Degradação de Desempenho de Contramedidas Estáticas a Ataques de Negação de Serviço de Baixo Volume

Bruno M. dos Santos[1], Ian V. Bastos[2], Igor M. Moraes[1]

[1]Laboratório MidiaCom – IC/TCC/PGC
Universidade Federal Fluminense (UFF), Niterói – RJ – Brasil

[2]PEL – Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro, RJ – Brazil

bmsantos@id.uff.br, ian.bastos@eng.uerj.br, igor@ic.uff.br

Abstract. *Intrusion detection systems based on supervised machine learning algorithms face limitations due to their reliance on network traffic distributions observed during the training phase. This article evaluates the robustness of the XGBoost classifier against low-rate denial-of-service attacks with dynamic parameters. The results show that variations in transmission rate, duration, and attack cycle distance real-time traffic from the patterns seen during training, characterizing a concept drift. This phenomenon severely degrades the classifier's detection performance. Experiments demonstrate a significant reduction in recall and F1-score. While both metrics initially remained above 96%, following the concept drift and considering the worst-case scenario, these values declined to 65.99% and 79.34%, respectively. Therefore, static countermeasures fail to generalize to dynamic scenarios, necessitating continuous adaptation mechanisms to preserve classifier efficacy.*

Resumo. *Sistemas de detecção de intrusão baseados em algoritmos de aprendizado de máquina supervisionado enfrentam limitações decorrentes da dependência da distribuição do tráfego de rede observada na fase de treinamento. Este artigo analisa a robustez do classificador XGBoost frente a ataques de negação de serviço de baixo volume com parâmetros dinâmicos. Os resultados mostram que variações na taxa de transmissão, duração e ciclo do ataque distanciam o tráfego real daquele visto em treino, caracterizando uma deriva de conceito. Esse fenômeno degrada severamente o desempenho de detecção do classificador. Os experimentos demonstram uma redução significativa na revocação e na pontuação F1. Enquanto ambas as métricas mantinham-se inicialmente acima de 96%, após a deriva de conceito e considerando o pior cenário, os valores reduziram para 65,99% e 79,34%, respectivamente. Portanto, contramedidas estáticas falham em generalizar para cenários dinâmicos, exigindo mecanismos de adaptação contínua para preservar a eficácia do classificador.*

1. Introdução

Em 2025, o número total de ataques de negação de serviço (*Denial of Service* - DoS) reportado pela Cloudflare mais que dobrou, atingindo 47,1 milhões de incidentes [Yoachimik and Pacheco 2026]. Os ataques DoS têm como objetivo tornar inacessíveis os serviços providos pela vítima a usuários legítimos e, tipicamente, são volumétricos, ou seja, os atacantes buscam exaurir a largura de banda do canal de acesso à vítima

enviando dados a uma taxa mais alta do que a capacidade do canal. Recentemente, uma outra variação de DoS têm sido reportada entre as ameaças mais insidiosas. São os ataques de Negação de Serviço de Baixo Volume (*Low-Rate Denial of Service* - LDoS) [Wu et al. 2020]. Diferentemente dos ataques DoS volumétricos, o LDoS opera de forma furtiva, explorando vulnerabilidades no mecanismo de controle de congestionamento do *Transmission Control Protocol* (TCP). Ao enviar rajadas periódicas de tráfego malicioso seguidas de silêncio, o LDoS consegue reduzir a vazão de fluxos TCP legítimos. Esta estratégia de ataque favorece a ocultação perante contramedidas a ataques DoS baseadas em volume ou limiares estáticos.

Contramedidas aos ataques LDoS baseadas em Aprendizado de Máquina (*Machine Learning* - ML) e Aprendizado Profundo (*Deep Learning* - DL) têm sido propostas, especialmente para arquiteturas de Redes Definidas por Software (*Software-Defined Networks* - SDN), que permitem integrar mecanismos programáveis de detecção e resposta a ataques. Pérez-Díaz *et al.* apresentam uma arquitetura modular em SDN e avaliam diferentes modelos de ML, incluindo um *Multi-Layer Perceptron* (MLP), reportando acurácia em torno de 95% para detecção de ataques *Low-Rate Distributed Denial of Service* (LR-DDoS) em ambiente experimental controlado [Pérez-Díaz et al. 2020]. Contramedidas baseadas em ML, em sua maioria, seguem um fluxo de trabalho supervisionado clássico, no qual um conjunto de dados rotulado é coletado para treinamento *offline*, e o modelo resultante é posteriormente implantado para inferência em tempo real, geralmente no controlador SDN, sem mecanismos explícitos de adaptação *online* [Tang et al. 2022]. Portanto, se baseiam na hipótese de que a distribuição dos dados em operação é compatível com a observada no treinamento [Gama et al. 2014].

Contramedidas baseadas em ML e DL, comumente, validam seus modelos a partir de conjuntos de dados estáticos ou cenários com parâmetros de ataque fixos, como exemplificado em estudos recentes de detecção de LDDoS baseados em DL [Al-Shukaili et al. 2025]. Contudo, atacantes reais são agentes adaptativos e podem alterar dinamicamente parâmetros do ataque LDoS para contornar contramedidas conhecidas. O comportamento dinâmico do atacante induz o fenômeno conhecido na literatura de ML como Deriva de Conceito (*Concept Drift*), o que expõe uma vulnerabilidade crítica para contramedidas estáticas [Wahab 2022]. Modelos treinados em uma distribuição fixa tendem a sofrer degradação acentuada de desempenho quando expostos a variações nos padrões de entrada de dados, tornando-se obsoletos ao longo do tempo.

Este artigo avalia a degradação de contramedidas estáticas a ataques LDoS dinâmicos. O objetivo é quantificar a degradação de desempenho do modelo XGBoost, tradicionalmente eficaz quando submetido a variações paramétricas do ataque LDoS que não estavam presentes na fase de treinamento. A avaliação foi conduzida por meio da simulação de ciclos de ataque, alternando a intensidade e a frequência dos pulsos maliciosos. Os resultados revelam que a adaptabilidade do atacante neutraliza a capacidade preditiva do modelo, culminando em uma redução superior a 30% nas métricas de detecção. As principais contribuições deste artigo são: a sistematização de um cenário de ataque LDoS dinâmico, no qual os parâmetros do ataque variam ao longo do tempo, simulando um atacante adaptativo; uma análise de desempenho demonstrando que modelos com alta acurácia em cenários estáticos falham em manter a mesma alta acurácia em cenários dinâmicos, apresentando reduções superiores a 30% em revocação (*recall*); e a

demonstração da necessidade de transição para contramedidas a ataques LDoS baseadas em aprendizado *online* ou adaptativo.

Este artigo está estruturado da seguinte forma. A Seção 2 apresenta o referencial teórico detalhando conceitos essenciais para a compreensão dos mecanismos do ataque LDoS e da Deriva de Conceito. A Seção 3 aborda os trabalhos relacionados e contextualiza a incapacidade das soluções de contramedidas atuais em lidar com a variação dinâmica de parâmetros de ataque. A Seção 4 descreve o ambiente de avaliação, o cenário de ataque dinâmico e a metodologia de avaliação. A Seção 5 apresenta os resultados e discute a degradação de desempenho observada em modelos estáticos sob o ataque LDoS dinâmico. Por fim, a Seção 6 reúne as conclusões e indica direções para trabalhos futuros.

2. Fundamentação Teórica

Nesta seção, estabelece-se a base conceitual sobre o funcionamento do ataque LDoS e o fenômeno de deriva de conceito em ambientes dinâmicos.

2.1. Ataque LDoS

O ataque LDoS é uma ameaça furtiva que explora os mecanismos de controle de congestionamento e retransmissão do TCP. Diferente de ataques DoS volumétricos, no LDoS, são enviados periodicamente "pulsos" curtos de tráfego em alta taxa, conforme ilustrado na Figura 1. No exemplo, o fluxo TCP legítimo apresenta uma vazão em torno de 30 Mb/s até o instante do início do ataque em 30 s. A introdução dos pulsos programados (áreas sombreadas) força o TCP a interpretar que há um congestionamento na rede, em função da perda de pacotes. Isso faz com que o TCP reduza a janela de congestionamento e, consequentemente, reduza a vazão do fluxo legítimo de forma cíclica, chegando a menos de 10 Mb/s. A eficiência do ataque reside em manter uma taxa média de tráfego de ataque que não seja detectada por contramedidas em uso, enquanto degrada significativamente a vazão de fluxos TCP legítimos.

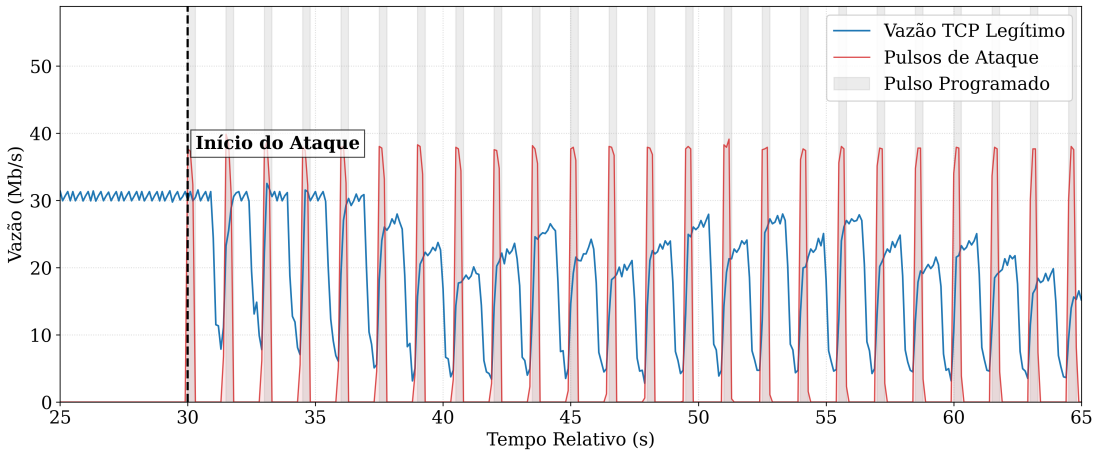

Figura 1. Um exemplo de ataque LDoS com taxa R = 55 Mb/s, período T = 1,5 s e duração do pulso L = 0,3 s.

2.2. Deriva de Conceito

Em ambientes reais de rede, os dados apresentam uma natureza dinâmica. As propriedades estatísticas do tráfego e dos padrões de ataque sofrem alterações ao longo do tempo devido, por exemplo, a flutuações na largura de banda ou adaptações estratégicas dos atacantes. A Figura 2 ilustra a resposta de uma contramedida estática frente a diferentes taxas de ataque. Inicialmente, observa-se um cenário de ataque na área sombreada em azul, com a taxa de transmissão de 60 Mb/s. Nesse estágio, a contramedida identifica a anomalia com precisão e aciona um mecanismo de bloqueio do tráfego malicioso, o que resulta na normalização da vazão TCP próxima de 30 Mb/s e na estabilização das retransmissões. Esse comportamento demonstra que, enquanto o padrão do ataque coincide com o conhecimento prévio do modelo, a contramedida é eficaz em preservar a vazão do fluxo TCP. Entretanto, a transição para um ataque de 30 Mb/s, iniciando na área sombreada em laranja, caracteriza a deriva de conceito. Nesta fase, a redução na taxa de ataque altera sua assinatura estatística, fazendo com que o comportamento malicioso mimetize oscilações típicas da rede. Como o detector estático não reconhece essa variação, o bloqueio não é acionado, permitindo que o ataque degrade a rede de forma persistente.

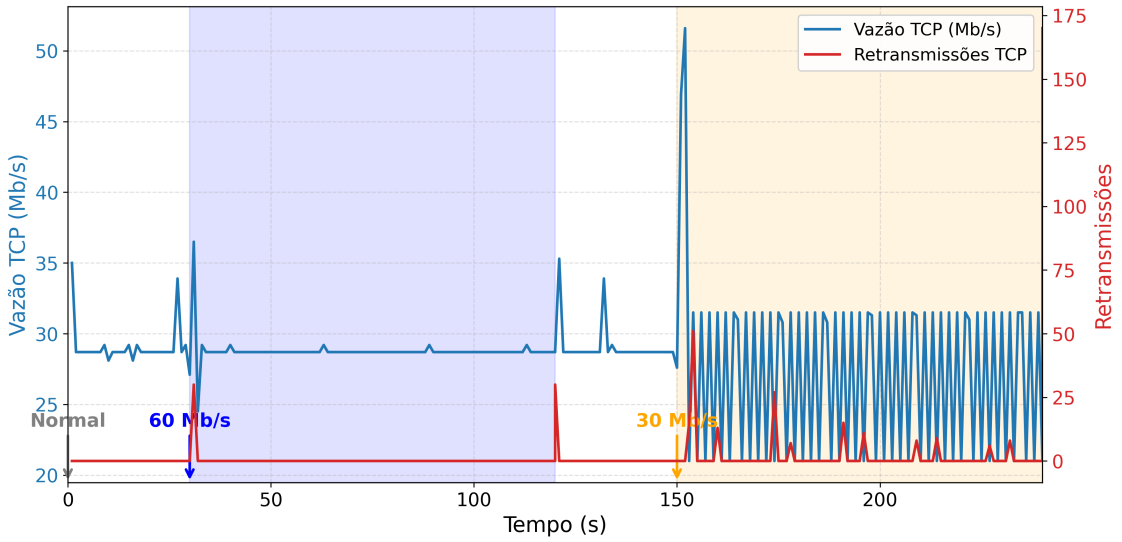

Figura 2. Um exemplo do impacto da deriva de conceito com a variação da taxa de ataque: na área sombreada em azul, a taxa de transmissão de ataque é 60 Mb/s e o ataque não é bem sucedido por conta da contramedida estática em uso; na área sombreada em laranja, a taxa de transmissão de ataque é 30 Mb/s e o ataque é bem sucedido porque a contramedida estática não acompanha a redução da taxa de ataque.

A compreensão desses conceitos é fundamental para o desenvolvimento de contramedidas que sejam resilientes a mudanças contextuais, evitando a degradação da rede quando as condições operacionais divergem do conjunto de treinamento original.

3. Trabalhos Relacionados

Esta seção analisa o estado da arte das contramedidas a ataques LDoS baseadas em ML [Kuzmanovic and Knightly 2003] e discute as limitações inerentes aos modelos estáticos contra cenários dinâmicos e deriva de conceito.

3.1. Detecção de LDoS Baseada em Aprendizado de Máquina e Aprendizado Profundo

A literatura consolidada demonstra que algoritmos de ML supervisionados superam as abordagens estáticas tradicionais na detecção de ataques LDoS. Wu *et al.* destacam em sua pesquisa que a complexidade e a furtividade do tráfego LDoS exigem a extração de características capazes de capturar a periodicidade e a curta duração dos pulsos de ataque, padrões que métodos lineares frequentemente falham em identificar [Wu et al. 2020]. Corroborando essa visão, Rios *et al*. revisam sistematicamente técnicas de mitigação em diferentes camadas e evidenciam o predomínio de classificadores como *Support Vector Machine* (SVM), *Random Forest* e Redes Neurais [Rios et al. 2022]. A análise do *survey* indica que tais modelos são, em sua maioria, treinados *offline* a partir de cenários experimentais com parâmetros de ataque previamente definidos, refletindo uma prática recorrente na literatura.

No que tange à aplicação de técnicas de *Deep Learning* no contexto de computação em nuvem, Fu *et al.* propõem um método de detecção baseado em características de tempo–frequência, no qual o modelo aprende a reconstruir sequências de tráfego normal em um espaço latente [Fu et al. 2022]. A relevância deste trabalho reside na demonstração de que, mesmo em abordagens complexas, a validação experimental tende a assumir condições estáticas próximas às de treinamento, sem discutir explicitamente a degradação de desempenho sob variações paramétricas do atacante.

Em SDN, trabalhos recentes têm buscado integrar detecção precisa, mitigação granular e rastreabilidade do atacante. Ma *et al.* introduzem o sistema BDTM (*Bidirectional Detection and Traceability Mitigation*), que inova ao correlacionar informações dos planos de controle e de dados, permitindo não apenas a detecção de ataques LDoS, mas também a localização progressiva da origem do ataque e a aplicação de isolamento de tráfego baseado no fechamento de portas dos comutadores SDN [Ma et al. 2025]. Paralelamente, Liu *et al.* propõem o *framework* ERT-EDR, cujo enfoque é na proteção do protocolo TCP via *Extended Random Trees*. Embora classificado como uma contramedida *online* devido à sua velocidade de resposta, o ERT-EDR permanece limitado por uma natureza estática, uma vez que a contramedida não prevê mecanismos de atualização contínua de seus parâmetros frente a novos fluxos de dados. Assim, o ERT-EDR exemplifica o estado da arte ao combinar eficiência computacional com alta acurácia na distinção de padrões em ambientes programáveis [Liu et al. 2024].

Um marco recente e fundamental para o presente trabalho é o esquema Trident [Tang et al. 2025] que combina monitoramento de estado de porta, estatísticas de fluxo e classificação via ML no controlador SDN, reportando resultados acima de 98% nos cenários experimentais definidos. Esse estudo é central para a pesquisa deste artigo por dois motivos: (i) define a metodologia de coleta e rotulação de tráfego LDoS aqui reproduzida como base comparativa; e (ii) representa o exemplo clássico de um classificador robusto treinado a partir de um conjunto finito de cenários, sem investigar o impacto da variação dinâmica dos parâmetros de ataque, além dessas condições pré-estabelecidas.

3.2. Limitações de Modelos Estáticos e Deriva de Conceito

A deriva de conceito é uma limitação de modelos de aprendizado de máquina empregados em ambientes operacionais não estacionários, nos quais a relação entre atributos e rótulos

pode se alterar ao longo do tempo [Gama et al. 2014]. No contexto mais amplo de mineração de dados e aprendizado contínuo, Lu *et al.* sistematizam o problema ao discutir as principais causas de deriva de conceito, suas manifestações e as famílias de estratégias para mitigação, incluindo sua detecção explícita, adaptação incremental, uso de janelas deslizantes e mecanismos de retreinamento seletivo [Lu et al. 2019]. Essa base é particularmente pertinente para segurança, na qual o tráfego e o comportamento adversarial evoluem de forma contínua e intencional.

Em Sistemas de Detecção de Intrusão (*Intrusion Detection Systems* - IDSes), essa limitação se torna ainda mais evidente. Shyaa *et al.* apresentam uma pesquisa abrangente sobre deriva de conceito e dinâmica de atributos em IDSes, discutindo como alterações na distribuição do tráfego e dos ataques degradam classificadores estáticos e como abordagens cientes da deriva de conceito (*drift-aware*) podem combinar seleção dinâmica de atributos, modelos adaptativos e mecanismos de detecção de deriva de conceito para manter desempenho ao longo do tempo [Shyaa et al. 2024]. Em conjunto, esses resultados sustentam que a degradação sob mudança de conceito não é um efeito colateral pontual, mas um desafio estrutural em detecção baseada em ML.

A literatura recente também ilustra arquiteturas práticas para adaptação contínua de contramedidas de detecção em cenários de segurança. Camarda *et al.* propõem um IDS *online* baseado em *Random Forest* incremental, acoplado a um módulo de aprendizado ativo e a um detector de deriva de conceito, armazenando janelas de dados recentes para atualizações incrementais e utilizando sinais de mudança para disparar retreinamentos direcionados [Camarda et al. 2025]. Na mesma linha, Leveni e Boracchi exploram o uso de *Isolation Forest* não apenas como detector de anomalias, mas como componente central de um *framework* de detecção e resposta à deriva de conceito em dados de segurança de rede, com estratégias explícitas de retreinamento quando um desvio estatístico é identificado [Leveni and Boracchi 2025].

Esses trabalhos são relevantes porque demonstram, em termos conceituais e arquiteturais, como mecanismos de detecção/adaptação à deriva de conceito podem ser acoplados a classificadores para reduzir a degradação da operação de contramedidas. Entretanto, apesar do amadurecimento do debate sobre o problema de deriva de conceito em IDS e em segurança de redes, a aplicação sistemática de mecanismos de aprendizado adaptativo para o cenário específico de ataques LDoS em SDN — nos quais variações nos parâmetros do ataque podem induzir mudanças sutis, porém críticas — ainda aparece como uma lacuna de investigação.

4. Metodologia

Para avaliar o impacto no desempenho de contramedidas estáticas ao ataque LDoS em SDN frente a variações dinâmicas, adota-se uma abordagem via emulação. A metodologia consiste na reprodução do cenário de referência e na implementação da contramedida Trident [Tang et al. 2025], utilizando o algoritmo XGBoost baseado em *Gradient Boosting* [Chen and Guestrin 2016], sob condições de estresse paramétrico. O modelo é configurado para priorizar alta precisão e reduzir a incidência de falsos positivos, condição necessária para viabilidade operacional em sistemas de detecção em tempo real.

4.1. Ambiente de Emulação e Topologia

O ambiente de avaliação é construído sobre a plataforma Mininet, gerenciada pelo controlador Ryu (*OpenFlow* 1.3). A topologia de rede implementada, ilustrada na Figura 3, segue um modelo linear composto por três comutadores (S1 a S3) e quatro estações (H1 a H4), definida para ter um enlace de gargalo (*bottleneck*) controlável.

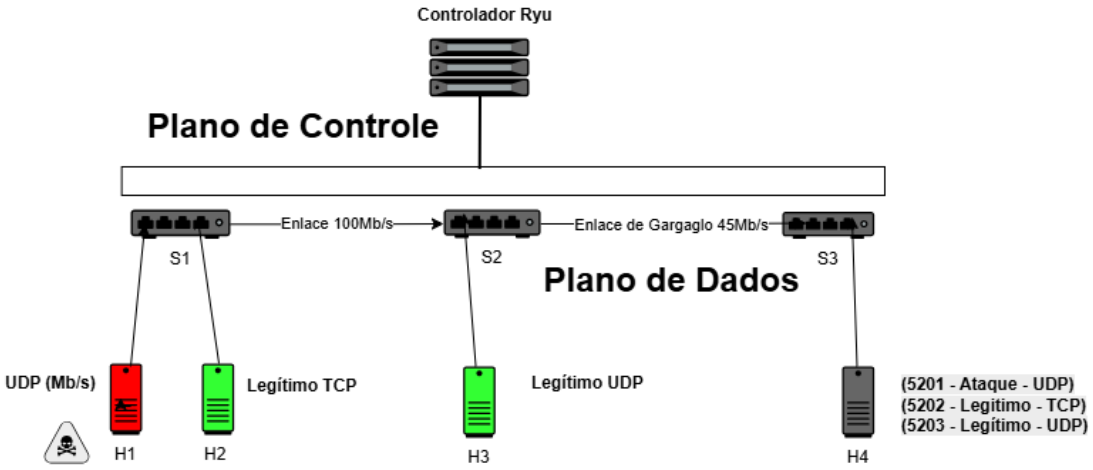

Figura 3. A Topologia de Rede do Ambiente de Avaliação.

Os enlaces de acesso e interconexão entre S1-S2 possuem largura de banda de 100 Mb/s. O enlace de gargalo, localizado entre S2-S3, está configurado com capacidade reduzida de 45 Mb/s, simulando o ponto de saturação alvo do ataque. A injeção e orquestração dos fluxos de dados foram realizadas utilizando a ferramenta Iperf3 para a geração do tráfego legítimo, enquanto a geração do ataque é executada via *sockets* Python, ambos controlados por *scripts* para modulação das rajadas conforme a distribuição detalhada a seguir:

- Tráfego Legítimo de Fundo: gerado pelas Estações H2 e H3 em direção ao Servidor H4. A Estação H2 envia um fluxo TCP (Porta 5202), enquanto a Estação H3 envia um fluxo UDP (Porta 5203). A soma do tráfego legítimo (35 Mb/s) ocupa aproximadamente 77% da capacidade do enlace de gargalo.
- Tráfego de Ataque: A Estação H1 atua como o atacante, enviando rajadas UDP (Porta 5201) em direção ao Servidor H4, visando exaurir a banda do enlace de gargalo e causar descarte de pacotes de fluxos legítimos TCP.

4.2. Modelagem do Ataque e Geração do Conjunto de Dados

O ataque LDoS é modelado matematicamente como uma função periódica de pulso. O fluxo de ataque $F_{atk}(t)$ é caracterizado por três parâmetros principais: taxa de ataque (R), período (T) e duração do pulso (L). Para avaliar a deriva de conceito no XGBoost, são conduzidos experimentos variando esses três parâmetros em relação à capacidade ociosa do enlace.

Diferente das contramedidas estáticas que treinam e testam o modelo com características idênticas ou parecidas, este estudo segrega os dados em dois cenários distintos. No primeiro cenário, o modelo é treinado com um conjunto de dados contendo 4.961 amostras, das quais 2.893 (58,31%) pertencem à classe normal e 2.068 (41,69%) à classe

ataque, resultando em um conjunto balanceado. Esse cenário envolve tráfego normal e de ataques LDoS de 45 e 55 Mb/s, nos quais a assinatura do ataque se apresenta de forma evidente. A fim de controlar o tráfego, é executado o algoritmo *Token Bucket Filter* (TBF) pela estação de origem, restringindo o tráfego TCP legítimo a uma taxa média de 30 Mb/s. No segundo cenário, o conjunto de testes é submetido a taxas de ataques próximas às de treinamento, com valores de 50 e 60 Mb/s. Essas taxas de ataque, bem como a duração dos pulsos, seguem os parâmetros utilizados na contramedida Trident [Tang et al. 2025]. Adicionalmente, foram incluídas as taxas de ataque de 25 e 30 Mb/s, sob diferentes combinações de período (T) e largura de pulso (L). O principal objetivo é simular um atacante que altera sua assinatura durante a execução do ataque para se distanciar do padrão aprendido pela contramedida na fase de treinamento. Dessa forma, avalia-se a capacidade do modelo treinado com parâmetros fixos de generalizar a detecção para configurações distantes do treinamento. Cada execução do experimento tem duração total de 8 minutos, estruturada em um intervalo inicial de 240 segundos de tráfego legítimo, seguido por dois intervalos de ataque LDoS de 120 segundos cada, iniciados em $t_1 = 240$ e $t_2 = 360$, respectivamente. A Tabela 1 resume os parâmetros utilizados na geração dos fluxos de tráfego.

Tabela 1. Parâmetros de Configuração do Experimento.

Parâmetros	**Configurações**
Duração do Experimento	8 minutos por cenário
Cenários de Ataque	Em $t_1 \rightarrow R$ = 60 Mb/s, T = 1,5 s, L = 0,2 s Em $t_2 \rightarrow R$ = 60 Mb/s, T = 2,5 s, L = 0,2 s
	Em $t_1 \rightarrow R$ = 50 Mb/s, T = 1,5 s, L = 0,4 s Em $t_2 \rightarrow R$ = 50 Mb/s, T = 2,5 s, L = 0,4 s
	Em $t_1 \rightarrow R$ = 30 Mb/s, T = 1,0 s, L = 0,4 s Em $t_2 \rightarrow R$ = 30 Mb/s, T = 1,4 s, L = 0,5 s
	Em $t_1 \rightarrow R$ = 25 Mb/s, T = 1,2 s, L = 0,5 s Em $t_2 \rightarrow R$ = 25 Mb/s, T = 1,5 s, L = 0,6 s
Carga Legítima TCP	30 Mb/s (H2 – H4)
Carga Legítima UDP	5 Mb/s (H3 – H4)
Largura de banda do Gargalo	45 Mb/s

4.3. Validação Estatística dos Resultados

Cada cenário de teste é submetido a 30 execuções independentes. A confiabilidade das métricas é aferida através de intervalos de confiança de 99%. Essa metodologia assegura a representatividade do comportamento médio da contramedida perante à natureza estocástica da rede, sendo crucial para distinguir degradações reais de flutuações aleatórias.

5. Análise de Desempenho

Nesta seção, avalia-se a validação do impacto do ataque LDoS na infraestrutura de rede e a quantificação da perda de desempenho do XGBoost em face de variações dinâmicas de intensidade e duração dos pulsos de ataque, como consequência da deriva de conceito.

5.1. Validação do Impacto em Fluxos Legítimos TCP

Antes de analisar a capacidade de detecção da contramedida implementada, é imperativo demonstrar que os ataques LDoS utilizados nos cenários de teste reduzem, de fato, a vazão de fluxos TCP legítimos. Uma crítica recorrente na literatura é a dificuldade das contramedidas estáticas em distinguir ataques furtivos de flutuações normais do tráfego de rede, frequentemente classificando o tráfego de baixa taxa de transmissão como ruído [Kuzmanovic and Knightly 2003]. Para quantificar o impacto dessas variações paramétricas do ataque, são analisadas duas métricas sob diferentes taxas de ataque (25, 30 e 50 Mb/s): o tamanho da janela de congestionamento do TCP (Cwnd), uma vez que o controle de congestionamento do TCP é o mecanismo diretamente explorado pelo ataque LDoS, e a vazão efetiva percebida pelo cliente, que reflete o efeito do ataque sobre a vazão de um fluxo legítimo TCP. As Figuras 4 e 5 mostram o impacto dos ataques nas intensidades em questão.

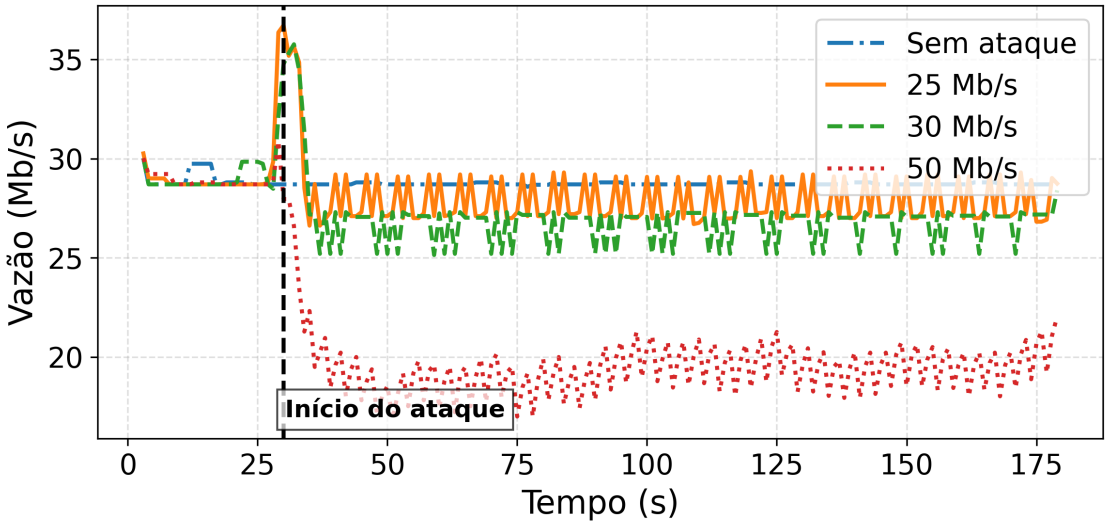

Figura 4. Impacto do Ataque LDoS na Vazão do Fluxo Legítimo TCP.

A análise da vazão revela que, no instante $t = 30\ s$, início do ataque, ocorre uma degradação imediata desta métrica, associada à reação do mecanismo de controle de congestionamento do TCP. No cenário de alta intensidade (50 Mb/s), observa-se uma redução abrupta e persistente da vazão para um patamar significativamente inferior ao inicial, em torno de 20 Mb/s, indicando a manutenção do protocolo em regime de congestionamento contínuo. No cenário de média intensidade (30 Mb/s), a vazão sofre quedas periódicas e se estabiliza em torno de 26–27 Mb/s, apresentando oscilações sincronizadas com os pulsos do ataque. Já no cenário de menor intensidade analisado (25 Mb/s), a vazão mantém-se próxima de 27–28 Mb/s, porém com flutuações frequentes, evidenciando a interferência recorrente do ataque sobre a vazão do fluxo TCP legítimo. Contudo, o impacto mais insidioso é observado na análise da dinâmica da janela de congestionamento TCP.

- **Cenário de 50 Mb/s:** Observa-se uma redução drástica e sustentada da janela de congestionamento, que passa a oscilar em valores significativamente baixos ao longo de toda a duração do ataque. Esse comportamento indica que o TCP entra repetidamente em fases de recuperação e partida lenta, sem conseguir restabelecer sua janela original, mesmo na ausência de saturação contínua do enlace.

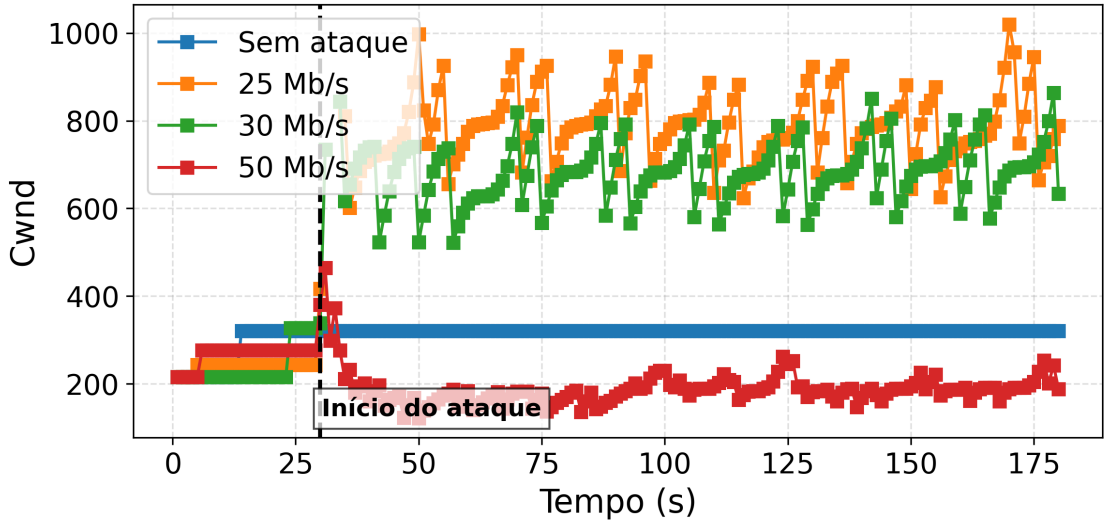

Figura 5. Impacto do Ataque LDoS na Janela de Congestionamento do Fluxo Legítimo TCP.

- **Cenário de 30 Mb/s:** Observa-se oscilações frequentes da janela de congestionamento, com quedas abruptas sincronizadas aos pulsos do ataque. Esse comportamento impede a estabilização em valores elevados, resultando em variações constantes da vazão e degradação da vazão do fluxo TCP legítimo. Tal característica é crítica, pois mantém a conexão ativa, porém em um regime instável, dificultando a detecção por mecanismos tradicionais baseados exclusivamente em altos volumes de tráfego.
- **Cenário de 25 Mb/s:** Apresenta um comportamento intermediário, no qual a janela de congestionamento sofre reduções periódicas seguidas de tentativas de recuperação. Apesar dessas recuperações parciais, a janela não retorna de forma estável aos níveis observados antes do ataque, mantendo-se em um regime oscilatório. Esse padrão evidencia que o ataque LDoS consegue manter o TCP em um estado permanente de contenção, limitando a utilização eficiente da largura de banda disponível, ainda que a vazão média aparente permaneça relativamente elevada.

Esses resultados confirmam que, mesmo quando a taxa de ataque não provoca a saturação total do enlace, a manipulação sistemática da janela de congestionamento força o fluxo TCP legítimo a operar em regimes ineficientes de transmissão, validando a eficácia do ataque LDoS em reduzir a vazão deste fluxo por meio da exploração do mecanismo de controle de congestionamento do TCP.

5.2. Análise Comparativa da Degradação de Detecção

O próximo passo é avaliar a resposta da contramedida implementada ao ataque LDoS. Esta seção detalha como a alteração dos parâmetros do ataque afeta o modelo XGBoost quando opera sob a restrição de treinamento estático. A Figura 6 apresenta o comparativo de desempenho das seguintes métricas: Acurácia, Precisão, Revocação e Pontuação F1 para os quatro cenários de ataque definidos na Tabela 1.

Conforme pode ser observado, o modelo mantém altas métricas de desempenho

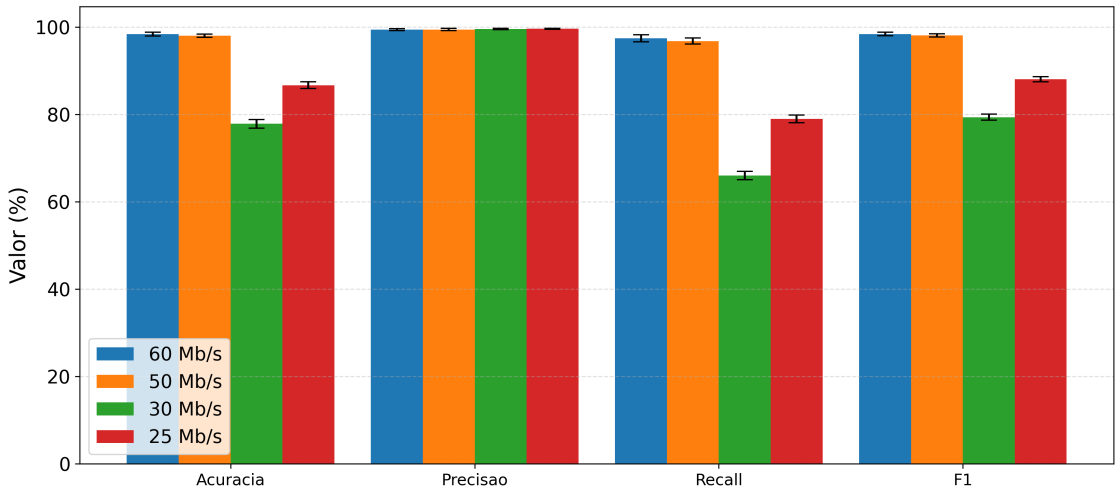

Figura 6. Métricas de desempenho do XGBoost para os diferentes cenários de ataques LDoS (Tabela 1) com variação dos parâmetros período T e largura de pulso L durante o ataque.

quando os parâmetros de ataque são próximos aos parâmetro em que ele foi treinado (45 e 55 Mb/s). Porém, sofre uma queda acentuada em suas métricas quando submetido a padrões diferentes do que foi visto, como nos cenários de 25 e 30 Mb/s, nos quais a revocação cai para 78,91% e 65,99%, respectivamente. Isso significa que, aproximadamente, entre 21% a 34% do tráfego malicioso furtivo é ignorado pelo modelo. O resultado é crítico para a conclusão do estudo, pois demonstra que a natureza estática do treinamento é um fator determinante para a falha na detecção de ataques LDoS dinâmicos. O modelo XGBoost tem suas métricas de desempenho reduzidas se não possuir mecanismo de adaptação contínua.

5.3. Discussão

Os resultados obtidos confirmam a hipótese central deste estudo: contramedidas de alto desempenho, quando baseadas na premissa de estacionariedade — isto é, treinados sob um conjunto fixo de padrões — tornam-se vulneráveis a adversários que exploram a dinâmica dos parâmetros do ataque LDoS. A análise a seguir discute as implicações teóricas e práticas da degradação na taxa de detecção, interpretando a deriva de conceito sob a ótica da segurança adaptativa.

A degradação acentuada de desempenho do XGBoost nos cenários de ataque com taxas de 25 e 30 Mb/s, com valores de revocação atingindo 65,99%, ilustra o problema de generalização sob mudança de distribuição. Os resultados obtidos apresentam intervalos de 99% estreitos e estáveis, mesmo nos regimes de pior desempenho. Esse comportamento indica que a perda de eficácia não está associada ao aumento da variabilidade estatística ou à instabilidade entre execuções, mas sim a um viés consistente do modelo. Em termos práticos, o modelo passa a operar de maneira confiante em regiões do espaço de características que não foram adequadamente representadas durante o treinamento, reforçando a hipótese de que o modelo supervisionado aprendeu padrões fortemente associados a rajadas evidentes e contrastes bem definidos entre tráfego normal e malicioso.

Ao alterar a taxa e a duração do ataque, as amostras maliciosas deslocam-se para

regiões do espaço de características que não foram mapeadas durante a fase de treinamento. Nessa condição, o vetor de características do ataque dinâmico cruza a fronteira de decisão fixa por não corresponder aos padrões previamente aprendidos, sendo interpretado erroneamente como tráfego normal. Embora a redução do limiar de detecção possa aumentar a revocação, a flexibilização do critério de decisão eleva a taxa de falsos positivos, acarretando aumento do custo operacional e redução de confiança na contramedida. Esse resultado demonstra que a limitação observada não decorre de um problema de calibração do modelo, mas de representatividade: uma contramedida de decisão estática é incapaz de distinguir classes cuja separação varia dinamicamente ao longo do tempo.

Em um cenário operacional real, um atacante adaptativo pode empregar estratégias de sondagem, reduzindo gradualmente a taxa de ataque e variando ciclos de pulsos, até cessarem os bloqueios de uma contramedida estática, mantendo-se sistematicamente abaixo do limiar de detecção enquanto continua a exaurir recursos da rede. Os resultados apresentados indicam que esse comportamento é viável em contramedidas baseadas em modelos *offline*, uma vez que a fronteira de decisão permanece fixa e não reage à evolução do padrão do ataque. Nesse contexto, a contramedida avaliada neste artigo, mostra-se inadequada para proteção de longo prazo, pois exige intervenção humana recorrente para retreinamento e reimplantação do modelo. Essa rotina de manutenção manual, além de custosa, ocorre em uma escala temporal incompatível com a velocidade de adaptação dos ataques modernos, reforçando a necessidade de contramedidas capazes de aprenderem e de se ajustarem continuamente em operação.

6. Conclusão e Trabalhos Futuros

Este trabalho investigou a fragilidade estrutural do modelo XGBoost baseado em treinamento estático frente a ataques LDoS dinâmicos. Diferente da literatura predominante, que foca na maximização de métricas em cenários controlados, a avaliação deste artigo priorizou a análise de vulnerabilidade decorrente da deriva de conceito induzida pela variabilidade dos ataques durante sua execução.

Os resultados obtidos permitem traçar duas conclusões principais. Primeiramente, validou-se que, neste cenário implementado, ataques LDoS de baixa intensidade (25–30 Mb/s), com variações de duração e ciclo, degradam o tráfego legítimo TCP da rede, reduzindo a vazão legítima em aproximadamente 20%. Em segundo lugar, demonstrou-se que o modelo supervisionado XGBoost, quando treinado estaticamente, sofre degradação de desempenho em cenários diferentes do que aprendeu, com a revocação caindo para níveis de aproximadamente 65%.

Conclui-se, portanto, que a premissa de estacionariedade adotada nas contramedidas atuais representa uma vulnerabilidade crítica. A memória congelada das contramedidas torna-se obsoleta assim que o adversário altera os parâmetros do ataque, exigindo uma mudança de paradigma na defesa.

Como trabalhos futuros, pretende-se investigar a aplicação de técnicas de aprendizado *online* para a mitigação de LDoS. O objetivo é avaliar uma contramedida adaptativa capaz de atualizar seu modelo de decisão em tempo real. Busca-se, com isso, superar a rigidez das contramedidas estáticas, permitindo a detecção contínua de ataques que evoluem dinamicamente ao longo do tempo.

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq, CAPES, FAPERJ e RNP e faz parte do INCT de Redes de Comunicação e Internet das Coisas Inteligentes (ICoNIoT), financiado pelo CNPq (405940/2022-0) e pela CAPES (88887.954253/2024-00). Os autores usaram as ferramentas de inteligência artificial generativa ChatGPT e Gemini exclusivamente para correção ortográfica e gramatical do texto e auxílio técnico à codificação de *scripts*, não substituindo a autoria ou a análise crítica dos resultados pelos autores.

Disponibilidade de Artefatos

Alinhado aos princípios e práticas de Ciência Aberta, o conjunto de dados e o código-fonte utilizado nesta pesquisa estão disponíveis publicamente para fins de reprodutibilidade através do link: `https://github.com/brumesan/LDoS-SDN-2026`.

Referências

Al-Shukaili, N. A., Kiah, M. L. M., and Ahmedy, I. (2025). Optimizing feature selection and deep learning techniques for precise detection of low-rate distributed denial of service (LDDoS) attack. *Discover Internet of Things*, 5(80).

Camarda, S., Musumeci, F., and Torre, G. (2025). Managing concept drift in online intrusion detection systems. *Joint National Conference on Cybersecurity*.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, New York, NY, USA. ACM.

Fu, Z., Li, M., Wu, Y., and Liu, Q. (2022). Low-rate denial of service attack detection method based on time-frequency characteristics. *Journal of Cloud Computing*, 11.

Gama, J., Žliobait schoole, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, 46(4):1–37.

Kuzmanovic, A. and Knightly, E. W. (2003). Low-rate tcp-targeted denial of service attacks: the shrew vs. the mice and elephants. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 75–86.

Leveni, F. and Boracchi, G. (2025). Online isolation forest. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pages 26858–26876. PMLR.

Liu, B., Tang, D., Chen, J., and Liang, W. (2024). ERT-EDR: Online defense framework for TCP-targeted LDoS attacks in SDN. *Expert Systems with Applications*, 254:124356.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., and Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, pages 2346 – 2363.

Ma, X., Li, X., He, Y., Qi, Q., and Li, H. (2025). BDTM: Bidirectional detection and traceability mitigation of LDoS attacks in SDN. *IEEE Transactions on Network and Service Management*, 20:6826 – 6839.

Pérez-Díaz, J. A., Valdovinos, I. A., Choo, K.-K. R., and Zhu, D. (2020). A flexible SDN-based architecture for identifying and mitigating low-rate DDoS attacks using machine learning. *IEEE Access*, 8:155859–155872.

Rios, V. M., Inácio, P. R. M., and Maimó, D. (2022). Detection and mitigation of low-rate denial-of-service attacks: A survey. *IEEE Access*, 10:76648–76668.

Shyaa, W., Gharaibeh, H., and Rawashdeh, M. (2024). Evolving cybersecurity frontiers: A comprehensive survey on concept drift and feature dynamics-aware machine learning and deep learning in intrusion detection systems. *Engineering Applications of Artificial Intelligence*.

Tang, D., Yan, Y., Zhang, S., Chen, J., and Qin, Z. (2022). Performance and features: Mitigating the low-rate TCP-targeted DoS attack via SDN. *IEEE Journal on Selected Areas in Communications*, 40:428 – 444.

Tang, D., Yan, Y., Zhang, S., Chen, J., and Qin, Z. (2025). Trident: A low-rate DoS attack mitigation scheme based on port and traffic state in SDN. *IEEE Transactions on Computers*, 74:1758–1770.

Wahab, O. A. (2022). Intrusion detection in the IoT under data and concept drifts: Online deep learning approach. *IEEE Internet of Things Journal*, 9(20):19706 – 19716.

Wu, Z., Zhang, L., and Yue, M. (2020). Low-rate DoS attacks, detection, defense, and challenges: A survey. *IEEE Access*, 8:43920–43943.

Yoachimik, O. and Pacheco, J. (2026). 2025 q4 DDoS threat report: A record-setting 31.4 Tbps attack caps a year of massive DDoS assaults. Technical report, Cloudflare. Disponível em `https://blog.cloudflare.com/ddos-threat-report-2025-q4/`.