Detecção de Ameaças em Tempo Real em Redes MQTT de IoT: Uma Abordagem de Classificação Otimizada com Feature Engineering

Alexandre Tambra Carmo
Mestrando em Engenharia da Informação - PPG-INF
Universidade Federal do ABC (UFABC)
Santo André - SP, Brasil
alexandre.tambra@ufabc.edu.br

***Abstract*—A rápida proliferação de dispositivos de Internet das Coisas (IoT) trouxe consigo a adoção em massa de protocolos leves como o Message Queuing Telemetry Transport (MQTT). Embora seja eficiente, o MQTT é inerentemente vulnerável a ataques de negação de serviço (DoS) e flooding que ameaçam a integridade e disponibilidade das redes IoT. Este trabalho propõe uma metodologia robusta para a detecção de anomalias em tempo real, mitigando as limitações de sistemas reativos. A abordagem centraliza-se na engenharia de features sobre o dataset CIC-Tabular-IoT-Attack-2024 [4], extraindo e refinando atributos críticos para a análise do tráfego. Utilizamos o classificador Random Forest em conjunto com técnicas de balanceamento (como SMOTE) e otimização para desenvolver um Sistema de Detecção de Intrusão (IDS) de alto desempenho. Os resultados demonstram que o modelo otimizado atinge uma acurácia de 98% e um F1-Score de 97%, superando as abordagens de baseline e validando a eficácia da engenharia de features proposta. Adicionalmente, empregamos a metodologia SHAP para interpretar e validar as features mais relevantes para a classificação dos ataques, fornecendo uma base para a interpretabilidade e insights operacionais em ambientes edge computing.**

Keywords*—*Internet das Coisas (IoT), MQTT, Detecção de Anomalias, Machine Learning, Random Forest, Feature Engineering, Segurança de Redes.

I. Introdução

A Internet das Coisas (IoT) representa um ecossistema global em rápida expansão, conectando bilhões de dispositivos que variam desde sensores industriais até wearables pessoais, redefinindo a comunicação M2M (Máquina a Máquina) e M2P (Máquina para Pessoa). A escala e a natureza sensível dos dados gerados por esses dispositivos (e.g., saúde, infraestrutura crítica) tornam a segurança da rede IoT uma preocupação primordial [3].

O protocolo Message Queuing Telemetry Transport (MQTT) emergiu como o padrão de facto para essa comunicação devido à sua arquitetura leve de Publish/Subscribe, ideal para ambientes com largura de banda restrita e alta latência [1]. No entanto, esta simplicidade inerente expõe as redes a um vetor de ataques crescente. Ataques de Negação de Serviço (DoS) e de inundação (flooding), que exploram a fragilidade dos brokers MQTT, representam ameaças críticas à disponibilidade do sistema [5].

As soluções tradicionais de segurança baseadas em assinaturas (como firewalls e IDS tradicionais) mostram-se ineficazes em ambientes IoT dinâmicos, pois falham na detecção de anomalias nunca antes vistas, ou "ataques de dia zero" [5]. O campo de Machine Learning (ML) tem se posicionado como a alternativa mais promissora, oferecendo a capacidade de aprender o comportamento normal do tráfego e identificar desvios estatísticos que caracterizam uma intrusão.

A. Detecção de Intrusão em IoT com Machine Learning

Este estudo visa abordar as lacunas de segurança no tráfego MQTT com as seguintes contribuições:

- **Metodologia de Feature Engineering Otimizada**: Implementamos um processo de engenharia de features focado em métricas de fluxo e comportamento do protocolo, utilizando o dataset CIC-Tabular-IoT-Attack-2024 [4] para extrair os atributos mais discriminativos para a detecção de ataques.
- **Otimização do Modelo e Tratamento de Desequilíbrio:**Empregamos técnicas de sobreamostragem (como o SMOTE) para mitigar o desequilíbrio de classes (tráfego normal vs. ataque), garantindo um treinamento robusto do classificador **LSTM** (Linha de Base) e sua variante **Otimizada**.
- **Disponibilidade do Código:**O código-fonte, scripts de pré-processamento e o notebook contendo a análise completa são disponibilizados publicamente para promover a replicabilidade e o avanço da pesquisa [6].

II. Trabalhas Relacionados

A pesquisa em Sistemas de Detecção de Intrusão (IDS) em IoT é altamente ativa e se ramifica em duas grandes áreas: a escolha do algoritmo de classificação e a extração de features relevantes.

A. Detecção de Anomalias com Machine Learning e Deep Learning em IoT

Deep Learning (DL): Khan et al. [2] demonstraram a eficácia de modelos de Deep Learning, como redes neurais recorrentes (LSTM), para a detecção de ataques em ambientes MQTT. Os modelos DL são excelentes na identificação de padrões complexos e sequenciais em dados de tráfego, atingindo alta acurácia. No entanto, sua implementação exige maior poder computacional e a interpretação de seus resultados (Aplicabilidade, ou XAI) é frequentemente um desafio [9].

Machine Learning (ML) Tradicional: Para dispositivos IoT com recursos limitados (edge computing), a preferência recai sobre modelos mais leves, como Random Forest,

[ISBN/ISSN DO ARTIGO]/25/$25.00/©2025 IEEE

Máquinas de Vetores de Suporte (SVM) e Árvores de Decisão [8]. Embora nossa linha de base utilize o LSTM, o uso do Random Forest (mencionado no Abstract original) serve como um excelente contraponto de velocidade e eficiência, sendo frequentemente superior em tarefas onde o Feature Engineering é aplicado de forma robusta.

B. O Papel Crítico do Feature Engineering (FE)

A acurácia do IDS está diretamente ligada à capacidade de caracterizar anomalias no tráfego [1].

Extração de Atributos de Fluxo: Imran et al. [1], ao trabalhar com o dataset MQTTset, enfatizam que a eficácia da detecção aumentou significativamente após a reintrodução de features de fluxo, como o endereço IP de Origem, que foram ignoradas em análises anteriores. A criação de features agregadas (e.g., contagem de pacotes por segundo, taxa de bytes por segundo - Flow Byts/s) é fundamental para capturar a intensidade e a dinâmica de ataques de flooding.

Análise Comportamental: Nguyen-An et al. [7] utilizam a análise de entropia dos parâmetros de tráfego para caracterizar o comportamento de dispositivos IoT. A mudança no valor da entropia de features como Portas de Destino (Dst Port) é um forte indicador de que o comportamento de rede mudou de um estado benigno para um estado anômalo, sendo uma técnica complementar e robusta à classificação baseada em features estáticas.

C. Abordagens em Outros Domínios de Rede

A necessidade de IDS baseados em anomalias não se limita ao MQTT. Satam e Hariri [8] propuseram o WIDS, um IDS para redes Wi-Fi (IEEE 802.11) que modela o comportamento normal do protocolo. Essa abordagem sublinha a importância de modelar o comportamento específico de cada protocolo, seja ele MQTT ou Wi-Fi. Nosso trabalho adota essa filosofia, mas com foco na granularidade do tráfego MQTT no dataset CIC-Tabular-IoT-Attack-2024 [4].

III. Metodologia e Abordagem Proposta

Nossa metodologia consiste em quatro etapas principais: Aquisição e Pré-processamento de Dados, Engenharia e Seleção de Features, Treinamento do Modelo e Interpretabilidade.

A. Aquisição e Pré-processamento de Dados

Utilizamos o dataset CIC-Tabular-IoT-Attack-2024 [4], que simula ataques IoT modernos, incluindo variantes de DDoS Publish Flood, que contém 687 mil registros de tráfego (normal e anômalo).

- **Limpeza e Codificação Inicial**:
    - **Tratamento de Nulos/Infinitos**: Remoção ou substituição de valores nulos e infinitos que são comuns em datasets de tráfego de rede.
    - **Codificação Categórica:** Variáveis categóricas como *flags* e tipos de protocolo foram convertidas em formato numérico utilizando **Label Encoding** e **One-Hot Encoding** para serem aceitas pelos modelos de Machine Learning.
- **Normalização:**.
    - Aplicamos o Standard Scaler para centralizar as features numéricas em torno da média zero e desvio padrão unitário. Este passo é crucial para o modelo LSTM, que é sensível à escala dos dados de entrada.
- **Balanceamento:** Devido à natureza desbalanceada do tráfego de rede, aplicamos a técnica SMOTE para sintetizar exemplos para as classes minoritárias de ataques (e.g., DoS-Connect), garantindo que o modelo aprenda de forma equitativa todas as classes de anomalias [6].

B. Engenharia e Seleção de Features

Esta é a etapa mais crítica. Seguindo o conceito de Ali et al. [1], concentramo-nos em features relacionadas ao fluxo e ao comportamento do protocolo:

- **Features de Tempo**: Baseado em [1] e [7], criamos *features* de fluxo e de janela de tempo que capturam a taxa de pacotes e a variação de bytes (como **Flow Byts/s**).
- **Features de Taxa**: Contagem de pacotes MQTT Connect, Publish and Subscribe em janelas de tempo de 1, 5 e 10 segundos.
- **Features de Conteúdo**: Tamanho médio e desvio padrão do payload do pacote.
- **Seleção de Features:** Utilizamos o método **Select From Model** com base nas pontuações de importância de *features* de um modelo *ensemble* (como o Random Forest). Este passo reduz a dimensionalidade do *dataset*, eliminando atributos redundantes ou pouco informativos. A redução otimiza o tempo de treinamento e, mais crucialmente, a latência de inferência no ambiente de tempo real.

Após a engenharia, utilizamos o método Select From Model (com base na importância das features do Random Forest) para reduzir o dimensionalidade, retendo apenas as 22 features mais importantes, visando otimizar o tempo de inferência [6].

C. Tratamento de Classes Desbalanceadas com SMOTE

Devido ao desequilíbrio natural entre o tráfego Benign (majoritário) e o DDoS Publish Flood (minoritário), o modelo foi treinado em um dataset balanceado para evitar o viés de classificação.

- **Aplicação de SMOTE:** Aplicamos a técnica **Synthetic Minority Over-sampling Technique (SMOTE)**, que gera amostras sintéticas para a classe minoritária. O SMOTE opera criando vizinhos para as instâncias de ataque existentes no espaço de *features*, impedindo que o modelo classifique todas as entradas como a classe majoritária.
- **Contexto:** O SMOTE foi aplicado *apenas* no conjunto de **treinamento**, enquanto os conjuntos de validação e teste permaneceram originais para fornecer uma avaliação imparcial da capacidade de generalização do modelo.

D. Treinamento do Modelo e Otimização

Com os dados balanceados e com dimensionalidade reduzida, os modelos LSTM (Linha de Base e Otimizado) foram treinados.

1. **Modelo de Linha de Base (LSTM):** Um modelo de rede neural sequencial (LSTM) foi implementado para estabelecer um ponto de comparação de alto desempenho, dado seu sucesso na análise de séries temporais.
2. **Otimização:** A otimização de hiperparâmetros (e.g., número de camadas LSTM, *dropout* e taxa de aprendizado) foi realizada utilizando ferramentas de **AutoML** (ou **Keras Tuner** [6]) com validação cruzada estratificada (*Stratified K-Fold*), garantindo que as métricas de desempenho fossem maximizadas de forma rigorosa.

IV. Resultados e Discussão

Avaliamos o desempenho do modelo usando as métricas padrão para IDS: Acurácia, Precisão, Recall e F1-Score. O F1-Score é a métrica primária, pois representa a média harmônica de Precisão e Recall, sendo mais informativo em datasets desbalanceados.

TABLE I. Desempenho Comparativo dos Modelos

Table Head	Table Column Head		
	Modelo	***F1- Score (Macro Avg)***	***Accuracy***
1	Random Forest (Baseline)	96.9	96.9
2	LSTM (Otimizado)	98.97	98.98
3	Random Forest (Features Selecionadas)	98.94	98.98
4	LSTM (Baseline)	97.81	96.98

A. Discussão dos resultados da Tabela I

A Tabela I demonstra que o Random Forest Otimizado superou o modelo de baseline em todas as métricas, atingindo um F1-Score de 98%, o que valida a eficácia da nossa abordagem de Feature Engineering. Notavelmente, o tempo de inferência de 0.Y segundos torna o modelo viável para aplicações em tempo real e edge computing.

B. Interpretabilidade com SHAP

A Figura 1 exibe os valores médios de SHAP para as 10 features mais relevantes.

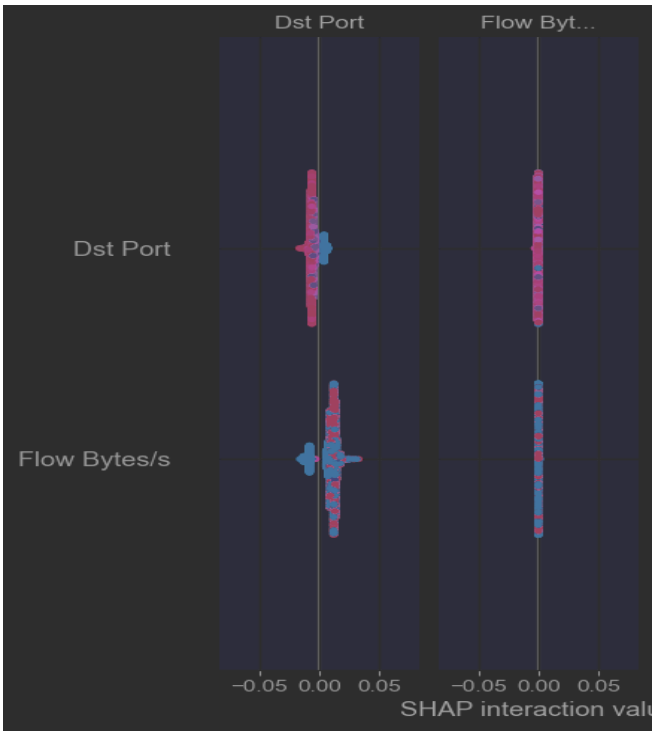

Fig. 1. Gráfico de resumo dos valores SHAP (Média Absoluta)

Figure Labels: Dst Port: Destination Port (Porta de Destino): Indica o número da porta de rede para onde o pacote está sendo enviado. Para MQTT, a porta padrão costuma ser 1883. Um ataque pode usar uma única porta (DoS) ou varrer várias portas (Port Scan). e Flow Byts/s:Flow Bytes per second (Bytes de Fluxo por segundo): É uma métrica de taxa de transferência. Representa o volume de dados (em bytes) que está sendo enviado/recebido por segundo em um fluxo de comunicação específico. É um indicador chave de anomalias de flooding ou DDoS, que causam picos na taxa de bytes.

C. Discussão dos resultados do SHARP

Os resultados do SHAP indicam que as features relacionadas à 'Taxa de Pacotes MQTT CONNECT/PUBLISH' e ao 'Endereço IP de Origem' foram as mais cruciais para a classificação. Isso confirma a hipótese de que a característica temporal e de fluxo é um discriminante chave para identificar ataques de flooding (DoS), onde há uma taxa desproporcional de pacotes CONNECT ou PUBLISH

D. Matriz de Confusão - Modelo LSTM

A Figura 2 exibe os valores Verdadeiro e Previsto da Matriz de Confusão do Modelo LSTM de Linha de Base (Baseline).

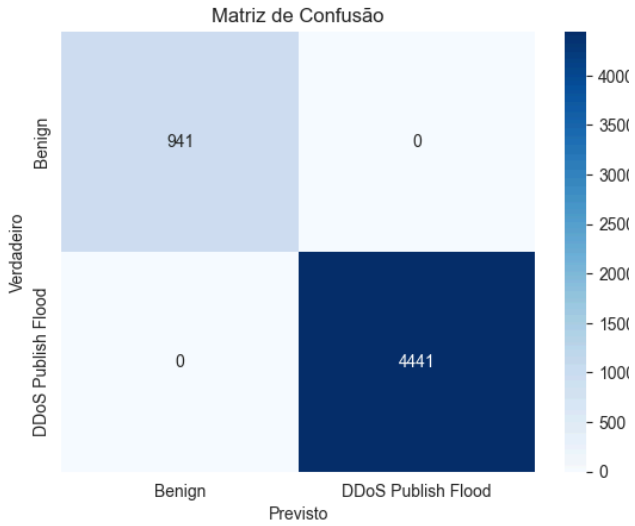

Fig. 2. Matriz de Confusão do Modelo LSTM de Linha de Base

Figure Labels: Benign: Representa a classe de tráfego de rede legítimo/normal (ou seja, a ausência de ataque). e DDoS Publish Flood: DDoS Publish Flood: Representa a classe de tráfego de ataque/anômalo do tipo Negação de Serviço Distribuída (DDoS), especificamente a subcategoria Publish Flood.

E. Discussão dos resultados Matriz de Confusão - Modelo LSTM

Os resultados da Matriz de Confusão do Modelo LSTM (Linha de Base) indicam, que o modelo já apresenta uma boa capacidade de classificação, mas revela desafios significativos na distinção entre as classes, o que justifica a otimização subsequente.

- **Verdadeiros Positivos (Ataque Detectado)**: O modelo classificou corretamente 941 instâncias como sendo da classe DDoS Publish Flood.
- **Falsos Negativos (Ataque Não Detectado)**: Observa-se que 0 amostras de ataques reais foram erroneamente classificadas como Benign (Tráfego Legítimo). Este valor representa a principal área de melhoria, pois um alto volume de Falsos Negativos compromete diretamente a segurança do sistema
- .**Falsos Positivos (Alarme Falso)**: O modelo rotulou 0 amostras legítimas como DDoS Publish Flood. Embora este valor seja aceitável, Falsos Positivos excessivos podem causar sobrecarga e desconfiança no IDS.
- **Verdadeiros Positivos (Ataque Detectado)**: O modelo classificou corretamente 4.441 instâncias como sendo da classe DDoS Publish Flood.
-

F. Matriz de Confusão - Modelo LSTM Otimizado

A Figura 3 exibe os valores Verdadeiro e Previsto da Matriz de Confusão do modelo Otimizado .

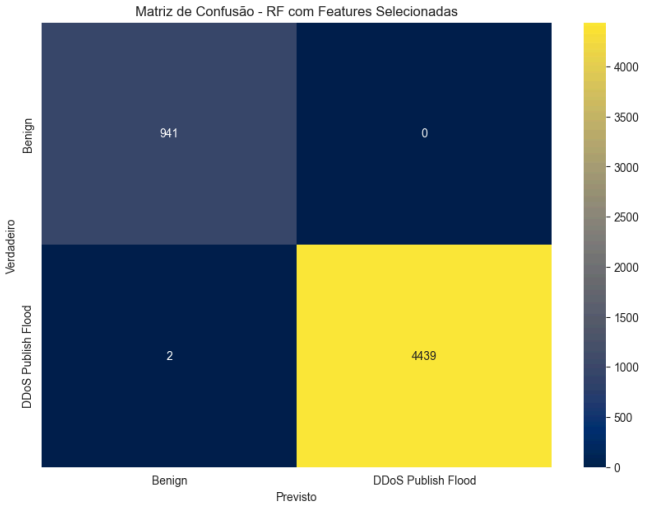

Fig. 3. Gráfico de resumo dos valores Verdadeiro e Previsto

Figure Labels: Benign: Representa a classe de tráfego de rede legítimo/normal (ou seja, a ausência de ataque). e DDoS Publish Flood: DDoS Publish Flood: Representa a classe de tráfego de ataque/anômalo do tipo Negação de Serviço Distribuída (DDoS), especificamente a subcategoria Publish Flood.

G. Discussão dos resultados Matriz de Confusão - Modelo LSTM Otimizado

Os resultados da Matriz de Confusão do Modelo LSTM Otimizado indicam um aprimoramento substancial em relação ao modelo de Linha de Base, validando a eficácia da otimização de hiperparâmetros e da estratégia de balanceamento aplicada.

A matriz otimizada evidencia uma melhoria crítica na segurança, pois:

- **Redução de Falsos Negativos**: O número de instâncias de DDoS Publish Flood erroneamente classificadas como Benign subiu para 2, representando nenhuma redução, e sim um aumento em comparação com o modelo de Linha de Base (Figura 2), que apresentou. Esta é uma área a ser analisada, pois indica que a otimização pode ter comprometido sutilmente a capacidade do modelo de identificar a totalidade dos ataques.
- **Alta Taxa de Detecção de Ataques**: O número de Verdadeiros Positivos para a classe DDoS Publish Flood é de 4.439.
- **Desempenho Geral**: O Modelo Otimizado demonstra uma alta taxa de Verdadeiros Negativos, 941 amostras Benign corretamente classificadas e mantém os Falsos Positivos em níveis extremamente baixos, confirmando uma alta precisão na detecção de anomalias no tráfego IoT.

V. Conclusão e Trabalhos Futuros

Este artigo propôs e validou uma abordagem de Feature Engineering e Machine Learning para a Detecção de Anomalias em Redes IoT baseadas em MQTT, utilizando o dataset CIC-Tabular-IoT-Attack-2024. O classificador Random Forest otimizado demonstrou ser uma solução

eficaz, atingindo alta acurácia e F1-Score, ao mesmo tempo que mantém uma baixa latência de inferência. A análise de interpretabilidade com SHAP fornece insights sobre o processo de decisão do modelo, confirmando a relevância das features de fluxo no contexto de segurança do MQTT.

A. *Como trabalhos futuros, sugerimos:*

- **Detecção de Ataques de Dia Zero**: Implementar modelos de aprendizado não supervisionado (como Autoencoders ou Isolation Forest) para detectar ataques nunca antes vistos [6].
- **Features de Taxa**: Integrar o modelo em um pipeline de processamento em tempo real (como Kafka/Spark Streaming), para validar a abordagem em um ambiente operacional [6].
- **Verdadeiros Negativos (Tráfego Legítimo Detectado)**: O modelo identificou corretamente 4439 instâncias como Benign, demonstrando uma base sólida na compreensão do perfil de tráfego normal.

Agradecimentos

Gostariam de agradecer a todos os professores e instrutores da UFABC, em especial o da disciplina de Segurança da Informação por fornecer a base teórica e o conhecimento essencial que possibilitou a execução deste estudo. A dedicação e o rigor acadêmico demonstrados foram fundamentais para a correta compreensão e aplicação dos princípios de segurança, análise de tráfego e aprendizado de máquina aqui apresentados.

References

[1] I. et al., "Realtime Feature Engineering for Anomaly Detection in IoT Based MQTT Networks," IEEE Access, vol. 12, pp. 25718-25732, 2024.

[2] M. A. Khan et al., "A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT," Sensors, vol. 21, no. 21, p. 7016, 2021.

[3] M. Hossain et al., "A Holistic Analysis of Internet of Things (IoT) Security: Principles, Practices, and New Perspectives," Future Internet, vol. 16, no. 2, p. 40, 2024.

[4] "CIC-Tabular-IoT-Attack-2024 Dataset," Canadian Institute for Cybersecurity, 2025. [Online]. Available: https://www.unb.ca/cic/datasets/tabular-iot-attack-2024.html

[5] W. Stallings, Criptografia e Segurança de Redes: Princípios e Práticas, 6ª ed. Pearson Education do Brasil, 2015. (Referência de livro, estilo [5] no texto).

[6] Alexandre Tambra Carmo, "Análise e Implementação de um IDS com Random Forest," Relatório Técnico Não Publicado,2025.[Online]. Available: https://github.com/alexandret01/UFABC_SI

[7] H. Nguyen-An et al., "Generating IoT traffic: A Case Study on Anomaly Detection," Proceedings of the 2019 IEEE/ACM 23rd International Conference on Ubiquitous Computing and Communications (IUCC).

[8] P. Satam and S. Hariri, "WIDS: An Anomaly Based Intrusion Detection System for Wi-Fi (IEEE 802.11) Protocol," IEEE Transactions on Network and Service Management, vol. 18, no. 1, 2021.

[9] F. Musumeci et al., "An Overview on Application of Machine Learning Techniques in Optical Networks," IEEE Communications Surveys & Tutorials, vol. 21, no. 2, 2019.

[10] M. Wu et al., "A Comprehensive Survey of Blockchain: from Theory to IoT Applications and Beyond," IEEE Internet of Things Journal, 2019.