Avaliação de estratégias para o aperfeiçoamento da detecção de anomalias no tráfego DNS

Mayara R. E. Santos[1], **Raquel S. M. Santos**[1], **Italo V. S. Brito**[1], **Leobino N. Sampaio**[1]

[1] Instituto de Computação – Universidade Federal da Bahia (UFBA)
Salvador, BA – Brasil

{mayara.rodrigues, raquelsms, italovalcy, leobino}@ufba.br

Abstract. *The increasing complexity of cyber threats makes the detection of anomalies in DNS traffic crucial for ensuring network security. Although several studies have been conducted to explore this process, the presence of false positives remains a significant challenge in the analysis. This work aims to validate DNS anomaly detection techniques with low false positive rates, using data from threat intelligence sources and domain analysis techniques applied in similar studies. The validation was performed using datasets publicly available and widely used. Despite the fact that the results did not meet the expectations, they indicate the need to evaluate the data using other techniques, and these aspects can be explored in future works.*

Resumo. *A crescente complexidade das ameaças cibernéticas torna a detecção de anomalias no tráfego DNS essencial para a segurança da rede. Apesar de diversos estudos explorando o processo, a presença de falsos positivos continua sendo um desafio nas análises. Este trabalho buscou validar estratégias de detecção de anomalias DNS com baixas taxas de falsos positivos, utilizando informações de inteligência de ameaça e técnicas de análise de domínios aplicadas em trabalhos similares. A validação foi realizada com dados de fontes confiáveis, utilizados por outros autores. Os resultados, embora não tenham atendido às expectativas, indicam a necessidade de validar os dados e aplicar outras técnicas, aspectos que podem ser explorados em trabalhos futuros.*

1. Introdução

O DNS (*Domain Name System*) é um protocolo de comunicação fundamental para o funcionamento da internet e amplamente utilizado no processo de resolução de nomes de domínios para endereços IP (*Internet Protocol address*). Devido à ampla utilização desse protocolo, o sistema DNS é um alvo comum de ataques cibernéticos. *Distributed Denial of Service*, *Cache Poisoning* e consultas anormais são algumas das anomalias que podem ocorrer no tráfego DNS. Existem medidas preventivas que visam identificar esses comportamentos atípicos, uma delas é o monitoramento do tráfego por meio de ferramentas especializadas para detecção de irregularidades ou padrões incomuns na rede.

Embora existam técnicas e ferramentas eficazes que auxiliam no processo de detecção, um dos principais desafios relacionados à identificação de anomalias no tráfego DNS é a probabilidade de que ocorra um falso positivo nas consultas. A redução de falsos positivos no monitoramento de segurança do DNS é recorrente alvo de investigação na literatura, isso porque existem inúmeros atacantes que utilizam técnicas para fazer com

que suas atividades pareçam legítimas, fazendo com que uma parte do tráfego benigno seja avaliado como malicioso. Na tentativa de oportunizar o tratamento de falsos positivos, foram consideradas as técnicas de entropia dos domínios acessados e a análise de bi-gramas. Essas técnicas foram aplicadas em conjunto para possibilitar uma análise mais detalhada do tráfego DNS detectado pelo Zeek. A combinação dessas estratégias, aliada ao uso de dados fornecidos por uma base de inteligência de ameaças, pode contribuir para a redução da ocorrência de falsos positivos na detecção de tráfego DNS anômalo, o que favoreceu a otimização do processo de detecção de anomalias, por meio de diferentes parâmetros de detecção aplicados às requisições DNS.

Para isso, foi realizada a análise da reputação de listas de domínios potencialmente maliciosos e a validação da legitimidade das consultas DNS, utilizando dados de uma base de inteligência de ameaças e outras fontes complementares que fornecem parâmetros para a análise de tráfego benigno e maligno. O objetivo foi aprimorar os métodos de detecção de irregularidades no fluxo de consultas DNS. Assim, a resolução do problema proposto busca melhorar as técnicas conhecidas de detecção de anomalias no tráfego DNS, tornando essa detecção mais precisa e eficiente, além de otimizar recursos computacionais essenciais para garantir a segurança nas trocas de pacotes DNS. As heurísticas apresentadas nesta pesquisa também contribuirão para o aperfeiçoamento de métodos utilizados na análise comportamental do tráfego em trabalhos futuros. O restante deste artigo está organizado da seguinte maneira: a seção 2 apresenta uma revisão de trabalhos que estão relacionados ao problema proposto nesta pesquisa; a seção 3 demonstra uma análise detalhada da metodologia utilizada; a seção 4 evidencia um levantamento e discussão dos resultados alcançados com os experimentos realizados; e a seção 5 apresenta as conclusões e projetos futuros.

2. Trabalhos Relacionados

A detecção de anomalias no tráfego DNS, especialmente em relação a túneis DNS, possui uma ampla documentação. [Wang et al. 2021] dissertam variadas técnicas de detecção de tunelamento DNS, que têm como uma das principais bases a análise de características de *payload*, como a entropia de caracteres, utilizada neste trabalho. Esta pode ser definida como a aleatoriedade de uma palavra, e domínios relacionados a túneis DNS costumam apresentar alta entropia devido a combinações complexas e ilegíveis de caracteres, que visam ocultar os dados transmitidos. Por outro lado, [Sharma and Swarnkar 2023] apresentam um *framework* para detectar túneis DNS do tipo *zero-day* usando análises de n-gramas. A validação foi realizada com tráfego DNS de bases de dados públicas, demonstrando boa eficiência e embasando o uso de bigramas neste trabalho. Já [AlSabeh et al. 2024] discutem a detecção de domínios gerados aleatoriamente, típicos de túneis e anomalias de DNS, utilizando switches P4, que oferecem flexibilidade e alta capacidade de processamento, junto a modelos de inteligência artificial para identificar anormalidades em pacotes e tráfego.

Os falsos positivos são uma preocupação na detecção de anomalias no tráfego DNS, pois atacantes usam técnicas para disfarçar suas atividades. Nesse contexto, a solução de [Sharma and Swarnkar 2023] apresentou baixa taxa de falsos positivos, devido à sua precisão em distinguir tipos de domínios e à eficiência dos switches P4, que minimizam a interferência na qualidade do tráfego. Por outro lado, [AlSabeh et al. 2024] apresentam um *framework* eficiente e com baixas taxas de falsos positivos para a detecção

de domínios gerados aleatoriamente a partir da utilização de modelos de IA que analisam características diversas de tráfego e pacotes. Essa abordagem valida a aplicação de diversos parâmetros para otimizar a detecção de anomalias, sendo utilizada neste trabalho para integrar diferentes bases de informações e garantir uma detecção eficiente de anomalias DNS. Adotando outra estratégia, [Ishikura et al. 2021] exploram o fato de consultas relacionadas a tunelamento DNS gerarem respostas do tipo *cache missing*, indicando que o domínio não foi acessado anteriormente, o que se deve à geração de domínios no tunelamento. Eles utilizaram detecção por regras e por aprendizado de máquina, alcançando boa eficiência e baixos falsos positivos. Tais trabalhos são alguns dos quais ilustram a relevância da temática, além de promoverem embasamento teórico para a utilização de parâmetros aplicados em testes relativos a este trabalho.

3. Metodologia

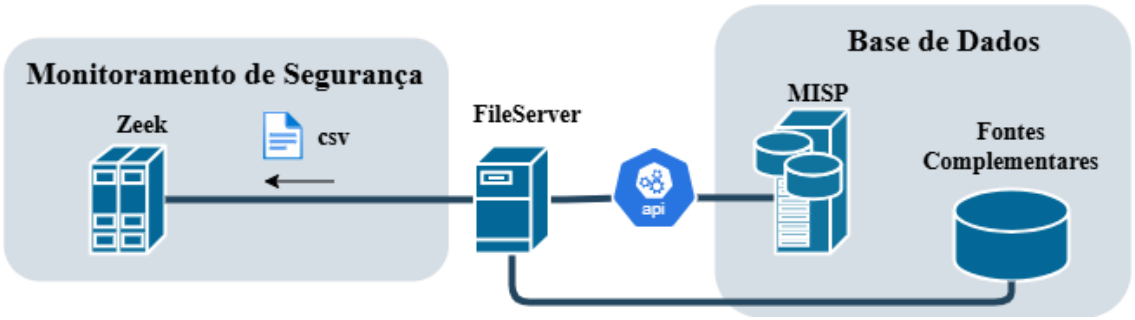

Figura 1. ***Fluxo de Execução***

Visando promover a detecção de anomalias DNS com baixas taxas de falsos positivos, a metodologia dessa pesquisa exploratória envolveu a criação de uma estrutura computacional composta por um módulo de detecção, representado pela ferramenta de análise de tráfego Zeek, e um módulo de inteligência de ameaça, representado pelo *Malware Information Sharing Platform* (MISP). Essa abordagem objetiva aprimorar a detecção de anomalias na camada de tradução de domínios, integrando a análise das características do tráfego com informações de inteligência de ameaça, aumentando assim as chances de uma identificação eficaz de comportamentos anômalos. Foi aplicado o método qualitativo para definição das técnicas de análise das consultas DNS, de forma que foram empregues parâmetros previamente utilizados em trabalhos correlatos para detecção de anomalias no tráfego DNS, como descrito na seção de Trabalhos Relacionados, além da utilização de dados da base de inteligência de ameaça utilizada nesta pesquisa. Nesse contexto, foi utilizada a seguinte sequência de análise:

1. Verificação da presença de domínios analisados na base de dados do MISP;
2. Entropia dos domínios acessados;
3. Análise de bigramas, por meio do cálculo de pontuação;

Sendo que, neste contexto, a pontuação é a soma das frequências dos bigramas de um domínio, e será explicada mais detalhadamente na subseção de Testes Realizados. Visando atestar a eficácia da estrutura desenvolvida, a métrica de avaliação utilizada foi a verificação das taxas de falsos positivos e positivos negativos. Tal escolha foi feita levando em consideração o fato de que métricas similares são frequentemente usadas em trabalhos similares, pela sua capacidade em promover distinção entre dados correta e incorretamente avaliados.

3.1. Processo de Integração

Para integrar o módulo de monitoramento de segurança com a base de dados, foram criados *scripts* para automatizar o processo de exportação e importação de dados, assegurando a integridade das informações coletadas. Nesse sentido, o Zeek é um analisador de tráfego *open source* amplamente utilizado para monitoramento de segurança, o qual permite a análise personalizada do fluxo de pacotes, característica que o levou a ser utilizado no contexto desse trabalho e que será explorada mais adiante.

Diante da crescente complexidade das ameaças cibernéticas, é crucial utilizar uma base de conhecimento aprofundada para classificá-las. Sistemas como o *Collective Intelligence Framework* (CIF), *TheHive* e MISP oferecem recursos para detecção, análise e compartilhamento de ameaças. A arquitetura do MISP inclui uma interface web, um banco de dados centralizado e recursos de importação e exportação que facilitam a integração com ferramentas de segurança cibernética. A escolha dessa plataforma se justifica por suas características técnicas, como o mecanismo de correlação automática, que relaciona eventos e dados de fontes externas, assim como o suporte a importação e exportação de dados em diversos formatos, facilitando, dessa forma, a integração com de sistemas de detecção de intrusão em rede (NIDS), como o Suricata, Snort e Zeek. Ao fornecer dados do MISP ao módulo de monitoramento de segurança, portanto, obtemos parâmetros mais relevantes para as análises efetuadas nesta pesquisa.

Os dados coletados incluíam domínios e IPs com um nível significativo de comprometimento. O dataset utilizado na pesquisa está disponível em: https://github.com/hackinsdn/sbrc2025-wtg. A base de inteligência de ameaça foi fundamental para a filtragem de falsos positivos, utilizando a instância MISP, que recebe informações validadas por organizações reconhecidas. Para a exportação dos IPs, foi aplicada a taxonomia *Admiralty Scale* para avaliar a confiabilidade das fontes e a credibilidade das informações. Na etapa de exportação dos dados da base de inteligência de ameaças, foram utilizados scripts para buscar atributos do tipo IP (origem ou destino) e domínios classificados com um grau significativo de comprometimento. A biblioteca PyMISP possibilitou o acesso à API REST do MISP, permitindo a geração de um arquivo CSV com uma lista de domínios comprometidos publicados em eventos de organizações previamente selecionadas. Além disso, a implementação de um *shell script* possibilitou a coleta de uma lista de IPs comprometidos. Para essa coleta, foram aplicadas tags associadas à taxonomia MISP *Admiralty-Scale* como parâmetro de busca, resultando em IPs contidos em eventos classificados pela comunidade com um nível considerável de confiabilidade da fonte e credibilidade da informação.

Mediante a aplicação dos analisadores do Zeek, módulos responsáveis por examinar o tráfego de protocolos específicos e gerar logs sobre a atividade de rede, foi possível desenvolver um sistema de detecção capaz de realizar análises personalizadas por meio de *scripts* customizados. Esses *scripts* foram fundamentais para a concretização dos objetivos estabelecidos neste estudo, uma vez que, por meio deles, foi possível explorar funcionalidades específicas dos *frameworks* do Zeek voltadas para esse tipo de implementação, como a análise detalhada de dados de rede. Importa destacar, ainda, que a unidade de monitoramento de segurança opera em dois modos distintos: análise *offline* de arquivos de tráfego e análise em tempo real de uma interface de rede específica.

Dessa forma, os eventos do sistema de monitoramento são acionados automati-

camente ao serem detectados padrões específicos no tráfego, podendo ser referenciados nos *scripts* para permitir a execução de comandos e a geração de logs conforme o comportamento identificado. Tendo isto em conta, após a exportação dos dados do MISP, a etapa seguinte foi a criação de um *script* no Zeek que foi desenvolvido para possibilitar o processamento e análise do arquivo gerado na exportação. Para tal, utilizou-se o *Input Framework*, que permite a extração desejada no escopo de *scripts* para tabelas cujas informações podem ser utilizadas no processamento de dados.

3.2. Experimentos Realizados

Nos testes realizados para validar a estrutura criada, foram utilizados os seguintes dados:

1. **Domínios populares da Alexa:** Alexa é um site que reúne domínios populares seus dados já foram usados por autores como [Wang et al. 2021] e estão disponíveis em [Ghodke 2016];
2. **Arquivos de tráfego benigno:** Arquivos produzidos por parceria entre o *Canadian Institute for Cybersecurity* e a *Bell Canada* [Mahdavifar et al. 2021];
3. **Arquivos de tráfego malicioso:** Arquivos relacionados a malwares baseados em DGAs de diferentes famílias, originados das plataformas VirusTotal, VirusShare, Traige e Malpedia [Alsabeh 2021]. Os foram coletados por [AlSabeh et al. 2023].

Sendo que a primeira fonte foi utilizada para extrair bigramas e obter dados que refletissem comportamento benigno. Por outro lado, as outras duas foram utilizadas para a realização de testes para validar a estrutura criada, pelo fato de conterem informações de tráfego que refletem comportamentos de usuários benignos e maliciosos. Nesse sentido, uma das dificuldades encontradas nos experimentos realizados foi o excesso de requisições DNS repetidas no *dataset* de tráfego malicioso. Isso se deve às características do tráfego DNS, que pode incluir múltiplas requisições para o mesmo domínio, porém dificultou os testes pela extensão dos arquivos. Ao analisar a base de dados, foram identificadas cerca de 3300000 requisições para pouco mais de 97000 domínios únicos, que foram extraídos para facilitar os testes. Em um ambiente isolado, foram feitas requisições para esses domínios, gerando um arquivo PCAP com informações de tráfego. Nas análises realizadas, foi essencial o cálculo de entropia dos domínios, utilizando a seguinte fórmula [Kovar 2015]:

$$H = -\sum p(x) \log p(x)$$

Onde H representa a entropia e $p(x)$ representa a frequência de um caractere em um domínio, i.e., a sua quantidade de ocorrências dividida pela quantidade total de caracteres do domínio, e o log calculado é na base 2. Em relação ao processo de cálculo de pontuação de um domínio, foram extraídos os bigramas dos domínios de [Ghodke 2016] e calculadas suas frequências, que correspondem à razão entre suas ocorrências e o total de bigramas. Ao analisar um domínio com o Zeek, seus bigramas são extraídos e, se estiverem entre os já obtidos, a frequência do bigrama é somada à pontuação. As frequências foram multiplicadas por 100000 para reduzir casas decimais e facilitar a manipulação dos dados pelo Zeek. Adicionalmente, foram calculados os valores de entropia e pontuação média para os dois grupos de domínios, objetivando encontrar valores para diferenciá-los. Contrariamente ao esperado, os valores de entropia foram extremamente próximos, como

se pode ver na Tabela 1, indicando similaridade entre os grupos e capacidade de disfarçar tráfego malicioso. Além disso, levou à adoção de estratégias de detecção que considerassem as duas variáveis, com o objetivo de promover uma análise mais abrangente, descritas a seguir:

1. **Priorização da Pontuação:** Nesta abordagem, dividimos os valores de pontuação em subgrupos de forma que, se um domínio tem uma pontuação que o coloca em um determinado grupo, avaliamos a entropia desse domínio para identificar seu tipo. Os subgrupos variaram em tamanho, com valores entre 2500 e 20000, e cada grupo continha 500 unidades. Em outras palavras, foi analisada como a pontuação de um domínio pertencer a um grupo específico pode ajudar a entender melhor suas características.
2. **Priorização da Entropia:** Estratégia análoga à anterior, mas com inversão da ordem das variáveis. Foram formados subgrupos que variaram entre 2 e 4, com cada grupo abrangendo 0.25 unidades. Isso significa que estamos focando na entropia para classificar os domínios, em vez da pontuação.

Em ambas as estratégias, os subgrupos foram definidos de maneira a incluir a maior parte dos domínios, de forma a tornar as análises eficientes. Dessa forma, os subgrupos foram definidos a partir de valores que englobassem a maior parte dos dados.

	Domínios benignos	Domínios maliciosos
Entropia média	2,804702132	12237,87829
Pontuação média	2,799544774	7179,426385

Tabela 1. Comparação de valores médios de entropia e pontuação.

4. Resultados e Discussões

Após a realização de testes envolvendo as estratégias de priorização de pontuação e de entropia, a partir da análise *offline* de pacotes com informação de tráfego, foram obtidos os resultados apresentados na Tabela 2, os quais não correspondem fielmente às expectativas de eficiência na análise. Uma das possíveis justificativas pode ser a discrepância dos valores, que é maior no caso de priorização de pontuação, como se pode observar na Figura 2. Ou seja, em uma determinada faixa de valores de entropia, a diferença na pontuação média entre domínios benignos e maliciosos é maior do que a diferença nos valores de entropia dentro de uma faixa específica de pontuação. Tal contexto pode ser explicado pela similaridade entre os grupos de domínios, observada nos valores médios de entropia, que provocam proximidade nos valores de pontuação independente da faixa de entropia. Ou seja, pelo fato de os domínios serem parecidos e, consequentemente terem bigramas parecidos, os valores de pontuações foram mais próximos. Por outro lado, a maior variação na pontuação para um determinado conjunto de valores de entropia, principalmente nas faixas mais altas de entropia pode ser justificada por domínios maliciosos com características divergentes dos domínios benignos. Ou seja, apesar de os valores de entropia serem semelhantes devido à estrutura do domínio, uma pontuação menor pode indicar que os bigramas não são comuns no tráfego legítimo. Por outro lado, as faixas de entropia em que os domínios maliciosos tem maior pontuação podem sugerir

manipulação deles visando legitimidade, utilizando palavras comuns. Domínios legítimos podem incluir tanto bigramas comuns quanto não comuns, enquanto domínios maliciosos podem conter apenas bigramas comuns, o que eleva artificialmente a pontuação e ajuda a diferenciar eles do tráfego benigno.

De forma geral, apesar de haver uma certa diferenciação dos dados, as taxas não se mostraram satisfatórias como planejado, indicando necessidade de adoção de novas estratégias, além de revisão das bases de dados, visando avaliar se a qualidade dos mesmos interferiu nos resultados. A não satisfação das expectativas se relaciona às alta eficiência de trabalhos relacionados, os quais têm sucesso em promover distinção precisa dos tipos de domínios. Adicionalmente, o arquivo contendo dados da base MISP continham pouco mais de 18000 domínios, um número relativamente baixo comparado ao de domínios presentes nos arquivos de tráfego usado nos testes, apesar de conter todos os domínios presentes na plataforma de inteligência de ameaça. Por conta dessa limitação numérica não houve correlação entre os domínios do MISP e os domínios dos arquivos de tráfego. Outro fator que pode ter influenciado nesse contexto é o fato de os domínios do MISP serem mais específicos, relacionados à seções específicas, enquanto os arquivos de tráfego possuem domínios mais ligados às páginas principais.

	Falsos positivos	Verdadeiros positivos
Priorização de pontuação	34,6%	49.9%
Priorização de entropia	31%	69.7%

Tabela 2. Apresentação dos resultados obtidos

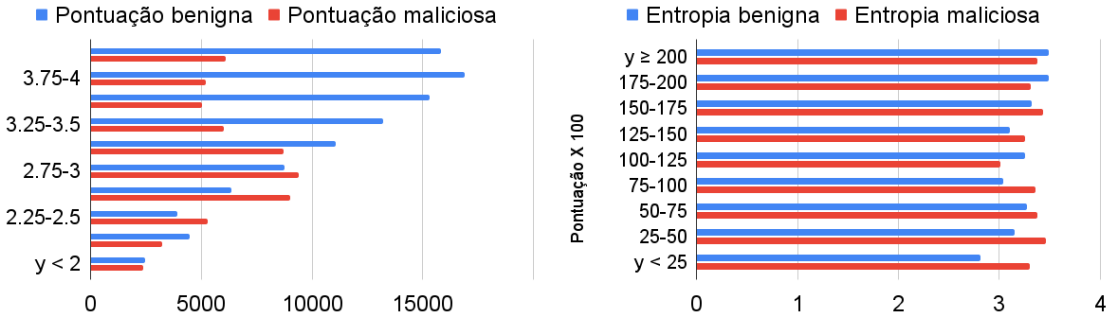

Figura 2. Classificação dos domínios de acordo com pontuação (A) e entropia (B).

5. Conclusões e Trabalhos Futuros

De forma geral, pode-se dizer que os resultados não corresponderam ao esperado por fatores diversos, como qualidade e limitações das bases de dados, bem como limitações ligadas às estratégias de detecção. Além disso, a correlação entre domínios maliciosos e dados obtidos do MISP, apesar de não observada no contexto desse trabalho, pode ser alcançada em trabalhos futuros com a utilização de outras bases de dados. É importante ressaltar que o MISP possui dados obtidos a partir de fontes confiáveis e que representam tráfego malicioso, sendo que a falta de correlação pode ser resultado das características

dos dados maliciosos utilizados. Como trabalhos futuros, planeja-se utilizar o processo de mineração de dados para validar os resultados obtidos nesta pesquisa, mediante a utilização de um modelo de classificação que possa auxiliar na identificação e redução de falsos positivos. Devido às características e complexidades dos dados analisados neste trabalho, para a construção do modelo de classificação pretende-se testar diferentes algoritmos de aprendizado supervisionado como *Rondom Forest*, *Support Vector Machine* (SVM) e o *Decision Tree* a fim de comparar os resultados e assim garantir uma maior capacidade de generalização, precisão e interpretação, garantindo, dessa forma, *insights* relevantes para o aprimoramento do desempenho geral do processo de classificação dos dados. Adicionalmente, considera-se a utilização de outras fontes de dados confiáveis e validação utilizando dados reais de produção.

Agradecimentos

Os autores agradecem o apoio da Rede Nacional de Ensino e Pesquisa (RNP), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

Referências

Alsabeh, A. (2021). P4-dga-multiclass. Github Repository. Acessed: March 19, 2025.

AlSabeh, A., Friday, K., Crichigno, J., and Bou-Harb, E. (2023). Effective dga family classification using a hybrid shallow and deep packet inspection technique on p4 programmable switches. In *ICC 2023 - IEEE International Conference on Communications*, pages 3781–3786.

AlSabeh, A., Friday, K., Kfoury, E., Crichigno, J., and Bou-Harb, E. (2024). On dga detection and classification using p4 programmable switches. *Computers Security*, 145:104007.

Ghodke, S. (2016). Alexa top 1 million sites: Rankings of the top 1 million websites in the world. Acessed: March 19, 2025.

Ishikura, N., Kondo, D., Vassiliades, V., Iordanov, I., and Tode, H. (2021). Dns tunneling detection by cache-property-aware features. *IEEE Transactions on Network and Service Management*, 18(2):1203–1216.

Kovar, R. (2015). Random words on entropy and dns. Splunk's Website Article. Accessed: March 19, 2025.

Mahdavifar, S., Maleki, N., Lashkari, A. H., Broda, M., and Razavi, A. H. (2021). Classifying malicious domains using dns traffic analysis. In *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pages 60–67.

Sharma, N. and Swarnkar, M. (2023). Optituned: An optimized framework for zero-day dns tunnel detection using n-grams. In *Proceedings of the 23rd IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6. IEEE.

Wang, Y., Zhou, A., Liao, S., Zheng, R., Hu, R., and Zhang, L. (2021). A comprehensive survey on dns tunnel detection. *Computer Networks*, 197:108322.