolveu a configuração de um ambiente controlado em computação em nuvem na Amazon Web Services (AWS), projetado para simular uma rede corporativa real. A topologia foi estruturada em uma rede privada virtual com segmentação lógica entre as redes de gerenciamento, do atacante e da vítima, utilizando um dispositivo de segurança (firewall) FortiGate como ponto central de roteamento e coleta de tráfego.

O tráfego foi gerado através de componentes específicos: uma máquina Ubuntu atuou como estação de trabalho da vítima para tráfego benigno e rotinas de exfiltração; um Kali Linux concentrou as ferramentas do atacante, como o dnscat2 para Do53 e o DoHC2 para DoH; e um servidor Windows realizou as capturas em portas espelhadas. Essa estrutura permitiu a extração consistente de arquivos contendo fluxos reais de comunicação.

Os dados recebidos consistem em uma série de arquivos que representam diferentes comportamentos de rede, divididos entre tráfego benigno (consultas DNS convencionais) e tráfego malicioso (cenários de exfiltração). A Tabela 2 detalha o inventário dos arquivos brutos utilizados.

Estes arquivos abrangem quatro cenários fundamentais, projetados para oferecer uma base comparativa sólida:

Tabela 2 – Relação de Arquivos PCAPNG Recebidos.

Nome do Arquivo	**Protocolo**	**Rótulo**	**Tamanho (KB)**
Dia1_Do53_Benigno	DNS (Porta 53)	0	718.200
Dia2_DoH_Benigno	DoH (Porta 443)	0	355.839
Dia3_Do53_Benigno	DNS (Porta 53)	0	692.132
Dia4_DoH_Benigno	DoH (Porta 443)	0	280.781
Dia5_Do53_Benigno	DNS (Porta 53)	0	694.146
Dia6_DoH_Benigno	DoH (Porta 443)	0	261.390
Dia7_Do53_Benigno	DNS (Porta 53)	0	694.021
Dia8_DoH_Benigno	DoH (Porta 443)	0	204.930
Dia9_Do53_Benigno	DNS (Porta 53)	0	694.076
Dia10_DoH_Benigno	DoH (Porta 443)	0	1.348.075
Dia11_Do53_Benigno	DNS (Porta 53)	0	2.322.626
Dia12_Do53_Exfiltração	DNS (Porta 53)	1	51.526
Dia13_Do53_Exfiltração	DNS (Porta 53)	1	45.900
Dia14_Do53_Exfiltração	DNS (Porta 53)	1	90.074
Dia15_Do53_Exfiltração	DNS (Porta 53)	1	1.060.473
Dia16_DoH_Exfiltração	DoH (Porta 443)	1	212.175

Fonte: Do autor.

- **Do53 Benigno:** Consultas DNS tradicionais (porta 53/UDP) automatizadas via *scripts*.
- **Do53 com Exfiltração (dnscat2):** Canal de Comando e Controle (C2) utilizando a ferramenta *dnscat2*.
- **DoH Benigno:** Requisições DNS sobre HTTPS (porta 443/TCP) para provedores públicos.
- **DoH com Exfiltração (DoHC2):** Uso da ferramenta *DoHC2* para exfiltrar arquivos através de requisições HTTPS.

Essa diversidade assegura que o conjunto de dados final contenha as variações necessárias para que modelos de detecção possam aprender padrões distintos de exfiltração. Com os dados em mãos, o próximo passo consiste na extração das informações relevantes para a detecção de exfiltração de dados.

4.3 Extração de Fluxos via NFStream

Após a aquisição dos arquivos brutos, a etapa seguinte é o processamento das capturas PCAPNG para a extração de fluxos. Para esta tarefa, foi utilizado o *framework* NFStream (BENEDICTIS, 2024), uma ferramenta de análise de rede baseada em Python que permite a conversão de tráfego binário em registros tabulares estruturados. Diferente

da inspeção individual de pacotes, a análise por fluxos agrupa sequências de comunicações que compartilham a mesma tupla de identificação, permitindo uma visão comportamental da rede.

4.3.1 Atributos Selecionados

Para melhorar a capacidade discriminatória do conjunto de dados proposto, o processo de extração não ficou apenas atrelado às métricas nativas da ferramenta. Foi desenvolvido um roteiro de automação que implementa uma camada de Engenharia de Atributos, criando indicadores específicos para a detecção de anomalias no protocolo DNS.

Um dos pilares desta análise é o cálculo da Entropia de Shannon aplicado ao campo do nome de domínio solicitado (`dns_query_entropy`). Matematicamente, a entropia mensura o grau de incerteza ou aleatoriedade das informações, sendo definida pela seguinte equação:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \tag{4.1}$$

Neste contexto, enquanto nomes de domínio legítimos tendem a seguir padrões linguísticos, dados exfiltrados, que são frequentemente codificados em formatos como Base64, apresentam uma distribuição de caracteres muito mais aleatória, resultando em valores de entropia mais elevados.

Outro atributo implementado foi o comprimento da consulta (`dns_query_length`), o qual é fundamental para mensurar o tamanho total do nome de domínio solicitado. Em cenários de exfiltração, os atacantes buscam maximizar o transporte de informações em cada transação, o que resulta em nomes de domínio consideravelmente mais longos do que os observados em resoluções convencionais, caracterizando assim um caso típico de tunelamento.

Complementarmente, a proporção de caracteres numéricos (`dns_numerical_ratio`) analisa a densidade de dígitos presentes no domínio. Uma vez que a exfiltração de arquivos geralmente envolve a transmissão de dados binários codificados, o surgimento frequente de numerais é um forte indicativo de que o tráfego não representa uma navegação humana comum.

Por fim, o último atributo que foi implementado é a contagem de subdomínios (`dns_subdomain_count`), responsável por monitorar a fragmentação da consulta. Esta métrica é estratégica para identificar tentativas de burlar limites de tamanho de pacotes DNS. Através da criação de múltiplas camadas de subdomínios, scripts de ataque conseguem particionar grandes volumes de dados em diversas etiquetas dentro de uma única consulta, uma estratégia recorrente para manter o canal de comunicação malicioso sem

interromper o fluxo de dados.

4.3.2 Estratégia de Processamento e Padronização

Considerando o elevado volume de dados dos arquivos originais, o código foi estruturado para realizar o processamento em lotes (*chunks*) de 50.000 registros. Essa abordagem de engenharia de *software* foi muito importante para a gestão eficiente da memória RAM, mitigando gargalos de processamento que impediam a finalização da tarefa devido à densidade das informações contidas nos arquivos PCAPNG. A implementação dessa estratégia garantiu a escalabilidade do fluxo de trabalho e a integridade dos dados, permitindo o processamento de grandes volumes de tráfego sem a interrupção por falta de recursos computacionais.

Por fim, estabeleceu-se uma lista de Atributos, garantindo que tanto os fluxos de tráfego convencional (Do53) quanto os criptografados (DoH) fossem exportados para o formato CSV com uma estrutura idêntica. Essa padronização é fundamental para assegurar a consistência do recurso informacional, permitindo que futuras investigações utilizem o conjunto de dados de forma integrada.

A seleção dos atributos que compõem este conjunto de dados foi orientada por uma abordagem híbrida, combinando métricas consagradas na literatura de detecção de exfiltração de dados com capacidades avançadas de análise de fluxo oferecidas pelo *framework* NFStream.

Primeiramente, foram selecionados por meio do NFStream os atributos considerados identificadores de fluxos, estes servem para contextualizar a comunicação, identificando as entidades envolvidas na troca de dados. Embora nem todos sejam usados diretamente no treinamento do modelo, eles são essenciais para a fase de rotulagem e auditoria. Os atributos pertencentes a esta categoria são: `src_ip`, `dst_ip`, `src_port`, `dst_port` e `protocol`.

Em seguida, ainda utilizando o NFStream foram definidos os atributos de fluxo e temporais, que constituem a base comportamental da metodologia. Diferente da análise estática, estas métricas permitem observar a dinâmica da comunicação ao longo do tempo, sendo fundamentais para identificar a persistência característica de exfiltrações via DNS. Atributos como `bidirectional_duration_ms` e `bidirectional_packets` são cruciais, pois, enquanto consultas legítimas são geralmente efêmeras, a exfiltração de dados exige uma sessão prolongada e um fluxo constante de pacotes para transportar informações. Complementarmente, os atributos de volume, como `bidirectional_bytes`, `src2dst_bytes` e `dst2src_bytes`, permitem detectar a assimetria típica de ataques de exfiltração, onde há um envio massivo de dados para o exterior da rede. Também foram incluídos os registros de `start_time` e `end_time` para garantir a precisão cronológica dos

fluxos extraídos.

Adicionalmente, utilizando as consultas, conseguimos por meio de engenharia de dados, criar alguns atributos sintáticos, que analisam a estrutura textual das consultas DNS. Esta categoria é amplamente fundamentada na literatura de segurança, pois atacantes frequentemente utilizam codificações para embutir dados dentro dos campos de query, alterando as propriedades naturais da linguagem. O atributo `dns_query_entropy` é utilizado para medir a aleatoriedade da string, enquanto o `dns_query_length` identifica consultas anormalmente longas, típicas de túneis que buscam maximizar a taxa de transferência. Complementam esta análise os atributos `dns_subdomain_count` e `dns_numerical_ratio`, que auxiliam na detecção de domínios com múltiplos níveis hierárquicos ou alta densidade de caracteres numéricos, padrões comuns em ferramentas de exfiltração.

Por fim, incluíram-se os atributos de contexto e aplicação, derivados do motor de inspeção profunda (Deep Packet Inspection - DPI) integrado ao framework NFStream. Os campos `application_name` e `application_category_name` fornecem uma camada extra de inteligência ao classificar o tráfego com base em assinaturas de rede conhecidas. O atributo `requested_server_name` e a variável alvo `label` encerram a estrutura do conjunto de dados, permitindo que o modelo correlacione o comportamento do fluxo com o serviço de rede solicitado e a respectiva classificação entre tráfego benigno ou malicioso.

Como resultado do processo de extração e transformação, o conjunto de dados consolidado totaliza 20 atributos. Esta seleção integra identificadores de rede, métricas temporais, volumétricas e atributos customizados de engenharia de dados. A Tabela 3 apresenta a relação completa dessas variáveis e suas respectivas representações técnicas.

Tabela 3 – Relação Final de Atributos que Compõem o Conjunto de Dados.

Atributo (*Feature*)	**Representação Técnica**
`src_ip`	Endereço IP de origem da conexão.
`dst_ip`	Endereço IP de destino da conexão.
`src_port`	Porta lógica de origem utilizada pelo cliente.
`dst_port`	Porta lógica de destino (ex: 53 para Do53 ou 443 para DoH).
`protocol`	Número do protocolo de transporte (ex: 6 para TCP, 17 para UDP).
`start_time`	Data e hora exata (timestamp) do início do fluxo.
`end_time`	Data e hora exata (timestamp) do término do fluxo.
`bidirectional_bytes`	Volume total de bytes trafegados em ambas as direções.
`bidirectional_packets`	Quantidade total de pacotes trocados no fluxo.
`bidirectional_duration_ms`	Duração total da comunicação em milissegundos.
`src2dst_bytes`	Volume de bytes enviados da origem para o destino.
`dst2src_bytes`	Volume de bytes enviados do destino para a origem.
`application_name`	Nome da aplicação identificado pelo NFStream.
`application_category_name`	Categoria da aplicação (ex: *Network Service*).
`requested_server_name`	Nome do servidor ou domínio solicitado no fluxo.
`dns_query_length`	Comprimento total do domínio consultado.
`dns_query_entropy`	Grau de aleatoriedade dos caracteres no domínio.
`dns_subdomain_count`	Quantidade de subdomínios identificados na consulta.
`dns_numerical_ratio`	Mensura a proporção de algarismos presentes no nome do domínio em relação ao seu comprimento total.
`label`	Variável alvo para classificação binária. Indica a natureza do fluxo: 0 para tráfego benigno e 1 para exfiltração de dados.

Fonte: Do autor.

4.3.3 Limitações na Extração de Atributos DoH

Durante a fase de processamento, identificou-se uma limitação crítica nos arquivos PCAPNG referentes à exfiltração via DNS over HTTPS (DoH). Ao submeter essas capturas ao *framework* NFStream, verificou-se que a ferramenta não foi capaz de extrair nenhum registro de fluxo das amostras maliciosas, resultando em arquivos de saída vazios.

Esse comportamento foi causado por instabilidades na camada de transporte (TCP) durante a geração do tráfego de ataque. A presença massiva de sinalizações de reinicialização de conexão (RST) impediu o estabelecimento do aperto de mão entre o cliente e o servidor. Como o NFStream prioriza a extração de fluxos estruturados e funcionais, as tentativas de conexão interrompidas foram descartadas pela ferramenta.

Diferente do cenário de exfiltração, o tráfego DoH benigno foi processado sem intercorrências, confirmando que a falha estava restrita à integridade das capturas do

cenário de ataque. Diante da impossibilidade de extrair métricas consistentes para a classe maliciosa de DoH e visando evitar um desequilíbrio artificial no conjunto de dados, optou-se por restringir o escopo deste trabalho à detecção de exfiltração sobre o protocolo Do53. Esta decisão assegura que o conjunto de dados final seja composto integralmente por fluxos válidos e estatisticamente representativos para o treinamento dos modelos de detecção.

Embora a impossibilidade de extrair fluxos do cenário malicioso de DoH tenha imposto uma restrição de escopo, este evento reforçou a robustez e o comportamento dos mecanismos de monitoramento. Foi evidenciado a sensibilidade operacional do NFStream: a ferramenta demonstrou um rigoroso critério de filtragem ao descartar conexões incompletas, o que garante que o conjunto de dados final não tenha tentativas de conexão falhas, conexões essas que não representam uma comunicação de dados real.

4.4 Rotulagem de dados

Após a extração dos atributos, os dados estavam fragmentados em diversos arquivos CSV, correspondentes a cada dia de captura. Para transformar esses registros isolados em um conjunto de dados pronto para o aprendizado de máquina, foi desenvolvido um roteiro de automação em Python focado em rotulagem binária, filtragem e gestão de recursos computacionais.

4.4.1 Atribuição de Rótulo e Filtros

A atribuição da variável alvo (*label*) foi realizada de forma programática e determinística, utilizando uma abordagem de rotulagem automática baseada na fonte dos dados. Este processo é fundamental para garantir a integridade da metodologia de aprendizado supervisionado, onde a "verdade fundamental"é estabelecida antes do treinamento do modelo.

A distinção entre as classes foi definida a partir do contexto de captura dos arquivos brutos PCAPNG fornecidos para este estudo:

- **Tráfego Benigno (Classe 0):** Definido a partir de capturas realizadas simulando cenários de atividade convencional de rede. Estes arquivos continham interações legítimas, sem a execução de ferramentas de tunelamento ou scripts de ataque.
- **Tráfego Malicioso/Exfiltração (Classe 1):** Definido a partir de capturas onde houve a execução controlada de ferramentas de exfiltração via DNS. Nestes arquivos, o tráfego DNS apresenta o comportamento específico de um tráfego malicioso.

Além da rotulagem, o código implementou uma camada de filtragem para assegurar a aderência ao escopo da pesquisa. Foram aplicados os seguintes critérios de seleção via biblioteca *Pandas* (TEAM, 2020):

- **Filtragem por Portas:** Foram mantidos apenas os fluxos cujas portas de origem ou destino fossem 53 (Do53) ou 443 (DoH), descartando comunicações de outros serviços de rede irrelevantes para o estudo;
- **Protocolos de Transporte:** O conjunto foi restrito aos protocolos de Controle de Transmissão (TCP) e de Datagrama de Usuário (UDP);
- **Consistência de Dados:** Removeram-se fluxos com contagem de pacotes igual a zero e aplicou-se o tratamento de valores nulos, substituindo-os por 0 para evitar erros matemáticos durante o treinamento dos modelos.

4.4.2 Engenharia de Processamento e Unificação

Dado o volume massivo de dados e a necessidade de preservar a memória RAM, o roteiro utilizou estratégias de limpeza de memória através do módulo *Garbage Collector* (`gc`). Esta abordagem de engenharia de *software* garantiu que cada arquivo fosse processado e pré-analisado individualmente antes da etapa final de concatenação.

A consolidação consistiu na unificação dos múltiplos arquivos intermediários em um único arquivo. Esta unificação foi realizada através da leitura sequencial de fluxos, preservando o cabeçalho original e concatenando todos os registros em uma estrutura tabular única no formato CSV. O resultado final deste processo é um conjunto de dados robusto e padronizado, composto por 20 atributos (19 preditores e 1 variável alvo), pronto para ser utilizado em fases subsequentes de análises.

4.5 Caracterização e Disponibilização do Conjunto de Dados

Esta seção detalha o resultado final do processo metodológico, apresentando as propriedades quantitativas do conjunto de dados gerado e as diretrizes para sua reprodutibilidade.

4.5.1 Características do Conjunto de Dados

O conjunto de dados consolidado apresenta um volume de dados robusto, estruturado com um desbalanceamento de classes planejado para simular o comportamento real de redes de computadores, onde ataques de exfiltração são eventos raros em comparação ao tráfego legítimo. A Tabela 4 resume as propriedades técnicas e a composição do recurso final.

Tabela 4 – Propriedades Técnicas do Conjunto de Dados Gerado

Propriedade	**Valor**
Nº de instâncias	109.359
Nº de atributos	20 (incluindo a *label*)
Nº de instâncias de exfiltração (Classe 1)	9.364
Nº de instâncias benignas (Classe 0)	99.995
Tipo de atributos	Numéricos, Categóricos e Temporais
Tamanho do dataset (formato .csv)	~19.4 MB
Ferramentas utilizadas	NFStream, Python, Pandas

Fonte: Do autor.

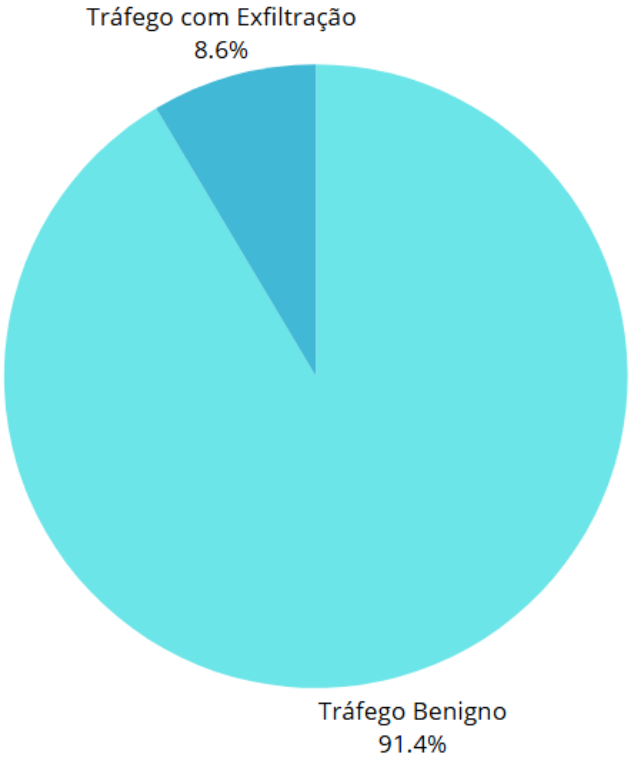

Figura 3 – Gráfico que Representa o Desbalanceamento do Conjunto de Dados.

Fonte: Do Autor

A distribuição das instâncias entre as classes no conjunto de dados final é apresentada na Figura 3. Observa-se que o conjunto de dados é deliberadamente desbalanceado, com a classe benigna representando aproximadamente 91,4% do total, enquanto a classe de exfiltração corresponde a 8,6%.

Ao manter essa proporção de 10:1, o modelo de aprendizagem de máquina é desafiado a minimizar falsos positivos, garantindo que o sistema de detecção seja robusto

o suficiente para operar em ambientes de alta criticidade sem gerar alertas excessivos e irrelevantes.

Para ilustrar a estrutura final do recurso gerado e a organização dos atributos extraídos pelo NFStream, a Tabela 5 apresenta um exemplo de uma instância extraída do conjunto de dados. O registro demonstra um fluxo de tráfego DNS típico, consolidando os metadados de rede, métricas temporais e as características sintáticas calculadas.

Tabela 5 – Exemplo de um registro individual do conjunto de dados.

Atributo	**Valor de Exemplo**
`src_ip`	10.0.30.186
`dst_ip`	8.8.8.8
`src_port`	52577
`dst_port`	53
`protocol`	17
`start_time`	2025-06-14 03:11:14.399
`end_time`	2025-06-14 03:11:14.401
`bidirectional_bytes`	239
`bidirectional_packets`	2
`bidirectional_duration_ms`	2
`src2dst_bytes`	82
`dst2src_bytes`	157
`application_name`	DNS
`application_category_name`	Network
`requested_server_name`	github.io.ec2.internal
`dns_query_length`	22
`dns_query_entropy`	3.7544418457133446
`dns_subdomain_count`	4
`dns_numerical_ratio`	0.0454545454545454
`label`	0

Fonte: Do Autor.

4.5.2 Reprodutibilidade e Acesso

Com o intuito de contribuir para a comunidade de segurança cibernética e garantir a reprodutibilidade científica deste estudo, todo o ecossistema de desenvolvimento foi estruturado para livre acesso, dividido em dois pilares fundamentais:

4.5.2.1 Ambiente e Ferramentas

O processamento dos dados foi realizado em sistema operacional *Windows 11*, utilizando o *Windows Subsystem for Linux*. Essa configuração permitiu a execução nativa de ferramentas baseadas em *kernel Linux*, como o framework **NFStream**, em um ambiente de desenvolvimento integrado via *Visual Studio Code*. A linguagem utilizada foi o *Python*

3.10.12, contando com a biblioteca **Pandas** para a manipulação de estruturas de dados e rotulagem automatizada. A execução dos *scripts* ocorreu em uma estação de trabalho com processador *Intel Core i7* e 32GB de memória RAM, garantindo a eficiência necessária para o processamento dos arquivos PCAPNG.

4.5.2.2 Repositório de Código e Scripts

Os roteiros de automação estão hospedados em um repositório público na plataforma GitHub[1]. O repositório contém quatro *scripts* fundamentais:

1. `extrair_pcapng.py`: Responsável por ler os arquivos brutos e gerar os fluxos bidirecionais via NFStream;
2. `gerador_amostra_benigna.py`: Script de subamostragem desenvolvido para processar arquivos CSV de tráfego benigno fornecidos como entrada. O script extrai uma amostra aleatória de 100.000 instâncias, permitindo consolidar um conjunto de dados final desbalanceado que representa com maior fidelidade os cenários reais.
3. `rotulagem_csv.py`: Responsável por filtrar e realizar o tratamento de valores ausentes, além de rotular o campo *label* como benigno (Calsse 0) ou malicioso (Classe 1);
4. `analise_exploratoria.py`: Responsável pela geração de métricas estatísticas, matrizes de correlação e visualizações gráficas.

O repositório inclui um arquivo `README.md` com instruções de instalação das dependências, permitindo a replicação integral do trabalho proposto.

4.5.2.3 Repositório de Dados

Devido às dimensões do arquivo consolidado e prezando pela rigorosidade científica, o conjunto de dados final foi publicado no repositório *Mendeley Data*. O recurso está registrado sob o DOI **10.17632/v2cy9y58t7.3**[2]. A escolha desta plataforma garante a persistência dos dados e facilita a citação deste trabalho em futuras pesquisas que utilizem a mesma base para modelos de aprendizagem de máquina ou análises estatísticas.

[1] Disponível em: <https://github.com/ViniciusDMRocha/dataset-exfiltracao-DNS>
[2] Link direto: <https://doi.org/10.17632/v2cy9y58t7.2>

5 Análise Exploratória dos Dados

Neste capítulo, o conjunto de dados consolidado é submetido a uma investigação para validar a hipótese de detecção de exfiltração de dados. Antes de tudo, é fundamental compreender a natureza dos atributos selecionados e o potencial analítico que cada um desempenha na distinção entre o tráfego benigno e o malicioso.

Pode-se definir a análise exploratória como o emprego de técnicas estatísticas para examinar e interpretar dados. O objetivo central é proporcionar ao pesquisador uma visão aprofundada sobre as informações coletadas, permitindo que ele identifique comportamentos atípicos ou padrões recorrentes e crie representações gráficas que sintetizem esses achados.

5.1 Potencial Preditivo dos Atributos

O conjunto de dados final é composto por 20 atributos que podem ser segmentados em grupos funcionais. Cada grupo oferece uma perspectiva diferente sobre o comportamento da rede:

5.1.1 Atributos Sintáticos de Domínio

Este grupo contém as variáveis customizadas via Engenharia de Dados e representa o núcleo da detecção baseada em conteúdo:

- **`dns_query_entropy`**: Baseado na Entropia de Shannon (SHANNON, 1948), possui elevado potencial teórico para identificar exfiltração. Espera-se que a aleatoriedade dos caracteres seja superior à de palavras em linguagem natural;
- **`dns_query_length`**: Atributo estratégico para identificar o tunelamento DNS. Para maximizar a eficiência, atacantes utilizam nomes de domínio extensos para carregar o máximo de informação por consulta;
- **`dns_subdomain_count`**: Quantifica o número de níveis ou rótulos contidos no nome de domínio. No contexto de exfiltração, esse valor tende a ser elevado, visto que protocolos de tunelamento costumam segmentar os dados codificados em múltiplos subdomínios para contornar restrições de tamanho de pacotes individuais.;
- **`dns_numerical_ratio`**: Analisa a proporção de números no domínio, sendo um forte indicador de dados binários codificados.

5.1.2 Atributos Volumétricos e de Fluxo

Estes atributos monitoram a intensidade da comunicação:

- **`src2dst_bytes` e `dst2src_bytes`**: Em um ataque de exfiltração, espera-se uma assimetria clara, com volume de saída superior ao de entrada;
- **`bidirectional_duration_ms`**: Importante para diferenciar uma consulta comum de uma sessão de exfiltração persistente.
- **`bidirectional_packets`**: Indica a quantidade total de pacotes trocados entre a origem e o destino durante a vigência do fluxo. Em cenários de exfiltração, esse número tende a ser elevado, pois grandes volumes de dados precisam ser fragmentados em múltiplas mensagens DNS devido às limitações de tamanho impostas pelo protocolo.

5.1.3 Atributos de Contexto e Identificação

Variáveis como `src_ip`, `dst_ip`, `src_port` e `protocol` atuam como metadados. Embora essenciais para auditoria e filtragem de escopo, possuem baixo potencial preditivo para modelos de generalização, pois são identificadores voláteis que não descrevem o comportamento da ameaça em si.

5.2 Resultados da Análise Estatística

Após a definição dos atributos, foi realizada a execução de rotinas computacionais para a extração de métricas estatísticas e geração de gráficos. Esta etapa busca comparar as expectativas teóricas com a realidade dos dados capturados em ambiente real.

5.2.1 Matriz de Correlação de Pearson

A Matriz de Correlação de Pearson foi gerada para mensurar a força da relação linear entre as variáveis independentes e a variável alvo (*label*). Conforme ilustrado na Figura 4, observa-se um comportamento atípico em relação à literatura clássica de cibersegurança.

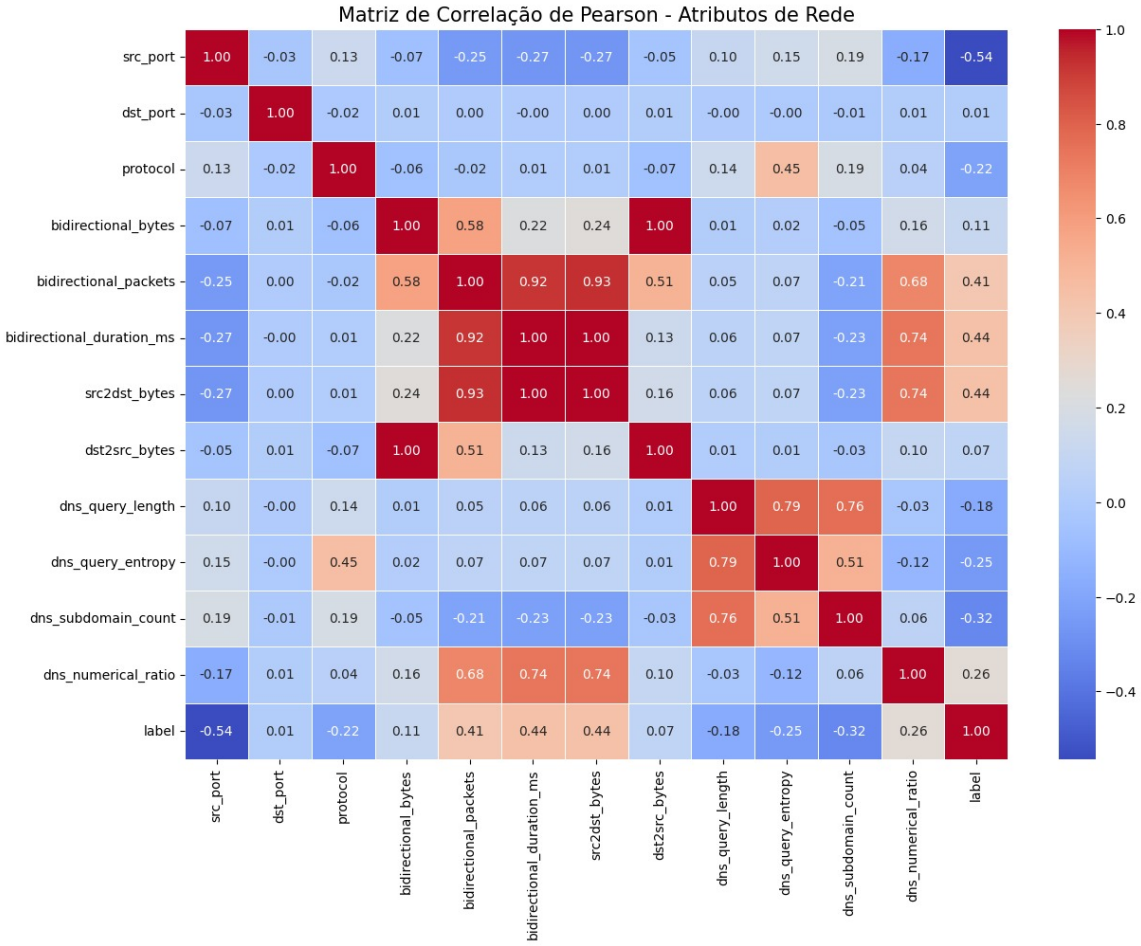

Figura 4 – Matriz de Correlação de Pearson entre atributos e classe alvo.

Fonte: Do Autor

A análise da Matriz de Correlação de Pearson, apresentada na Figura 4, permite identificar quais variáveis possuem maior poder discriminatório em relação à variável alvo (label). Os resultados revelam comportamentos distintos entre as características sintáticas e as comportamentais de fluxo.

5.2.1.1 Atributos com Maior Correlação com a Classe

Diferente do esperado para técnicas tradicionais, os atributos sintáticos como `dns_query_entropy` e `dns_query_length` apresentaram uma correlação negativa com a label (aproximadamente -0,25 e -0,18, respectivamente). É importante ressaltar que a correlação negativa não indica ausência de vínculo, mas sim uma relação inversa: à medida que a complexidade sintática da consulta aumenta, a probabilidade de pertencer à classe maliciosa diminui neste conjunto de dados específico. Este fenômeno revela que, no ambiente de rede capturado, o tráfego benigno manifestou uma complexidade estrutural superior à do tráfego de exfiltração.

Em contrapartida, os atributos de fluxo consolidaram-se como os indicadores mais confiáveis. As variáveis `bidirectional_duration_ms` $(0,44)$, `src2dst_bytes` $(0,44)$ e `bidirectional_packets` $(0,41)$ obtiveram as maiores correlações positivas. Estes valores demonstram que a persistência temporal e o volume de dados transmitidos são características muito mais marcantes de um ataque de exfiltração do que a simples "estranheza"do nome de domínio.

Os resultados apresentados trazem implicações significativas para o desenvolvimento de Sistemas de Detecção de Intrusão:

- **Insuficiência da Análise Estática:** A correlação negativa dos atributos sintáticos sugere que sistemas baseados fortemente em assinaturas ou na "entropia"de nomes de domínio podem gerar um número elevado de falsos positivos em redes modernas. Isso ocorre porque aplicações legítimas (como telemetria de software e serviços de nuvem) utilizam nomes de domínio complexos e ofuscados que mimetizam o comportamento de malwares.
- **Eficiência da Abordagem por Fluxos:** A forte correlação positiva dos atributos temporais e volumétricos valida a escolha metodológica deste trabalho. Detectar a exfiltração torna-se um problema de identificar a persistência e a assimetria do tráfego. Um ataque de exfiltração DNS raramente se denuncia por um único pacote, mas sim pelo comportamento mantido ao longo de um fluxo bidirecional.
- **Redução de Custo Computacional:** Identificar que variáveis como `dst_port` e `src_port` possuem correlação próxima a zero com a classe maliciosa permite a futura simplificação de modelos de Aprendizado de Máquina, focando o processamento apenas nos atributos que realmente impactam a detecção.

5.2.1.2 Análise de Multicolinearidade

Além da relação com a variável alvo, a matriz de correlação (Figura 4) permite identificar atributos altamente correlacionados entre si. Esta análise é fundamental para simplificar o modelo e evitar redundâncias que possam causar sobreajuste ou consumo desnecessário de recursos computacionais.

Observa-se uma correlação positiva quase perfeita ($r \approx 0,98$) entre as variáveis `bidirectional_packets` e `src2dst_bytes`. Este comportamento é justificável pela mecânica do protocolo DNS: uma vez que o tamanho máximo de um pacote UDP é limitado, o envio de um volume maior de dados exige, obrigatoriamente, um aumento proporcional no número de mensagens enviadas. Da mesma forma, a variável `bidirectional_duration_ms` apresenta forte correlação com o volume de pacotes. Isso ocorre porque o processo de exfiltração via DNS é inerentemente lento devido às limitações de banda do protocolo; logo, quanto mais dados são transferidos, maior é o tempo de permanência do fluxo na rede.

A identificação dessas correlações aponta para redundâncias informacionais que trazem implicações diretas para a modelagem:

- **Seleção de Atributos:** Atributos com correlação muito alta (geralmente acima de $0,9$) carregam praticamente a mesma informação estatística. Em cenários de detecção em tempo real, onde a velocidade de processamento é crítica, pode-se optar por manter apenas um deles (como o `src2dst_bytes`) para representar a intensidade do tráfego, descartando o `bidirectional_packets` sem perda significativa de acurácia.
- **Estabilidade do Modelo:** A presença de variáveis redundantes pode inflar artificialmente a importância de certos comportamentos em algoritmos de Aprendizado de Máquina. Ao identificar que a "volumetria"e a "frequência"são faces da mesma moeda no tunelamento DNS, o pesquisador pode ajustar o peso dessas variáveis para garantir que o modelo seja resiliente e não dependa de uma única métrica redundante.

5.2.1.3 Análise da Matriz de Correlação sem redundância

Após a análise exploratória inicial da matriz de correlção, foi realizado um experimento adicional no qual foram removidas as variáveis altamente correlacionadas entre si. Isso pode acabar inflacionando artificialmente a importância de certos padrões e aumentar desnecessariamente o custo de processamento em sistemas de detecção em tempo real. Para mitigar esse problema, foi gerada uma nova matriz de correlação filtrada, utilizando como critério de descarte o limiar de correlação de Pearson de 0,90. O processo consistiu na aplicação de uma máscara booleana triangular superior sobre a matriz original, permitindo identificar pares de variáveis que compartilham quase a mesma informação estatística.

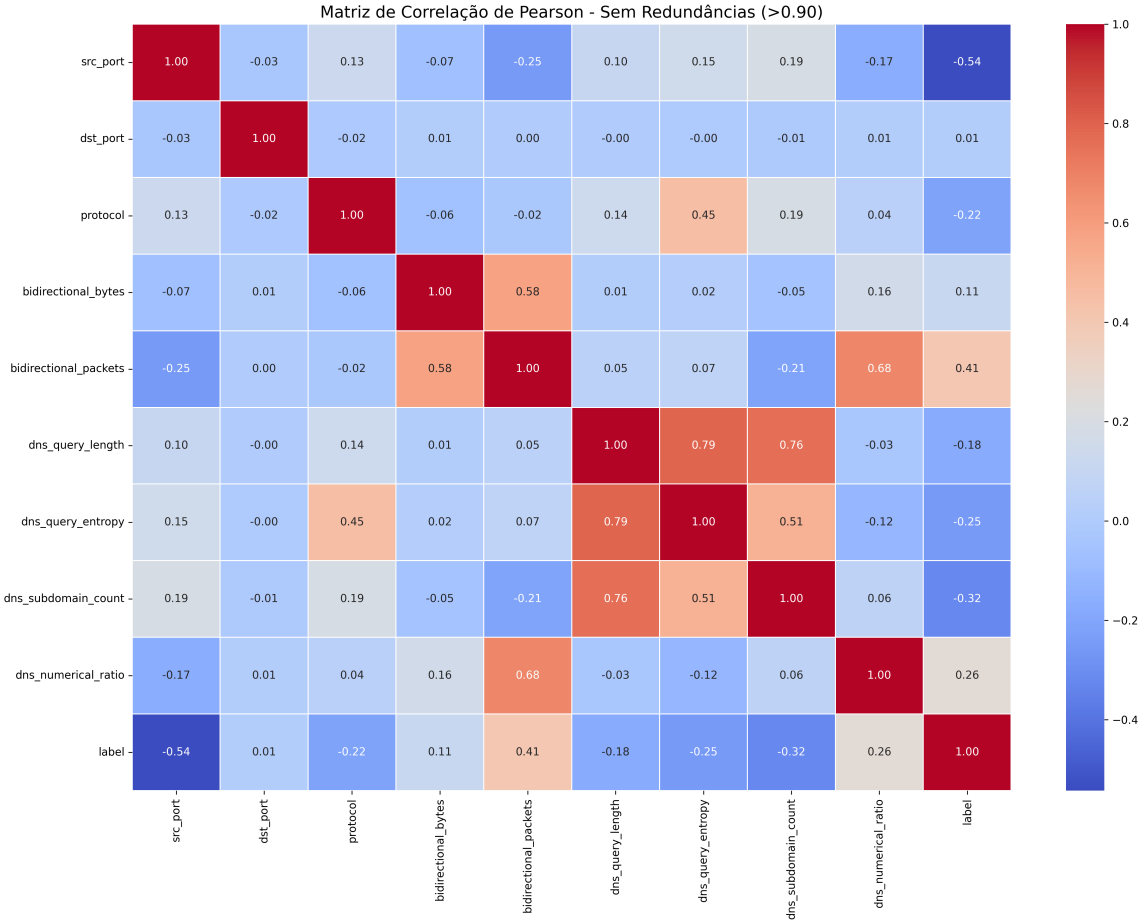

Figura 5 – Matriz de Correlação de Pearson - Sem Redundâncias.

Fonte: Do Autor

Conforme ilustrado na Figura 5, observa-se que a aplicação desse filtro resultou na exclusão de atributos como `src2dst_bytes` e `bidirectional_duration_ms`, que apresentaram colinearidade perfeita (1.00) ou superior a 0.92 com o atributo `bidirectional_packets`. Do ponto de vista técnico, essa redundância é justificada pela própria natureza do protocolo DNS sobre UDP.

Dessa forma, a nova matriz simplificada mantém a integridade da relação com a variável alvo e garante que o conjunto de dados seja mais robusto contra o sobreajuste. Essa redução de dimensionalidade prova que a detecção de exfiltração via DNS pode ser realizada de forma eficiente através de um núcleo reduzido de atributos comportamentais, facilitando a implementação de modelos de inteligência artificial mais leves e rápidos em infraestruturas de rede de alta performance.

5.2.2 Método de Tukey

A aplicação do método de Tukey, através de diagramas de caixa (*boxplots*), permitiu visualizar a dispersão de algumas das métricas e a presença de *outliers*. As figuras abaixo detalham essa distribuição para os quatro principais atributos sintáticos e os três principais atributos com maior correlação.

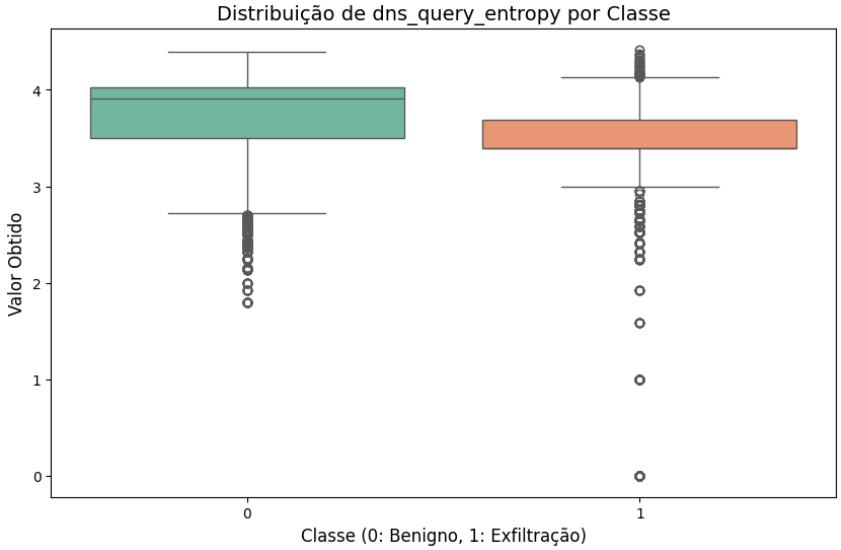

Figura 6 – Distribuição da Entropia de Shannon.

Fonte: Do Autor

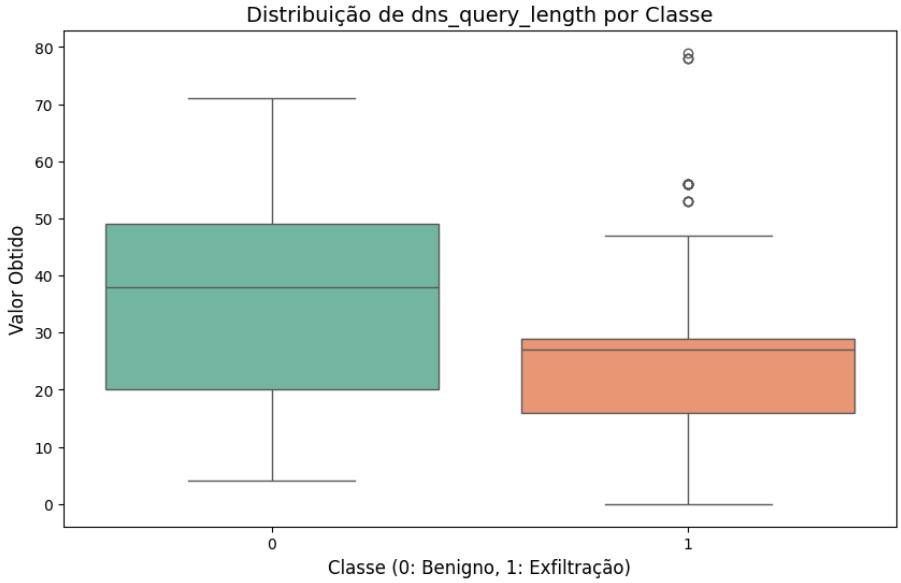

Figura 7 – Distribuição do Comprimento da Consulta.

Fonte: Do Autor

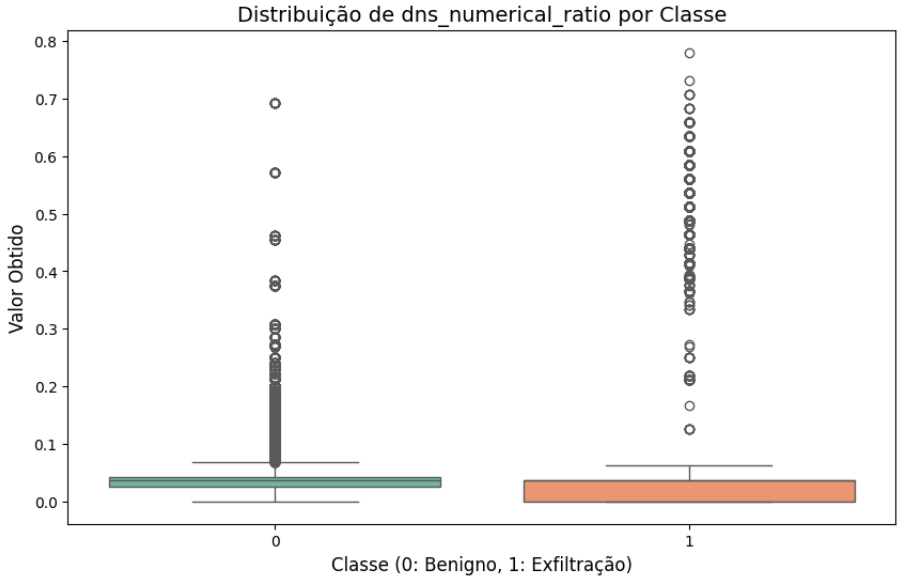

Figura 8 – Proporção Numérica nos Domínios.

Fonte: Do Autor

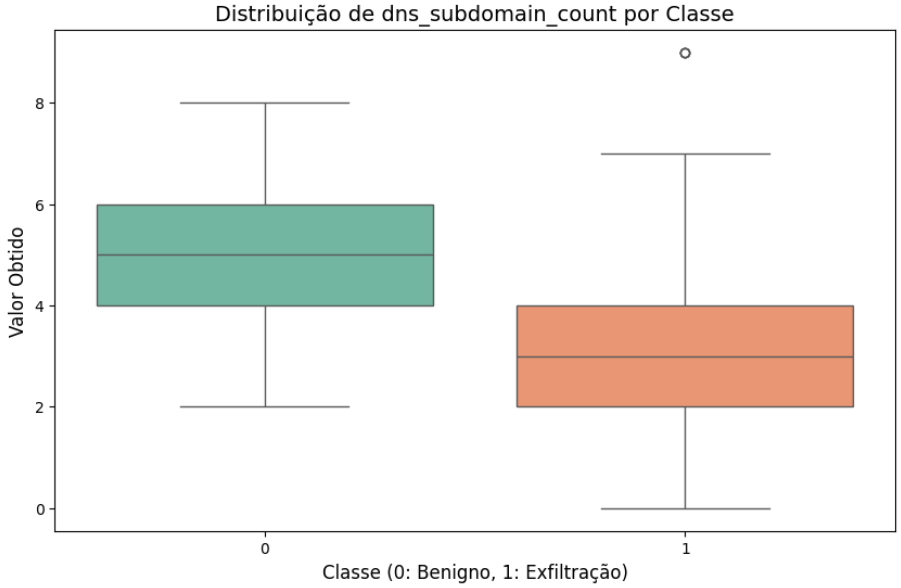

Figura 9 – Quantidade de Subdomínios.

Fonte: Do Autor

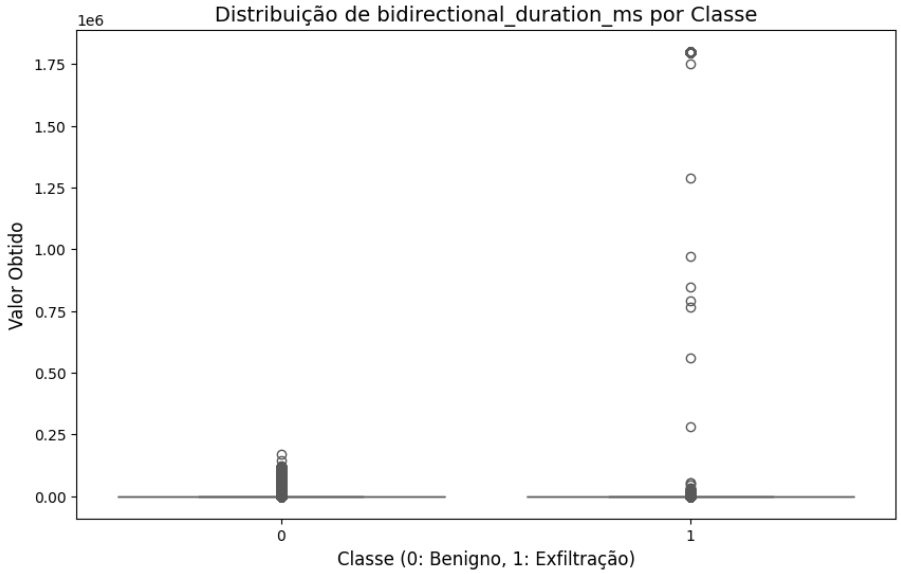

Figura 10 – Duração Bidirecional do Fluxo.

Fonte: Do Autor

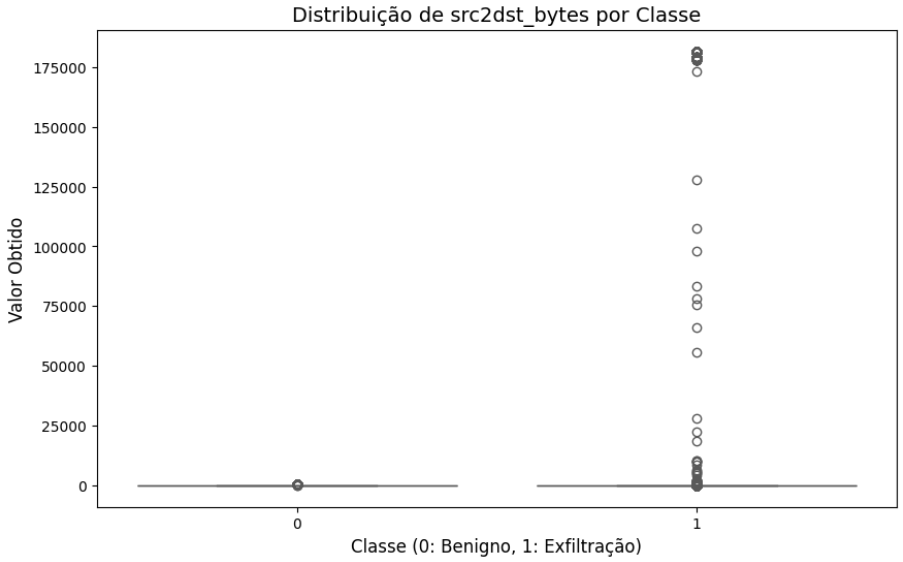

Figura 11 – Volume de Bytes Enviados (Origem para Destino).

Fonte: Do Autor

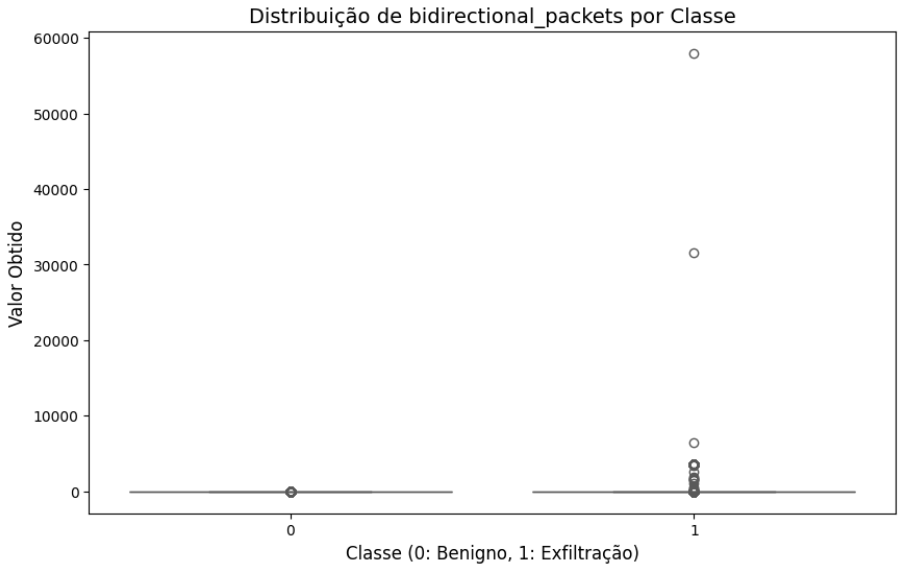

Figura 12 – Contagem Bidirecional de Pacotes.

Fonte: Do Autor

Os perfis de distribuição dos atributos sintáticos, apresentados nas Figuras 6 e 7, revelam uma sobreposição entre as classes. Na classe benigna (0), nota-se uma densidade de pontos discrepantes acima do limite superior. Isso confirma que o tráfego legítimo moderno, influenciado por serviços de telemetria e nomes de domínio ofuscados, gera um "ruído"estatístico que mimetiza as características comumente associadas à exfiltração.

Essa tendência de mimetização também é observada nas Figuras 8 e 9. No caso do `dns_subdomain_count`, a classe benigna apresenta uma mediana de rótulos superior à de exfiltração, justificando a correlação negativa identificada na matriz de correlação e reforçando o uso de nomes complexos em serviços legítimos. Já para o `dns_numerical_ratio`, nota-se que, embora a maioria das amostras se concentre em valores baixos, a classe de exfiltração (1) manifesta uma densidade maior de extremos, indicando o uso de caracteres numéricos como vetor para o transporte de dados binários codificados.

Em contraste, os atributos de fluxo evidenciam os pontos nos quais a exfiltração se torna estatisticamente visível. Conforme ilustrado nos gráficos de dispersão e distribuição das Figuras 10, 11 e 12, as variáveis `bidirectional_duration_ms`, `src2dst_bytes` e `bidirectional_packets` apresentam distribuições distintas, nas quais a classe de exfiltração (1) manifesta *outliers* extremos, confirmando a natureza persistente e volumétrica desse tipo de ataque.

5.3 Discussão

Nesta seção, sintetizam-se as interpretações dos resultados obtidos, confrontando a teoria da detecção de intrusão com os dados observados, além de documentar as limitações e contribuições práticas deste trabalho.

5.3.1 Interpretação e Aplicações do Conjunto de Dados

A análise exploratória demonstrou que a exfiltração moderna via DNS não se denuncia por pacotes isolados de alta entropia, mas sim pelo comportamento contínuo do fluxo. A principal aplicação deste dataset reside no treinamento de modelos de Aprendizado de Máquina que utilizam janelas temporais, em vez de inspeção estática de pacotes. O conjunto de dados permite que pesquisadores testem algoritmos de classificação em um cenário de desbalanceamento real (10:1), aproximando os modelos acadêmicos da realidade operacional dos centros de operações de segurança.

5.3.2 Limitações e Vieses

Embora os ataques registrados no dataset tenham sido executados com ferramentas amplamente reconhecidas no cenário de ameaças, é necessário considerar que os dados foram coletados em um ambiente controlado. Apesar do esforço em compor um cenário que inclua comportamentos típicos de redes modernas, como telemetria e navegação, existe uma fronteira natural entre dados laboratoriais e o tráfego real de uma rede em operação contínua.

Dessa forma, não se pode afirmar que o comportamento observado seja identicamente igual ao de uma rede de produção de larga escala. Variáveis como a presença de dispositivos legados, tráfego de redes IoT ou a imprevisibilidade do comportamento humano são fatores que poderiam alterar as correlações identificadas. Portanto, os resultados aqui apresentados devem ser interpretados como uma prova de conceito robusta, cujas premissas de detecção precisariam de validação adicional em ambientes de maior complexidade e heterogeneidade.

5.3.3 Lições Aprendidas

A construção deste recurso revelou que a fase de Engenharia de Atributos é consideravelmente mais crítica para o sucesso da detecção do que o volume bruto de dados coletados. A lição mais relevante foi a percepção da mimetização de tráfego: o tráfego legítimo atual tornou-se tão complexo que regras de segurança simplistas, baseadas apenas em entropia ou tamanho de nomes de domínio, tornaram-se obsoletas e geradoras de falsos positivos.

Nesse cenário, a principal contribuição deste trabalho foi o foco na dimensão temporal do fluxo. Observou-se que a exfiltração de dados é um processo inerentemente persistente. Enquanto consultas DNS comuns são esporádicas e efêmeras, o tunelamento exige uma sucessão coordenada de requisições que se estendem por longos períodos. Atributos como `bidirectional_duration_ms` revelaram-se indicadores muito mais robustos do que os atributos sintáticos, pois, embora o atacante possa manipular a aparência do pacote, ele dificilmente consegue ocultar a persistência temporal necessária para mover grandes volumes de informação.

A utilização do framework NFStream foi fundamental para esta análise, pois permitiu extrair essas métricas de tempo e bidirecionalidade de forma nativa e padronizada. A capacidade de tratar o tráfego como fluxos temporais contínuos, em vez de pacotes isolados, provou ser o caminho mais eficaz para diferenciar o ruído das aplicações modernas de uma tentativa real de exfiltração de dados.

6 Conclusão

Este trabalho cumpriu seu objetivo primordial de projetar, estruturar e validar um conjunto de dados inédito e de alta fidelidade para a detecção de exfiltração via DNS. Esta pesquisa entrega à comunidade de segurança um artefato consolidado com aproximadamente 110.000 registros e 20 atributos..

A principal contribuição deste trabalho reside na metodologia de construção do dataset, que utilizou o framework NFStream para transpor a análise puramente sintática e alcançar a camada de fluxos estatísticos. A validação desse conjunto de dados, realizada por meio de uma rigorosa Análise Exploratória (EDA), confirmou a utilidade prática e a relevância das métricas selecionadas:

1. **Consolidação de Atributos de Fluxo:** O resultado mais expressivo na criação deste conjunto foi a integração bem-sucedida de métricas comportamentais (como `bidirectional_duration_ms` e `src2dst_bytes`). A validação estatística provou que esses atributos são os verdadeiros discriminadores de ataques, garantindo que o dataset entregue tenha alto valor preditivo para modelos de Aprendizado de Máquina.

2. **Representatividade e Fidelidade:** A base de dados foi estruturada mantendo um desbalanceamento de 10:1, simulando com precisão o panorama real de redes de produção. Isso assegura que o conjunto de dados não seja apenas uma amostra de laboratório, mas um recurso capaz de treinar modelos para enfrentar o ruído real da telemetria moderna.

3. **Superação da Análise Estática:** A construção do dataset demonstrou que a "estética"da consulta (como a Entropia) é insuficiente quando isolada. Por isso, a entrega deste conjunto de dados com uma camada de engenharia de atributos avançada representa um avanço metodológico, fornecendo métricas temporais e volumétricas que superam as limitações das detecções baseadas apenas em nomes de domínio.

Em suma, este trabalho entrega um recurso de dados validado e pronto para uso, disponível publicamente no repositório Mendeley Data. O sucesso na criação deste dataset abre caminho para que futuras pesquisas utilizem essa base para o treinamento de classificadores inteligentes, elevando o estado da arte na proteção contra exfiltração de dados no protocolo DNS.

6.1 Trabalhos Futuros

Como desdobramentos desta pesquisa, propõe-se a continuidade do desenvolvimento do conjunto de dados e a exploração de novas frentes de análise. Primeiramente, planeja-se a ampliação do dataset mediante a inclusão de novos cenários de ataque. Um ponto relevante de evolução será a implementação efetiva do suporte ao protocolo DNS over HTTPS (DoH). Embora a inclusão desta variante cifrada tenha sido planejada e testada no escopo deste trabalho, limitações técnicas na captura e na decodificação de metadados específicos impediram sua integração completa na base atual. Dessa forma, a investigação de métodos para contornar tais restrições apresenta-se como um caminho promissor para manter a base atualizada frente às tendências de privacidade.

Adicionalmente, pretende-se realizar a aplicação do conjunto em modelos de detecção baseados em Aprendizado de Máquina e Aprendizado Profundo (*Deep Learning*). O objetivo dessa etapa futura será validar a eficácia dos atributos extraídos via NFStream na classificação em tempo real, testando a resiliência dos modelos frente ao desbalanceamento de classes e ao ruído de telemetria identificado na análise exploratória deste trabalho.

Em suma, este trabalho contribui para a área de segurança de redes ao disponibilizar um conjunto de dados aberto e documentado no *Mendeley Data*, apto para o treinamento de modelos de Inteligência Artificial mais robustos. A pesquisa sugere que a detecção de ameaças via DNS deve evoluir de uma análise estática para uma análise comportamental multidimensional, capaz de filtrar o ruído legítimo das aplicações modernas e isolar a persistência maliciosa de canais encobertos.

Referências

BENEDICTIS, M. D. **NFStream: A Fast, Flexible, and Extensible Network Data Analysis Framework**. 2024. <https://www.nfstream.org/docs/>. Acessado em: 9 de jan. de 2026. Citado 2 vezes nas páginas 17 e 26.

Fortinet. **Data Exfiltration - Cyber Glossary**. 2024. Accessed: 2025-03-24. Disponível em: <https://www.fortinet.com/resources/cyberglossary/data-exfiltration>. Citado na página 14.

HOFFMAN, P.; MCMANUS, P. **DNS Queries over HTTPS (DoH)**. [S.l.], 2018. Disponível em: <https://www.rfc-editor.org/info/rfc8484>. Citado na página 14.

HOFSTEDE, R.; CELEDA, P.; TRAMMELL, B.; DRAGO, I. et al. Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX. **IEEE Communications Surveys Tutorials**, v. 16, n. 4, p. 2036–2062, 2014. Citado na página 17.

IBM. **O que é rotulagem de dados?** 2024. <https://www.ibm.com/br-pt/think/topics/data-labeling>. Acessado em: 13 de jan. de 2026. Citado na página 19.

_____. **O que é um dataset?** 2024. <https://www.ibm.com/br-pt/think/topics/dataset>. Acessado em: 15 de jan. de 2026. Citado na página 18.

IBM. **Intrusion Detection System (IDS)**. 2025. Acesso em: 24 mar. 2025. Disponível em: <https://www.ibm.com/br-pt/topics/intrusion-detection-system>. Citado na página 15.

IBM Security. **Cost of a Data Breach Report 2023**. 2023. Accessed: 2024-03-06. Disponível em: <https://www.ibm.com/reports/data-breach>. Citado na página 10.

LASHKARI, A. H.; SALEH, S.; ELSAYED, M. S.; ZOHOUN, G. Detecting dns exfiltration traffic using machine learning. **Canadian Institute for Cybersecurity (CIC)**, 2021. Disponível em: CIC-Bell-DNS-EXF-2021 Dataset. Disponível em: <https://www.unb.ca/cic/datasets/dns-exf-2021.html>. Citado 2 vezes nas páginas 10 e 21.

MAHDAVIFAR, S.; SALEM, A.; VICTOR, P.; RAZAVI, A.; GARZON, M.; HELLBERG, N.; LASHKARI, A. H. Lightweight hybrid detection of data exfiltration using dns based on machine learning. In: **Proceedings of the 2021 11th International Conference on Communication and Network Security (ICCNS '21)**. [S.l.: s.n.], 2021. p. 80–86. Citado 4 vezes nas páginas 10, 13, 15 e 21.

MOCKAPETRIS, P. **Domain names - concepts and facilities**. [S.l.], 1987. Disponível em: <https://www.rfc-editor.org/info/rfc1034>. Citado na página 13.

_____. **Domain names - implementation and specification**. [S.l.], 1987. Disponível em: <https://www.rfc-editor.org/info/rfc1035>. Citado na página 13.

MONTAZERI-GHAFOORI, M.; BEIGI, M.; SAMANEH, M.; LASHKARI, A. H. **CIRA-CIC-DoHBrw-2020 Dataset**. Canadian Institute for Cybersecurity (CIC), 2020. Acessado em: 7 de mar. de 2026. Disponível em: <https://www.unb.ca/cic/datasets/dohbrw-2020.html>. Citado na página 10.

SHAFRANOVICH, Y. **Common Format and MIME Type for Comma-Separated Values (CSV) Files**. IETF, 2005. RFC 4180. Acessado em: 13 de jan. de 2026. Disponível em: <https://datatracker.ietf.org/doc/html/rfc4180>. Citado na página 18.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948. Citado na página 36.

TEAM, T. pandas development. **pandas-dev/pandas: Pandas**. Zenodo, 2020. Disponível em: <https://doi.org/10.5281/zenodo.3509134>. Citado na página 32.

TUEXEN, M.; RISSO, F.; BONGERTZ, J.; COMBS, G.; HJELMVIK, E. **PCAP Next Generation (pcapng) File Format**. 2024. <https://pcapng.com>. Acessado em: 8 de jan. de 2026. Citado na página 16.

TUKEY, J. W. **Exploratory Data Analysis**. Reading, MA: Addison-Wesley, 1977. Citado na página 19.

ZHAN, M.; LI, Y.; YU, G.; LI, B.; WANG, W. Detecting dns over https based data exfiltration. **Computer Networks**, v. 209, p. 108919, 2022. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128622001104>. Citado 3 vezes nas páginas 10, 13 e 21.

ŽIžA, K.; TADIć, P.; VULETIć, P. Dns exfiltration detection in the presence of adversarial attacks and modified exfiltrator behaviour. **International Journal of Information Security**, v. 22, p. 1865–1880, 2023. Disponível em: <https://doi.org/10.1007/s10207-023-00723-w>. Citado na página 10.