Um Método de Ofuscação para Proteger a Privacidade no Tráfego da Rede IoT

Bruna V. dos Santos[1], Andressa Vergutz[2], Michele Nogueira[2,3], Ricardo T. Macedo[1]

[1]UFSM - Departamento de Tecnologia da Informação
Universidade Federal de Santa Maria (UFSM-FW)

[2]Departamento de Informática
Universidade Federal do Paraná (UFPR)

[3]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)

bruna.vitoria@acad.ufsm.br, avergutz@inf.ufpr.br

michele@dcc.ufmg.br, rmacedo@inf.ufsm.br

Abstract. *The Internet of Things (IoT) connects objects to the Internet, offering intelligent services and applications. Network traffic analysis allows adversaries to identify devices, patterns and even user behavior, seriously compromising user privacy. In the literature, works use network traffic obfuscation techniques to avoid attacks. However, there are still breaches exploited by adversaries because not all network traffic is masked, not to mention the tradeoff between privacy and network overload. This work presents a method for obfuscating network traffic to improve user privacy in the face of IoT attacks. Particularly, the method is based on the technique of generating false traffic following different levels of obfuscation. Such levels allow a balance between network overhead and privacy. Results show the obfuscation of IoT traffic, reducing around 42% the identification precision of IoT devices.*

Resumo. *A Internet das Coisas (IoT) conecta objetos à Internet, oferecendo serviços e aplicações inteligentes. Por meio da análise do tráfego da rede, adversários identificam dispositivos, padrões e até mesmo os comportamentos dos usuários, comprometendo seriamente a privacidade. Na literatura, trabalhos propuseram o uso de técnicas de ofuscação do tráfego da rede para evitar ataques. Entretanto, ainda existem brechas exploradas pelos adversários visto que nem todo tráfego da rede é mascarado, sem mencionar o tradeoff entre privacidade e sobrecarga da rede. Este trabalho apresenta um método de ofuscação do tráfego da rede para melhorar a privacidade dos usuários enquanto mantém uma baixa sobrecarga da rede IoT. Particularmente, o método toma como base a técnica de geração de tráfego falso seguindo diferentes níveis de ofuscação. Tais níveis permitem flexibilizar a sobrecarga na rede enquanto melhora a privacidade. Os resultados mostram a ofuscação do tráfego dos dispositivos IoT, reduzindo em até 42% a precisão de identificação dos dispositivos da IoT.*

1. Introdução

O rápido avanço das tecnologias de comunicação sem fio possibilitou o surgimento da Internet das Coisas (IoT, do inglês *Internet of Things*). Esta permite a conexão de obje-

tos à Internet [Stoyanova et al. 2020] e a oferta de diferentes serviços e aplicações inteligentes através da troca de dados em tempo real [Prates Jr et al. 2019, He et al. 2018, Newaz et al. 2021]. Exemplos de aplicações consistem em casas e cidades inteligentes, além de um monitoramento contínuo da saúde dos usuários. De acordo com a Cisco [Cisco 2023] e Statista [Statista 2016], até 2023, o número de dispositivos IoT será de até três vezes a população mundial e, até 2025, haverá mais de 75 bilhões de dispositivos IoT em uso. A Cisco estima também que 48% dos dispositivos existentes até 2023 representarão dispositivos de casas inteligentes [Cisco 2023].

Há uma preocupação latente e crescente quanto à privacidade dos dados compartilhados por tais dispositivos, principalmente em ambientes residenciais [Prates et al. 2018, Newaz et al. 2021, Papadogiannaki and Ioannidis 2021]. Adversários podem analisar o tráfego de rede doméstica e identificar padrões no tráfego, nas atividades e comportamentos, violando assim a privacidade do usuário. Por exemplo, [Srinivasan et al. 2008] identificaram informações sobre os usuários de uma casa inteligente. A identificação ocorreu por meio da observação dos intervalos de tempo entre as transmissões realizadas pelos dispositivos IoT com o comportamento dos usuários. Isso resultou na identificação de cômodos, quantidade e comportamento dos usuários. Os ataques baseados na análise do tráfego da rede ferem diretamente a privacidade dos usuários de dispositivos IoT. No âmbito legal, o direito à privacidade é garantido pelo Art 2º da Lei nº 13.709, de 14 de agosto de 2018 [Brasil 2018], o que ressalta a importância de defesas e métodos de ofuscação do tráfego para proteger a privacidade.

Na literatura, as técnicas de ofuscação visam mascarar as características e comportamentos do tráfego da rede, como tamanho do pacote e *timestamp*, além de dificultar ataques. A técnica de ofuscação utilizada varia conforme objetivo e contexto. No contexto de casas inteligentes, alguns trabalhos utilizam técnicas de ofuscação do tráfego baseadas no preenchimento do tamanho dos pacotes [Pinheiro et al. 2021, Chaddad et al. 2021] e na geração de tráfego falso [Apthorpe et al. 2018, Dyer et al. 2012, Yu et al. 2021]. Apesar de ambas as técnicas ofuscarem uma porcentagem do tráfego da rede, ainda assim não são eficientes [Alyami et al. 2022]. Além disso, uma questão crítica na maioria das técnicas é a sobrecarga gerada à rede [Papadogiannaki and Ioannidis 2021], o que é crítico em redes compostas por dispositivos de baixo poder computacional, como a IoT. Portanto, torna-se imprescindível o desenvolvimento de métodos de ofuscação do tráfego da IoT, que considere o custo computacional e proteja a privacidade dos usuários.

Este trabalho apresenta o método MITRA (*A **M**ethod for **I**o**T** Nework **TR**affic Obfusc**A**tion)* para oferecer privacidade dos usuários diante dos ataques baseados no tráfego da rede no contexto de casas inteligentes. O método de ofuscação do tráfego proposto toma como base a técnica de geração de tráfego fictício (*dummy traffic*), particularmente, utiliza o tráfego verdadeiro da rede IoT e segue diferentes níveis de geração de tráfego para adaptar conforme contexto e evitar sobrecarga desnecessária da rede. Tais níveis se diferem pela quantidade de pacotes falsos gerados na rede. Dessa forma, o método MITRA oferece quatro níveis de ofuscação: baixo, médio, alto e randômico. No nível baixo gera-se uma menor quantidade de tráfego falso quando comparado ao nível alto. A quantidade e tipo de dispositivos fictícios varia conforme os dispositivos provenientes da captura original do tráfego. O método MITRA ofusca o tráfego e o comportamento da rede IoT, oferecendo diferentes níveis de tráfego para melhorar a privacidade dos usuários.

O método MITRA foi avaliado através de uma abordagem orientada a traços, seguindo duas principais etapas: *i)* análise do tráfego verdadeiro da IoT e *ii)* ofuscação do tráfego. Com o propósito de realizar uma avaliação em ambiente realista, foi utilizado o conjunto de dados *IoT Traffic Traces* que possui o tráfego de uma casa inteligente real [Sivanathan et al. 2021]. Na primeira etapa da avaliação, são extraídas as características de rede (ex., tamanho do pacote e endereço IP) e computadas as medidas estatísticas. Estas características são submetidas a classificadores supervisionados para a identificação dos dispositivos IoT. Na segunda etapa, são criados dispositivos IoT fictícios, que geram pacotes de tráfego falsos. O tráfego falso e o verdadeiro são embaralhados para a extração das características de rede e estatísticas e submetidos aos algoritmos de classificação para análise. Compara-se os resultados de identificação dos dispositivos IoT apenas com o tráfego original (sem ofuscação) *versus* tráfego com ofuscação. Por fim, comparou-se o resultado obtido pelo método MITRA com o resultado obtido com o uso da técnica de ofuscação por preenchimento de pacotes da literatura [Pinheiro et al. 2021], seguindo o mesmo *dataset*. Os resultados com o método MITRA reduziram em até 42% a identificação dos dispositivos IoT, um resultado com desempenho maior do que 20% quando comparado com a literatura [Pinheiro et al. 2021]. Além disso, o método resultou em menos de 1% de sobrecarga da rede.

O restante deste artigo procede como segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve o método proposto. A Seção 4 detalha a avaliação de desempenho juntamente com os resultados. Por fim, a Seção 5 conclui o trabalho e apresenta as direções futuras.

2. Trabalhos Relacionados

Na literatura, existem diversas técnicas de ofuscação de tráfego com o objetivo de evitar a exploração de características da rede e quebra de privacidade do usuário por adversários. Por exemplo, a técnica de ofuscação por preenchimento de pacotes [Prates Jr et al. 2019], que altera o instante de tempo (*timestamp*) dos pacotes de tráfego dos dispositivos IoT. Em [Chaddad et al. 2021, Pinheiro et al. 2021] foi utilizada a técnica de preenchimento por meio do balanceamento do tamanho dos pacotes da rede para ofuscar o tráfego dos dispositivos. Há ainda a técnica de ofuscação de geração de tráfego falso [Dyer et al. 2012, Apthorpe et al. 2018, Acar et al. 2020, Yu et al. 2021]. Esta técnica ofusca os dispositivos ao injetar pacotes falsos no tráfego. Alguns trabalhos da literatura [Datta et al. 2018, Barman et al. 2021] combinam as técnicas de ofuscação por preenchimento e por geração de pacotes. Contudo, a combinação das técnicas acarreta sobrecarga na rede. Por se tratar de um ambiente limitado de recursos, a IoT dificulta a ofuscação do tráfego da rede.

Os autores em [Yu et al. 2021] propuseram um sistema de defesa que ofuscou as atividades do usuário através da injeção de tráfego fictício na rede IoT. Entretanto, os autores não consideraram diferentes características do tráfego como por exemplo, *timestamp* e número de *bytes*, o que possibilita a realização de ataques com base nestas características. Os autores em [Dyer et al. 2012] utilizaram ambas as técnicas de geração de pacotes falsos e de preenchimento de pacotes, abordando o custo que as técnicas geram (sobrecarga da rede) *versus* a melhora da privacidade. Os autores apontaram ser essencial analisar e explorar os recursos disponíveis para compreender as particularidades do contexto em que a ofuscação será aplicada, por exemplo, considerar os poucos recursos disponíveis

na IoT. Além disso, os autores propuseram um mecanismo para inserir pacotes de tráfego falso com tamanho fixo. Porém, não foi considerada uma análise prévia para a tomada de decisão em relação ao tamanho do pacote, sendo que no tráfego de rede os pacotes possuem tamanhos variáveis. Portanto, seria mais eficaz inserir na rede pacotes falsos com tamanhos baseados nos tamanhos de pacotes identificados na rede.

Os autores em [Pinheiro et al. 2021] propuseram uma técnica de preenchimento para ofuscar o tráfego da rede IoT. Buscando atingir um equilíbrio entre privacidade e sobrecarga, eles criaram quatro níveis (100, 500, 700 e 900) de preenchimento de pacotes. Em cada nível foi comparado o número de *bytes* do pacote para incrementar o tamanho. Eles também compararam a técnica de preenchimento proposta com outras técnicas como a de Unidade Máxima de Transmissão (MTU, do inglês *Maximum Transmission Unit*), onde todos os pacotes recebem o tamanho da MTU da rede (ex. 1500 *bytes*) ou tamanho aleatório (*random*). Como resultado, os autores reduziram a possibilidade de classificação dos dispositivos IoT. No contexto de aplicativos móveis, os autores em [Chaddad et al. 2021] propuseram um sistema baseado na técnica de preenchimento para ofuscar o tamanho dos pacotes por meio de tamanhos aleatórios. A técnica seleciona o aplicativo de destino para alterar o tamanho dos pacotes do aplicativo de origem com base no de destino. Desta forma, a distribuição de probabilidade do aplicativo alvo se encaixa na distribuição de probabilidade do segundo aplicativo para confundir os classificadores de aprendizado de máquina (ML, do inglês *Machine Learning*). No entanto, as técnicas de preenchimento de pacotes tendem a aumentar consideravelmente o número de *bytes* dos pacotes da rede, incrementando a sobrecarga.

Neste sentido,este trabalho visa melhorar a privacidade dos usuários diante dos ataques baseados no tráfego da rede no contexto de casas inteligentes da IoT. Similar a [Pinheiro et al. 2021], o método proposto emprega diferentes níveis de ofuscação do tráfego da rede a fim de evitar sobrecarga na rede quando possível. Porém, diferente da literatura, o método emprega tais níveis de ofuscação em conjunto com a técnica de geração de tráfego falso e toma como base a captura original do tráfego da rede IoT. Essa captura original permite que o método crie dispositivos fictícios o mais similar possível aos dispositivos IoT originais, melhorando assim a ofuscação do tráfego.

3. MITRA: Método de Ofuscação do Tráfego da Rede IoT

Esta seção detalha o método de ofuscação MITRA que previne contra ataques baseados na análise do tráfego da rede (também conhecido como *traffic-based attacks*) e melhora a privacidade dos usuários da IoT. Diferente da literatura, MITRA toma como base o tráfego original da rede para a criação dos dispositivos fictícios e geração do tráfego fictício (*dummy traffic*). O método oferece diferentes níveis de geração de tráfego falso a fim de evitar sobrecargas desnecessárias na rede. A Figura 1 mostra as cinco etapas que compõem o método: *i)* coleta do tráfego da rede, *ii)* pré-processamento dos dados, *iii)* identificação dos dispositivos, *iv)* criação dos dispositivos fictícios e, *v)* geração do tráfego falso. As próximas subseções detalham estas etapas.

3.1. Coleta do Tráfego da Rede

Nesta primeira etapa, por meio de *sniffers*, o método MITRA captura o tráfego da rede em modo promíscuo e de forma passiva a fim de não alterar o comportamento da rede. Por

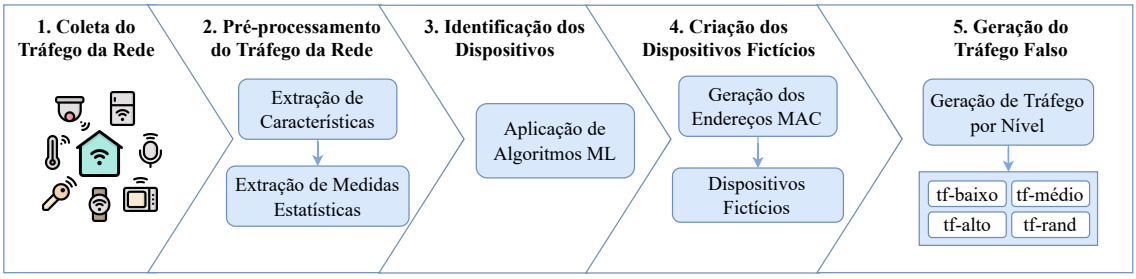


Figura 1. Etapas do Método MITRA

considerar um ambiente IoT, a captura ocorre no *gateway* ou no ponto de acesso da rede em que os dispositivos IoT estão conectados [Prates Jr et al. 2019]. Note que o dispositivo onde é realizada a coleta do tráfego da rede precisa ter acesso ou ser acessado pelos demais dispositivos conectados na rede visto que o método toma como base o tráfego capturado. A coleta do tráfego da rede serve como suporte para as demais etapas do método. Ele é exemplo para a criação dos dispositivos e geração de tráfego falso.

3.2. Pré-Processamento do Tráfego da Rede

O tráfego da rede capturado na etapa anterior serve como entrada para a etapa de pré-processamento. Nesta etapa, o método MITRA pré-processa e extrai as características de rede e estatísticas. Primeiro, ignoram-se os pacotes de rede pertencentes a comunicação *broadcast*, pacotes com tamanho zero e pacotes direcionados a servidor *Domain Name System* (DNS). Assim, são extraídas características do tráfego consideradas relevantes para as análises no âmbito da IoT. As características de rede extraídas compreendem o endereço IP do dispositivo de origem e destino, o protocolo da camada de transporte utilizado na comunicação, o instante de tempo (*timestamp*) em que os pacotes são transmitidos, o endereço MAC dos dispositivos de origem e destino, o tamanho do pacotes e o número das portas de origem e destino utilizada na comunicação.

Com o intuito de aumentar a granularidade e quantidade de informação, com base nas características de rede extraídas são realizados cálculos e extrações de medidas estatísticas. As medidas estatísticas consistem na média, mínima, máxima, variância, desvio padrão, entre outras. O método MITRA computa tais medidas estatísticas a partir das características de rede que possuem valores numéricos, como o tamanho dos pacotes e o instante de transmissão dos pacotes (*timestamp)*, considerando amostras de tamanho cinco. Na IoT existem dispositivos que transmitem uma baixa quantidade de pacotes de rede, por exemplo, dispositivos que medem a pressão do sangue enviam em média de três a dez pacotes por dia. Sendo assim, devido à natureza dos dispositivos IoT, calculam-se as medidas estatísticas a cada cinco pacotes de rede. Assim, tem-se para cada característica considerada suas respectivas medidas estatísticas. Além disso, o método MITRA filtra os dispositivos pelo endereço MAC para a rotulação dos mesmos. O endereço MAC também serve como *ground truth* para ter certeza a quem pertence o tráfego de rede. Essa informação é relevante para as comparações dos resultados na identificação dos dispositivos IoT por meio dos algoritmos de classificação.

3.3. Identificação dos Dispositivos

As características de rede e medidas estatísticas extraídas servem como entrada para a etapa de identificação dos dispositivos IoT. A identificação emprega algoritmos de apren-

dizagem de máquina indicados para problemas de multi-classificação, como Random Forest e *Decision Tree* [Papadogiannaki and Ioannidis 2021]. Uma vez que são criados rótulos para os dispositivos com base nos endereços MAC, a identificação dos mesmos utiliza algoritmos supervisionados. Assim, o método MITRA identifica os dispositivos conectados na rede por meio da aplicação de algoritmos supervisionados no conjunto de características extraído do tráfego, permitindo a adaptação em diferentes ambientes.

A aplicação dos algoritmos de aprendizagem de máquina segue o método *holdout* que consiste em dividir o conjunto de dados em treino e teste. Ou seja, neste método divide-se o conjunto de dados que contém o tráfego da rede com as características extraídas em 60-40, 60% para treino e 40% para teste. Sendo assim, a execução dos algoritmos faz o treinamento com a amostra de treino e em seguida é realizada a predição dos dados com a amostra de teste. Após a segmentação e tratamento do conjunto de dados, o método MITRA executa os algoritmos de ML. Dessa forma, o sucesso na identificação do tráfego dos dispositivos IoT simula o ataque baseado na análise do tráfego da rede IoT, ou seja, quanto maior a taxa de acerto, maior o sucesso do ataque.

3.4. Criação dos Dispositivos Fictícios

A criação dos dispositivos fictícios toma como base a identificação dos dispositivos IoT verdadeiros realizada na etapa anterior. Isso acontece devido ao objetivo de gerar um tráfego falso semelhante ao verdadeiro. Para tanto, os dispositivos fictícios são criados a partir das características de rede e estatísticas dos dispositivos IoT verdadeiros.

Para a geração dos dispositivos fictícios, considera-se uma parte do endereço MAC dos dispositivos IoT verdadeiros. Esta parte do endereço MAC corresponde ao código identificador do fabricante (OUI, do inglês *Organizationally Unique Identifier*). A Figura 3 ilustra o endereço MAC de um dispositivo verdadeiro e de um fictício. O código OUI do endereço MAC permanece igual para ambos os dispositivos, verdadeiro e fictício. Enquanto, os demais valores do endereço MAC fictício (dispositivo fictício) são gerados aleatoriamente e representados pelo controlador de interface de rede (NIC, do inglês *Network Interface Controller*). Assim, cria-se um endereço MAC fictício, utilizando o mesmo OUI, para cada endereço MAC dos dispositivos verdadeiros. Além do endereço MAC, os dispositivos fictícios recebem outros dados fictícios, como: o nome do dispositivo e o tipo de rede utilizada na comunicação, por exemplo *Wireless*.

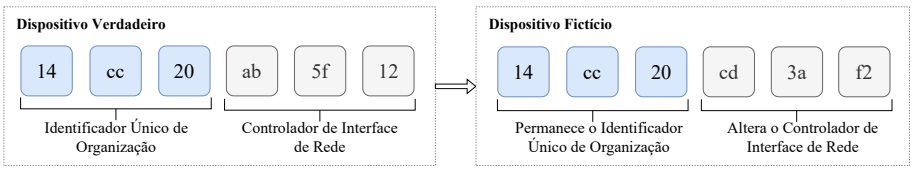


Figura 2. Geração dos Dispositivos Fictícios com base no Endereço MAC

3.5. Geração de Tráfego Falso

Nesta etapa, o método MITRA gera o tráfego falso a partir dos dispositivos fictícios criados com suas informações fictícias, como endereço MAC e o endereço IP. O método transmite todo o tráfego falso para o mesmo destino: o *Gateway* identificado na rede IoT, a fim de seguir o comportamento dos demais dispositivos IoT conectados na rede. Logo,

o dispositivo de destino recebe o endereço MAC do *Gateway* e um endereço IP aleatório. Além disso, o método MITRA segue um tamanho de pacote e número de porta de destino conforme captura original do tráfego da rede. Particularmente, o tamanho do pacote do tráfego falso segue o valor mais recorrente encontrado na captura original a fim de aproximar o comportamento fictício ao original. Desta forma, considerando que o método MITRA utiliza um destino fixo, ele preenche dinamicamente os dados referente ao dispositivo de origem (o dispositivo fictício) e estaticamente os dados do dispositivo de destino (o *Gateway*). Tais dados referem-se aos endereços MAC e IP, tamanho de pacote, número de porta, entre outros. Para a transmissão do tráfego falso, utiliza-se a mesma interface de rede empregada na captura do tráfego original. Portanto, o destino, tamanho de pacote e número de porta seguem o mesmo padrão para todo tráfego falso.

O método MITRA oferece quatro níveis de geração de tráfego falso para flexibilizar a sobrecarga na rede conforme cenário de rede. Sendo assim, o método segue os níveis: baixo (tf-baixo), médio (tf-médio), alto (tf-alto) e randômico (tf-rand) de tráfego falso, conforme mostra a Figura 3. A carga de tráfego falso gerada na rede depende do nível escolhido visto que cada nível segue uma quantidade específica de pacotes falsos. O nível tf-baixo gera uma menor quantidade de pacotes falsos quando comparado aos níveis tf-médio e tf-alto. O nível randômico alterna de forma aleatória entre um valor mínimo (valor de tf-baixo) e máximo (tf-alto) de pacotes falsos, alterando assim a quantidade de pacotes para cada dispositivo fictício. Tais níveis de geração de tráfego tem como objetivo diminuir a sobrecarga da rede em casos onde a inserção de tráfego falso possa gerar problemas no uso dos dispositivos conectados à rede. Assim, o método permite selecionar níveis de acordo com o estado atual da rede e conforme determinados cenários, viabilizando uma diminuição da sobrecarga da rede e melhoria da privacidade. Por exemplo, em rede ociosa, pode-se gerar maior quantidade de tráfego falso.

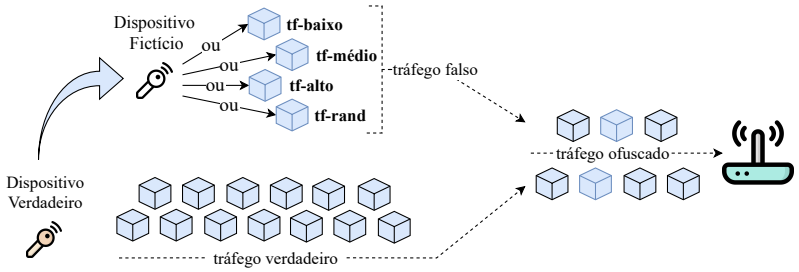


Figura 3. Funcionamento da Geração de Tráfego Falso do Método MITRA

A ofuscação do tráfego ocorre por meio da injeção do tráfego falso na rede IoT. Para isso, a geração do tráfego falso deve ser realizada na mesma rede onde encontram-se os dispositivos a serem ofuscados. Logo, deve-se selecionar um nível para a geração do tráfego de acordo com o estado da rede. De acordo com o nível escolhido, o método gera n pacotes de rede para cada dispositivo fictício criado, tal quantidade de pacotes varia conforme o nível de geração selecionado. Por fim, o tráfego falso gerado no nível selecionado é injetado na rede IoT a fim de ofuscar o tráfego original e inviabilizar a identificação dos dispositivos IoT verdadeiros. Isto, por consequência, melhora a privacidade dos usuários.

4. Avaliação de Desempenho

A avaliação do método MITRA passou pela seleção do conjunto de dados, análise e ofuscação do tráfego da rede. O método é projetado para funcionar *online*, mas para

fins de avaliação a abordagem seguida é orientada a traços, devido ao melhor controle no cenário de avaliação. Um conjunto de dados (*dataset*) serve como entrada para as duas principais etapas da avaliação: *i)* análise do tráfego de rede verdadeiro da IoT e *ii)* ofuscação do tráfego. Foi utilizado o conjunto de dados *IoT Traffic Traces* que possui o tráfego real de uma casa inteligente [Sivanathan et al. 2021]. A motivação para uso de tal conjunto de dados se deve ao fato de conter dispositivos IoT rotulados com endereço MAC, melhor compreensão dos dados e pelo amplo uso do conjunto de dados em outras pesquisas, permitindo verificações e fácil reprodução dos resultados.

Características da Base de Dados

O conjunto de dados *IoT Traffic Traces* [Sivanathan et al. 2021] possui a coleta do tráfego da rede de uma casa inteligente composta por 31 dispositivos IoT e não-IoT (incluindo o *gateway*). A coleta do tráfego da rede ocorreu do dia 22 de setembro de 2016 à 12 de outubro de 2016, totalizando 20 dias de captura com 32GB de tráfego de rede. A captura foi dividida em 20 arquivos pcaps referente a cada dia de coleta. A Figura 4 mostra a quantidade de pacotes gerados e o total de *bytes* dos principais dispositivos IoT presentes no *dataset*. O monitor de pressão arterial, seguido pelo detector de fumaça, são os dispositivos IoT que geraram uma menor quantidade de tráfego da rede. Em contrapartida, o *gateway*, câmera, assistente virtual e monitor de sono apresentaram uma maior taxa de tráfego gerado na rede. Esse comportamento de rede aponta uma disparidade na geração de tráfego falso pelos dispositivos IoT. Além dos dispositivos apresentados na Figura 4, o *dataset* possui assistentes virtuais, monitor de pressão arterial, balança Wi-Fi, notebooks, sensor de movimento, smartphones, porta-retrato digital, caixa de som portátil, interruptor de luz inteligente , entre outros. É importante salientar que independente do conjunto de dados utilizado para reprodução do método proposto, devem ser consideradas coletas compostas por dispositivos IoT e suas características de rede.

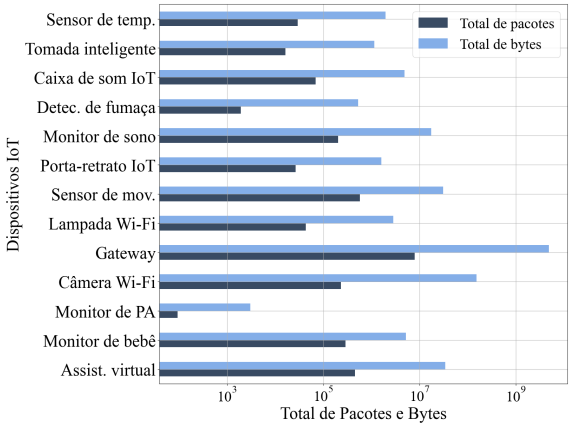


Figura 4. Volume de tráfego

Extração das Características, Identificação e Ofuscação do Tráfego dos Dispositivos

O *dataset* serve como entrada para a avaliação do método proposto. A primeira etapa da avaliação consiste da análise do tráfego de rede verdadeiro da IoT, logo, por meio da ferramenta Tshark[1] e das bibliotecas Numpy[2] e Pandas[3] em *scripts* Python v3, extraem-se as características da rede e computam-se as medidas estatísticas do tráfego bidirecional da IoT. As características de rede consistem do endereço MAC, endereço IP, protocolo da camada de transporte, tamanho dos pacotes, quantidade de pacotes, *timestamp* e tempo

[1]Tshark - Wireshark: https://www.wireshark.org/docs/man-pages/tshark.html. Acessado em Fev/2022.
[2]Biblioteca Numpy: https://numpy.org/. Acessado em Fev/2022.
[3]Biblioteca Pandas: https://pandas.pydata.org/. Acessado em Fev/2022.

entre pacotes. A partir das características de rede extraídas, computam-se as medidas estatísticas como média, mínima, máxima e desvio padrão. Criam-se amostras de tamanho cinco para calcular tais medidas, pois alguns dispositivos IoT geraram apenas cinco pacotes de rede em um dia de coleta (exemplo, o medidor de pressão do sangue). Assim, têm-se a média do tamanho dos pacotes, média dos *timestamps*, desvio padrão do tamanho dos pacotes, entre outras. Armazenam-se as características de rede e medidas estatísticas em um arquivo $.csv$ juntamente com os rótulos dos dispositivos IoT.

Em seguida, submete-se o arquivo $.csv$ à classificadores supervisionados, implementados pela biblioteca *Scikit-Learn*[4], para a identificação dos dispositivos IoT. Considerou-se os seguintes algoritmos supervisionados: *Extreme Gradient Boosting* (XGBoost), *Classification and Regression Trees* (CART), *Random Forest* e *Bootstrap Aggregating* (Bagging), visto que são comumente utilizados na literatura para a identificação de tráfego [Prates Jr et al. 2019, Acar et al. 2020, Pinheiro et al. 2021]. A análise do desempenho dos algoritmos de ML segue as métricas da acurácia, F1-Score, *recall* e precisão. Estas métricas de avaliação permitem analisar se é possível identificar os dispositivos IoT por meio da taxa de verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). Por fim, apresentam-se os resultados da identificação dos dispositivos por meio das métricas de avaliação. Esta primeira etapa da avaliação do método proposto simula o ataque baseado na análise do tráfego da rede IoT, onde o tráfego é analisado para identificar os dispositivos e inferir o comportamento do usuário.

A segunda etapa da avaliação compreende a ofuscação do tráfego dos dispositivos IoT por meio da geração de tráfego falso do método MITRA. Para isso, criam-se os dispositivos IoT fictícios com base no endereço MAC dos dispositivos verdadeiros. Cada dispositivo IoT verdadeiro identificado, com exceção do *gateway*, dá origem a um dispositivo fictício, totalizando assim 30 dispositivos fictícios. Estes dispositivos fictícios foram criados a partir da biblioteca *Python-generate_mac*[5] e a geração de tráfego foi realizada por meio da ferramenta Ostinato[6]. A geração do tráfego falso utiliza como endereço de origem o endereço MAC fictício criado e como endereço de destino os dados do *gateway*, visto que todo o tráfego da casa inteligente passa pelo *gateway*. O tráfego falso segue um valor estático para o tamanho de pacote de 120 *bytes* e o número da porta de destino 443, pois são os valores mais utilizados pelos dispositivos do *dataset*.

Além disso, a fim de evitar a sobrecarga da rede gerou-se tráfego falso seguindo os quatro níveis propostos pelo método: tf-baixo, tf-médio, tf-alto e tf-rand (randômico). Para cada nível gerou-se uma quantidade de pacotes específica por dispositivo fictício. Portanto, foram gerados 100, 600 e 1000 pacotes falsos por dispositivo para o nível tf-baixo, tf-médio e tf-alto, respectivamente. Enquanto no nível tf-rand gerou-se aleatoriamente entre 50 a 1000 pacotes por dispositivo, conforme mostra a Tabela 1. Dessa forma, em cada nível são gerados x pacotes por dispositivo fictício. Por exemplo, na geração do tráfego falso seguindo o nível tf-baixo foram gerados 100 pacotes por dispositivo fictício, logo totaliza 3000 pacotes falsos visto que foram criados 30 dispositivos fictícios. Por fim, o tráfego falso e verdadeiro são embaralhados em um mesmo conjunto de dados para a extração das características de rede e estatísticas, e submissão aos algoritmos de

[4]Biblioteca Scikit-Learn: https://scikit-learn.org/stable/. Acessado em Fev/2022.
[5]Python-generate_mac: https://pypi.org/project/python-generate-mac/. Acessado em Fev/2022.
[6]Ostinato: http://ostinato.org/. Acessado em Fev/2022.

classificação. Assim, compara-se os resultados de identificação dos dispositivos IoT apenas com o tráfego original (sem ofuscação) *versus* tráfego com ofuscação.

Tabela 1. Níveis de Geração de Tráfego Falso

Níveis	Quantidade de Pacotes por Dispositivo	Total de Pacotes
tf-baixo	100 pacotes falsos	3.000
tf-médio	600 pacotes falsos	18.000
tf-alto	1000 pacotes falsos	30.000
tf-rand	Entre 50 e 1000 pacotes falsos	Aleatório

Comparação com a Literatura

O método foi comparado com o trabalho de [Pinheiro et al. 2021], onde os autores propuseram uma técnica de preenchimento de pacotes para ofuscar o tráfego da rede IoT e atingir um equilíbrio entre a privacidade e a sobrecarga da rede.Os autores criaram quatro níveis de preenchimento de tamanho de pacotes: 100, 500, 700 e 900. Cada nível compara o número de *bytes* do pacote original para incrementar seu tamanho. Os números dos níveis se referem ao tamanho que os pacotes de rede terão após o preenchimento. Os pacotes de rede com tamanhos menores ou iguais a 100, por exemplo, passam a possuir o tamanho de 100 *bytes*. Os autores compararam esta técnica de preenchimento com outras abordadas na literatura como: linear, exponencial, *mouse elephant*, MTU, *random* e *random* 255. Além disso, os autores consideraram os algoritmos *Random Forest* e *Decision Tree* sob o mesmo *dataset* considerado neste trabalho. A partir da reprodução do trabalho de [Pinheiro et al. 2021], foram analisados e comparados os resultados obtidos com os resultados do método MITRA. A motivação da escolha do trabalho de [Pinheiro et al. 2021] se deve ao fato de ser na IoT, por utilizarem o mesmo *dataset* e possuírem abordagem similar de níveis de ofuscação, permitindo assim comparar as duas técnicas de ofuscação.

4.1. Resultados

Esta seção apresenta os resultados do método MITRA e a comparação com a literatura. Inicialmente, apresentam-se os resultados relacionados ao aumento da privacidade através do método proposto. Em seguida, compara-se a aplicação da técnica de ofuscação por preenchimento de pacotes de [Pinheiro et al. 2021], representativa da literatura, com o método MITRA. A Figura 5 apresenta o desempenho dos algoritmos de ML por meio das métricas da acurácia, precisão, *recall* e F1-score na identificação dos dispositivos IoT considerando apenas o tráfego original da rede. No geral, todos os classificadores apresentaram uma alta taxa de acerto na classificação do tráfego dos dispositivos IoT, com valores $\approx 65\%$ e $\approx 70\%$ de precisão, *recall* e F1-score. A acurácia atingiu $\approx 98\%$ em todos os classificadores. As outras métricas apresentaram valores menores que a acurácia devido à variabilidade do tráfego da rede dos dispositivos IoT. Entretanto, ainda assim, a identificação do tráfego dos dispositivos IoT atingiu valores maiores que 60%, possibilitando que atacantes aprendam o comportamento dos usuários por meio do

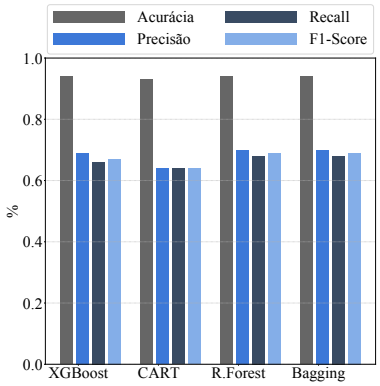
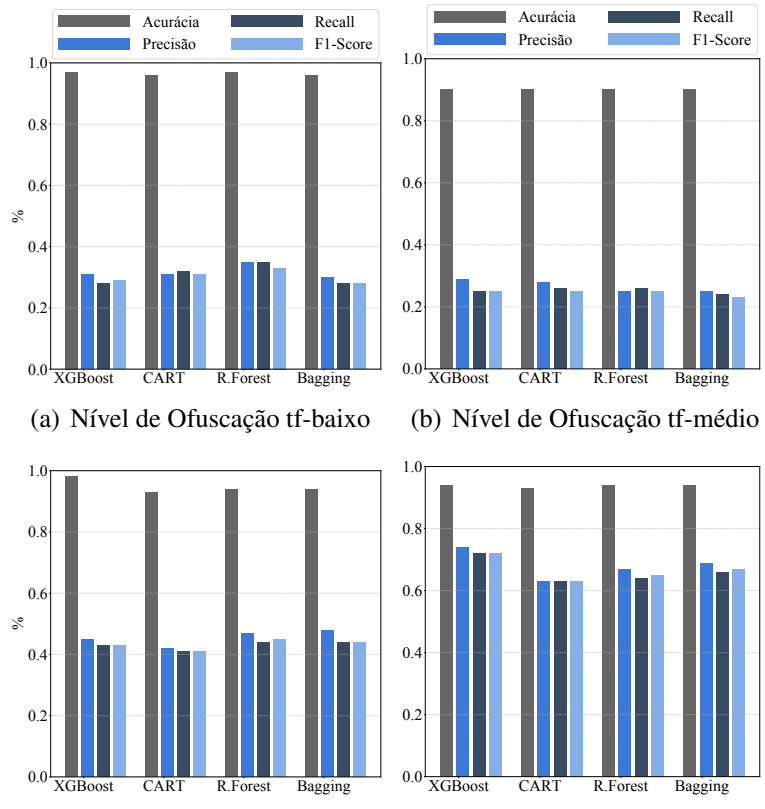


Figura 5. Identificação

tráfego dos dispositivos identificados. Uma vez identificado o dispositivo, o atacante monitora seu tráfego para aprender sobre a rotina do usuário.

A Figura 6 apresenta o desempenho do método MITRA na ofuscação do tráfego dos dispositivos IoT seguindo os quatro níveis de geração de tráfego falso (tf-baixo, tf-médio, tf-alto e tf-rand). A Figura 6(a) mostra os resultados da ofuscação seguindo o nível tf-baixo, onde o método MITRA obteve acurácia em $\approx 96\%$, precisão em $\approx 31\%$, *recall* em $\approx 30\%$ e F1-score em $\approx 30\%$. Ou seja, o método conseguiu reduzir o desempenho de classificação em até 30% quando comparado aos resultados de identificação da Figura 5. Na aplicação do nível tf-médio, o método atingiu $\approx 90\%$, $\approx 28\%$, $\approx 25\%$, $\approx 25\%$, de acurácia, precisão, *recall* e F1-score, respectivamente, conforme apresentado na Figura 6(b). Sendo assim, no nível tf-médio, o método reduziu ainda mais o desempenho de classificação ($\approx 40\%$ de redução), melhorando a privacidade dos usuários. Em contrapartida, na geração de 1000 pacotes falsos, seguindo o nível de ofuscação tf-alto, o método MITRA obteve $\approx 94\%$, $\approx 46\%$, $\approx 43\%$, $\approx 43\%$, de acurácia, precisão, *recall* e F1-score, respectivamente, conforme apresentado na Figura 6(c). Note que a acurácia permaneceu com resultados altos ($\approx 90\%$) por apresentar o resultado geral da classificação do tráfego, sem considerar separadamente os erros de classificação dos dispositivos que geraram pouca quantidade de dados. Dessa forma, com o aumento de pacotes falsos gerados, o método MITRA reduziu a eficiência na ofuscação.



(a) Nível de Ofuscação tf-baixo (b) Nível de Ofuscação tf-médio

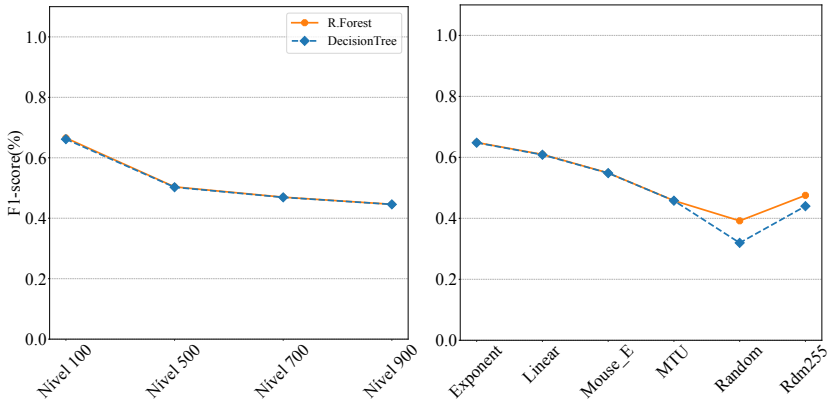
(c) Nível de Ofuscação tf-alto (d) Nível de Ofuscação tf-rand

Figura 6. Resultado da Ofuscação do Tráfego IoT do Método MITRA

A Figura 6(d) corrobora para que o aumento da quantidade de tráfego falso reduza a eficiência da ofuscação, atingindo em torno de 66% de precisão e F1-Score no nível de

ofuscação tf-rand. Nesse nível de ofuscação, o método MITRA gerou mais de 30.000 pacotes falsos, levando os classificadores a aumentar sua taxa de acerto na identificação do tráfego dos dispositivos por haver uma maior quantidade de amostra de dados. Dessa forma, o nível tf-médio obteve o melhor desempenho na ofuscação do tráfego, reduzindo de 18 dispositivos identificados para cerca de 8 dispositivos. Isso representa uma diminuição em 42% na identificação dos dispositivos ao aplicar o método MITRA. Além da redução significativa na identificação dos dispositivos, ao aplicar o nível tf-baixo e tf-médio utiliza-se menos recursos computacionais, sendo um bom resultado para ambiente IoT visto que quanto menor o tráfego falso adicionado na rede, menor será a sobrecarga.

A Figura 7(a) apresenta os resultados obtidos na ofuscação do tráfego da IoT seguindo as técnicas de preenchimento de pacotes da literatura [Pinheiro et al. 2021]. Para fins de comparação, a Figura 7(a) apresenta os resultados de ofuscação obtidos na replicação do trabalho de [Pinheiro et al. 2021], seguindo os níveis 100, 500, 700 e 900 de preenchimento do tamanho de pacote. O nível 900 de preenchimento de pacotes apresentou o melhor resultado de ofuscação, quando comparado aos outros níveis, obtendo $\approx$ 46% de F1-Score. Entretanto, este nível incrementa todos os tamanhos dos pacotes da rede para 900 *bytes*, aumentando assim a sobrecarga na rede. A Figura 7(b) apresenta os resultados da aplicação das técnicas de preenchimento amplamente conhecidas na literatura, como *Exponencial* (Exp), *Linnear* (Lin), *Mouse_Elephant* (M_E), *Maximum Transmission Unit* (MTU), *Random* (R) e *Random255* (R255). Entre elas, o *Random* e MTU apresentaram um melhor desempenho de ofuscação do tráfego, atingindo em torno de 30% e 40% de F1-Score, respectivamente. Quando comparado aos resultados de ofuscação do método MITRA, o nível tf-alto obteve resultados similares ao nível 900 de [Pinheiro et al. 2021] e MTU. No entanto, o nível 100 de [Pinheiro et al. 2021] que incrementa o tamanho de pacotes para 100 *bytes* não apresentou um resultado satisfatório de ofuscação ($\approx$ 63% de F1-Score). O nível tf-baixo do método MITRA atingiu um melhor resultado na ofuscação, dificultando a identificação dos dispositivos IoT.



(a) Ofuscação de [Pinheiro et al. 2021] (b) Técnicas da Literatura

Figura 7. Resultados da Literatura [Pinheiro et al. 2021]

A Tabela 2 mostra a sobrecarga da rede causada pelos níveis de ofuscação. O nível tf-médio do método MITRA gerou 600 pacotes por dispositivo fictício com o tamanho de 120 *bytes* cada, gerando assim um total de 0,82% de sobrecarga na rede. Em contrapartida, no nível 900 do método de [Pinheiro et al. 2021] todos os pacotes da rede passaram a ter o tamanho de 900 *bytes*, o que resultou em 266,8% de sobrecarga da rede, sendo um

aumento crítico para IoT. Portanto, conclui-se que o método MITRA apresentou melhor desempenho na ofuscação do tráfego da rede, quando comparado aos métodos de preenchimento de pacotes da literatura. Observa-se esses resultados no quesito de privacidade, pois as taxas de identificação dos dispositivos ficaram 20% menores do que no método MITRA. Além disso, no método MITRA a sobrecarga da rede foi menor do que 1%.

Tabela 2. Comparação da Sobrecarga de Rede dos Níveis de Ofuscação

Tipo de Tráfego	Tamanho do Tráfego	% de Sobrecarga da Rede
Tráfego Original	269,41 MB	0%
MITRA tf-baixo	269,79 MB	**0,13**%
MITRA tf-médio	271,65 MB	**0,82**%
MITRA tf-alto	273,13 MB	1,3%
[Pinheiro et al. 2021] 500	1622,9 MB	602,2%
[Pinheiro et al. 2021] 700	2271,49 MB	843,1%
[Pinheiro et al. 2021] 900	2920,49 MB	1084%

5. Conclusão

Este artigo apresentou o método MITRA, que ofusca o tráfego da rede IoT a fim de protegê-lo contra ataques baseados no tráfego. O método utiliza o tráfego verdadeiro da rede IoT e segue diferentes níveis de geração de tráfego falso (*dummy traffic*) para evitar sobrecarga desnecessária na rede. Sua avaliação seguiu uma abordagem orientada a traços tomando como entrada um conjunto de dados de uma casa inteligente, além de seus resultados serem comparados com resultados de técnicas representativas da literatura. Os resultados do método MITRA reduziram em até 42% a identificação dos dispositivos IoT, um resultado com desempenho maior do que 20% quando comparado com a literatura [Pinheiro et al. 2021], protegendo contra ataques. Além disso, o método apresentou menos do que 1% de sobrecarga na rede. Como trabalhos futuros espera-se expandir as análises sob ambientes experimentais IoT e reduzir ainda mais a eficiência de tais ataques a fim de melhorar a privacidade, sem sobrecarregar a rede.

Referências

Acar, A., Fereidooni, H., Abera, T., Sikder, A. K., Miettinen, M., Aksu, H., Conti, M., Sadeghi, A.-R., and Uluagac, S. (2020). Peek-a-boo: I see your smart home activities, even encrypted! In *ACM WiSec*, pages 207–218. ACM.

Alyami, M., Alharbi, I., Zou, C., Solihin, Y., and Ackerman, K. (2022). WiFi-based IoT devices profiling attack based on eavesdropping of encrypted wifi traffic. In *IEEE CCNC*, pages 385–392. IEEE.

Apthorpe, N., Huang, D. Y., Reisman, D., Narayanan, A., and Feamster, N. (2018). Keeping the smart home private with smart (er) IoT traffic shaping. *arXiv preprint arXiv:1812.00955*.

Barman, L., Dumur, A., Pyrgelis, A., and Hubaux, J.-P. (2021). Every byte matters: Traffic analysis of bluetooth wearable devices. *arXiv preprint arXiv:2105.11172*.

Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: `http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`. Acessado em Fevereiro, 2022.

Chaddad, L., Chehab, A., Elhajj, I. H., and Kayssi, A. (2021). Optimal packet camouflage against traffic analysis. *ACM Transactions on Privacy and Security (TOPS)*, 24(3):1–23.

Cisco (2018-2023). Cisco Annual Internet Report (2018–2023) White Paper. Disponível em: `https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html`. Acessado em Fevereiro, 2022.

Datta, T., Apthorpe, N., and Feamster, N. (2018). A developer-friendly library for smart home IoT privacy-preserving traffic obfuscation. In *ACM Workshop IoT S&P*, pages 43–48. ACM.

Dyer, K. P., Coull, S. E., Ristenpart, T., and Shrimpton, T. (2012). Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail. In *IEEE SSP*, pages 332–346. IEEE.

He, D., Ye, R., Chan, S., Guizani, M., and Xu, Y. (2018). Privacy in the Internet of things for smart healthcare. *IEEE Communication Magazine*, 56(4):38–44.

Newaz, A. I., Sikder, A. K., Rahman, M. A., and Uluagac, A. S. (2021). A survey on security and privacy issues in modern healthcare systems: Attacks and defenses. *ACM Transactions on Computing for Healthcare (HEALTH)*, 2(3):1–44.

Papadogiannaki, E. and Ioannidis, S. (2021). A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys (CSUR)*, 54(6):1–35.

Pinheiro, A. J., de Araujo-Filho, P. F., Bezerra, J. d. M., and Campelo, D. R. (2021). Adaptive packet padding approach for smart home networks: A tradeoff between privacy and performance. *IEEE Internet of Things Journal*, 8(5):3930–3938.

Prates, N., Pelloso, M., Macedo, R., and Nogueira, M. (2018). Ameaças de segurança, defesas e análise de dados em IoT baseada em SDN. *Minicursos SBSeg*.

Prates Jr, N., Vergütz, A., Macedo, R., and Nogueira, M. (2019). Um mecanismo de defesa contra ataques traffic side-channel temporais na IoT. In *SBSeg*, pages 323–336. SBC.

Sivanathan, A., Gharakheili, H. H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., and Sivaraman, V. (2021). IoT traffic traces. Available: `https://iotanalytics.unsw.edu.au/iottraces`. Accessed February, 2022.

Srinivasan, V., Stankovic, J., and Whitehouse, K. (2008). Protecting your daily in-home activity information from a wireless snooping attack. In *UBICOMP*, pages 202–211. ACM.

Statista (2016). Internet of things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). Available: `https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/`. Acessado em Jan/2022.

Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis, E., and Markakis, E. K. (2020). A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. *IEEE Communications Surveys and Tutorials*, 22(2):1191–1221.

Yu, K., Li, Q., Chen, D., Rahman, M., and Wang, S. (2021). Privacyguard: Enhancing smart home user privacy. In *IEEE/ACM IPNS*, pages 62–76. ACM.