SBSeg 2026 Armação dos Búzios - RJ
VEJA OS SELOS DE ARTEFATOS

Sentry: Uma Contramedida Adaptativa contra 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 – Brasil

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

Abstract. *Static machine learning algorithms render intrusion detection systems ineffective in dynamic environments, as these systems remain constrained to the data patterns observed during the training phase. This paper proposes Sentry, an adaptive countermeasure based on the Online Isolation Forest algorithm, aimed at identifying low-rate denial-of-service attacks in software-defined networks. Sentry performs continuous updates and incorporates selective criteria conditioned on the network state, preventing improper learning during attack periods and reducing the risk of model contamination. The experimental evaluation demonstrates that Sentry outperforms both the Online Isolation Forest with unrestricted updates and XGBoost. Sentry achieves recall and F1-score metrics of up to 98.89% and 98.53%, respectively. In contrast, XGBoost and Online Isolation Forest exhibit a sharp degradation in performance, with recall dropping to 17.95% and 10.50%, respectively.*

Resumo. *Algoritmos estáticos de aprendizado de máquina tornam os sistemas de detecção de intrusão ineficazes em ambientes dinâmicos, uma vez que esses sistemas permanecem restritos aos padrões de dados observados durante a fase de treinamento. Este artigo propõe o Sentry, uma contramedida adaptativa baseada no algoritmo Online Isolation Forest, voltada à identificação de ataques de negação de serviço de baixo volume em redes definidas por software. O Sentry realiza atualizações contínuas e incorpora critérios seletivos condicionados ao estado da rede, evitando aprendizado inadequado durante períodos de ataque e reduzindo o risco de contaminação do modelo. A avaliação experimental demonstra que o Sentry supera tanto o Online Isolation Forest com atualizações irrestritas quanto o XGBoost. O Sentry atinge métricas de revocação e pontuação F1 de até 98,89% e 98,53%, respectivamente. Em contrapartida, o XGBoost e o Online Isolation Forest exibem uma queda acentuada no desempenho, com a revocação caindo para 17,95% e 10,50%, respectivamente.*

1. Introdução

O cenário de cibersegurança em 2025 registrou um crescimento acentuado na agressividade das ameaças, com ataques de Negação de Serviço (*Denial-of-Service* – DoS) ultrapassando 47 milhões de ocorrências, representando um aumento superior a 100% em

relação ao período anterior [Yoachimik and Pacheco 2026]. Embora a maioria dos ataques DoS seja volumétrica e tenha como objetivo saturar a largura de banda, os ataques de Negação de Serviço de Baixo Volume (*Low-Rate Denial-of-Service* – LDoS) enviam rajadas periódicas de alta taxa para forçar o *Transmission Control Protocol* (TCP) a interpretar a perda de pacotes como congestionamento. Ao explorar o mecanismo de controle de congestionamento, o LDoS induz a redução cíclica da janela de congestionamento, degradando a eficiência de fluxos TCP legítimos [Rios et al. 2022]. Assim, os ataques LDoS contornam contramedidas baseadas exclusivamente na taxa de transferência ou no volume de dados, permanecendo furtivos.

As Redes Definidas por Software (*Software-Defined Networking* – SDN) impulsionaram o desenvolvimento de contramedidas para ataques LDoS baseadas em aprendizado de máquina (*Machine Learning* – ML) e aprendizado profundo (*Deep Learning* – DL). A SDN representa uma alternativa de defesa contra ataques LDoS, pois desacopla a lógica de controle do hardware de encaminhamento, fornecendo visibilidade centralizada da rede e programabilidade granular [Alashhab et al. 2023]. Pérez-Díaz *et al.*, por exemplo, propõem uma estrutura modular de detecção e prevenção de intrusão em SDN projetada para desacoplar os processos de identificação e mitigação da aplicação de rede. Os autores demonstram que modelos estáticos, como o *Multi-Layer Perceptron* (MLP), podem alcançar elevada acurácia em ambientes controlados [Pérez-Díaz et al. 2020]. Entretanto, a eficácia de contramedidas baseadas em modelos estáticos frequentemente depende de treinamento *offline* com conjuntos de dados rotulados, o que limita a capacidade de resposta dessas contramedidas a padrões de tráfego que divergem substancialmente do cenário inicial de treinamento [Tang et al. 2022]. De fato, contramedidas estáticas assumem que a distribuição dos dados durante a operação permanece consistente com aquela observada durante o treinamento [Gama et al. 2014].

Contramedidas adaptativas buscam superar as limitações das abordagens estáticas por meio de aprendizado por reforço [Pandiyakumari S and Suganya R 2025], ajuste de pesos de modelos [Rong et al. 2024] e refinamento de métricas estatísticas [Mathew and Vidhate 2024]. Entretanto, o comportamento dinâmico dos atacantes induz o fenômeno de deriva de conceito (*Concept Drift*), caracterizado por alterações nas propriedades estatísticas do tráfego que degradam a precisão de contramedidas que não acompanham tais mudanças [Wahab 2022]. Além disso, modelos com atualização irrestrita enfrentam o risco de envenenamento, assimilando padrões maliciosos como tráfego normal [Pérez-Díaz et al. 2020].

Em resposta às limitações dos modelos estáticos, este artigo propõe o Sentry, uma contramedida adaptativa baseada no algoritmo *Online Isolation Forest* (OIF) para identificar ataques LDoS em ambiente SDN. O Sentry é projetado para se adaptar continuamente à evolução do tráfego de rede. No entanto, algoritmos adaptativos irrestritos como o OIF enfrentam um desafio crítico: o risco de envenenamento do modelo, no qual padrões maliciosos são gradualmente considerados como comportamento legítimo. O objetivo é demonstrar a capacidade do Sentry de detectar variações dinâmicas de ataques LDoS em cenários de tráfego de rede não estacionários, mantendo a integridade de seu processo de aprendizado ao prevenir o envenenamento do modelo por fluxos maliciosos. Para isso, o processo de atualização do OIF é modificado por meio de um mecanismo de aprendizado seletivo que restringe as atualizações a condições de rede estáveis e legítimas, com base

na validação do comportamento do tráfego e em critérios de consistência temporal.

A validação do Sentry baseia-se em uma análise comparativa em relação ao XGBoost e ao OIF. O experimento simula um ataque LDoS com diferentes ciclos, intensidades e frequências de pulsos maliciosos. Os resultados revelam que o Sentry mantém elevados níveis de revocação e pontuação F1, alcançando valores de até 98,89% e 98,53%, respectivamente. Por outro lado, observa-se a vulnerabilidade tanto do XGBoost quanto do OIF. O XGBoost apresenta forte degradação de desempenho sob cenários de deriva de conceito, atingindo mínimos de 17,95% para revocação e 30,34% para pontuação F1. Enquanto o OIF degrada para valores tão baixos quanto 10,50% de revocação e 17,87% de pontuação F1 devido à contaminação do modelo. As principais contribuições deste artigo são: a implementação de uma contramedida adaptativa seletiva contra ataques LDoS em SDN; o desenvolvimento de um mecanismo de filtragem que impede o autoenvenenamento da contramedida por fluxos de ataque LDoS; e uma análise comparativa demonstrando a resiliência do Sentry frente a ataques LDoS com parâmetros dinâmicos em comparação ao OIF e ao XGBoost.

O restante deste artigo está estruturado da seguinte forma. A Seção 2 discute trabalhos relacionados e contextualiza a incapacidade das contramedidas atuais de lidar com variações dinâmicas de ataque. A Seção 3 detalha a visão geral do Sentry. A Seção 4 descreve o ambiente de avaliação e o cenário de ataque. A Seção 5 apresenta os resultados e discute a comparação de desempenho observada entre Sentry, OIF e XGBoost. Finalmente, a Seção 6 resume as conclusões e delineia direções para trabalhos futuros.

2. Trabalhos Relacionados

Esta seção analisa as contramedidas identificadas na literatura em (i) estudos que utilizam algoritmos clássicos de ML e DL para detectar ataques LDoS e (ii) estudos que adotam abordagens adaptativas para lidar com a deriva de conceito e melhorar o desempenho em ambientes dinâmicos.

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

A literatura mostra uma convergência em direção ao uso de algoritmos de ML e DL para superar as limitações dos métodos estatísticos tradicionais na detecção de ataques LDoS. Rios *et al*. revisam o predomínio de classificadores como *Support Vector Machine* (SVM) e Random Forest, destacando sua eficácia na captura da periodicidade dos ataques [Rios et al. 2022]. Paralelamente, Liu *et al.* propõem a estrutura *Extremely Randomized Trees* (ERT) - *Edit Distance on Real sequence* (EDR), que utiliza ERT para proteção do protocolo TCP. Embora o ERT-EDR apresente resposta rápida, ele exemplifica o estado da arte em contramedidas estáticas, onde a detecção é robusta, mas carece de mecanismos de atualização para lidar com mudanças no tráfego [Liu et al. 2024].

No que tange à aplicação de técnicas de DL no contexto da 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 uma abordagem focada na análise de fluxos bidirecionais, Ma *et al.* introduzem o sistema *Bidirectional Detection and Traceability Mitigation* (BDTM), que integra informações dos planos de controle e de dados. A inovação do BDTM reside em sua capacidade de rastreabilidade, permitindo a localização da fonte do ataque LDoS e o isolamento de portas específicas de comutadores SDN [Ma et al. 2025]. No entanto, similarmente a outros modelos contemporâneos, como o de [Al-Shukaili et al. 2025], a eficácia do BDTM é validada em cenários com parâmetros de ataque fixos, omitindo a análise da degradação do modelo sob variações dinâmicas do adversário.

Uma estrutura metodológica relevante para este trabalho é o esquema Trident [Tang et al. 2025]. Ao propor uma arquitetura de monitoramento baseada em estados de porta e classificação via algoritmo XGBoost. A importância do Trident para a presente pesquisa reside na robustez de sua metodologia de coleta e rotulagem de tráfego LDoS, que serviu como base para a estruturação de parte dos cenários experimentais aqui propostos. Entretanto, embora o Trident alcance alta acurácia em cenários predefinidos, ele opera sob uma premissa estática que não inclui mecanismos de adaptação contra a deriva de conceito. A incapacidade de modelos estáticos em generalizar para cenários com tráfego dinâmico foi quantificada na literatura [dos Santos et al. 2026]. O estudo demonstrou como variações paramétricas na taxa e duração dos pulsos de ataque LDoS provocam uma degradação severa na contramedida XGBoost treinado *offline*, evidenciando a necessidade de investigação de abordagens adaptativas.

2.2. Limitações dos Modelos Estáticos e Deriva de Conceito

A deriva de conceito é uma limitação dos modelos de ML implantados em ambientes operacionais não estacionários, nos quais a relação entre características e rótulos pode mudar ao longo do tempo [Gama et al. 2014]. Em Sistemas de Detecção de Intrusão (*Intrusion Detection Systems* - IDSes), essa limitação torna-se ainda mais evidente. Shyaa *et al.* apresentam um levantamento abrangente sobre deriva de conceito e dinâmica de características em IDSes, discutindo como as mudanças na distribuição de tráfego e ataques degradam os classificadores estáticos. Eles também exploram como abordagens que consideram a deriva de conceito podem combinar seleção dinâmica de características, modelos adaptativos e mecanismos de detecção para manter o desempenho ao longo do tempo [Shyaa et al. 2024]. Coletivamente, esses resultados apoiam o argumento de que a degradação sob deriva de conceito não é um efeito colateral isolado, mas um desafio estrutural na detecção baseada em ML.

A literatura recente ilustra arquiteturas práticas para a adaptação contínua de contramedidas de detecção em cenários de segurança. Camarda *et al.* propõem um IDSes *online* baseado em *Random Forest* incremental, acoplada a um módulo de aprendizado ativo e um detector de deriva de conceito. Este sistema armazena janelas de dados recentes para atualizações incrementais e utiliza sinais de mudança para acionar o retreinamento direcionado [Camarda et al. 2025]. Expandindo a fronteira do aprendizado adaptativo, Leveni e Boracchi introduziram o OIF, um método especificamente projetado para condições de fluxo contínuo que rastreia o processo de geração de dados à medida que ele evolui ao longo do tempo. Ao contrário das abordagens que dependem do retreinamento periódico para se adaptar ao contexto *online*, o OIF utiliza um conjunto de histogramas dinâmicos com mecanismos de aprendizado e esquecimento para acompanhar continuamente a distribuição dos dados [Leveni and Boracchi 2025].

Apesar de sua capacidade de rastrear a deriva de conceito, os modelos adaptativos irrestritos enfrentam desafios críticos em cibersegurança. Contramedidas baseadas em Aprendizado por Reforço Profundo (*Deep Reinforcement Learning* – DRL) permitem que os agentes aprendam estratégias de detecção adaptativas maximizando recompensas, demonstrando potencial para operação em ambientes SDN [Pandiyakumari S and Suganya R 2025]. Paralelamente, a estrutura AW-XGBoost utiliza pesos adaptativos para ajustar o modelo à medida que novas informações são coletadas [Rong et al. 2024]. De forma semelhante, técnicas que utilizam métricas de Tempo inter-chegada (*Inter-arrival Time* - IAT) e descarte de pacotes [Mathew and Vidhate 2024], embora adaptativas no cálculo de médias, ainda dependem de limiares paramétricos manuais. A dependência de limites fixos ou de realimentação irrestrita contrasta com a necessidade de uma contramedida que aprenda de forma autônoma com a estrutura intrínseca dos dados, mas que possua mecanismos de filtragem para garantir a integridade do modelo contra padrões maliciosos.

Nesse contexto, o aprendizado adaptativo simultaneamente contínuo e seletivo para ataques LDoS em SDN emerge como uma lacuna de investigação. O Sentry destaca-se por estender a arquitetura de isolamento *online* [Leveni and Boracchi 2025] com um critério de atualização condicionado ao estado da rede. Enquanto contramedidas como Trident e ERT-EDR oferecem alta precisão em uma estrutura estática [Tang et al. 2025, Liu et al. 2024], e abordagens baseadas em DRL fornecem adaptação irrestrita [Pandiyakumari S and Suganya R 2025], o Sentry garante que a evolução do modelo ocorra de forma segura. Ao filtrar as amostras que compõem a atualização do classificador, o sistema previne o autoenvenenamento e garante a estabilidade da detecção sob deriva de conceito sem comprometer a eficiência exigida pelo plano de dados.

3. Visão Geral do Sentry

A contramedida proposta para detecção de ataques LDoS é denominada Sentry. O Sentry integra coleta de métricas em tempo real, análise estatística em nível de porta e detecção de anomalias baseada em aprendizado seletivo utilizando o algoritmo OIF [Leveni and Boracchi 2025]. O principal objetivo do Sentry é manter um perfil de detecção robusto que evolua com mudanças legítimas na rede, permanecendo resiliente a ataques maliciosos. Para atingir esse objetivo, a arquitetura do Sentry, ilustrada na Figura 1, é dividida em três camadas funcionais interconectadas: coleta de dados, processamento de atributos e detecção com aprendizado seletivo.

3.1. Coleta de Dados

A coleta de dados do Sentry é realizada de forma contínua e periódica por meio de monitoramento ativo na interface *southbound*, baseando-se em consultas regulares ao plano de dados. O mecanismo de monitoramento executa solicitações de estatísticas com um intervalo de consulta definido por T_{int}, garantindo alta resolução temporal na observação do comportamento da rede. Especificamente, as mensagens `OFPPortStatsRequest` e `OFPFlowStatsRequest` são enviadas a todos os comutadores registrados, permitindo a aquisição de métricas detalhadas de porta e fluxo. As mensagens `OFPPortStatsRequest` retornam estatísticas agregadas por porta, como bytes e pacotes transmitidos e recebidos, enquanto as mensagens

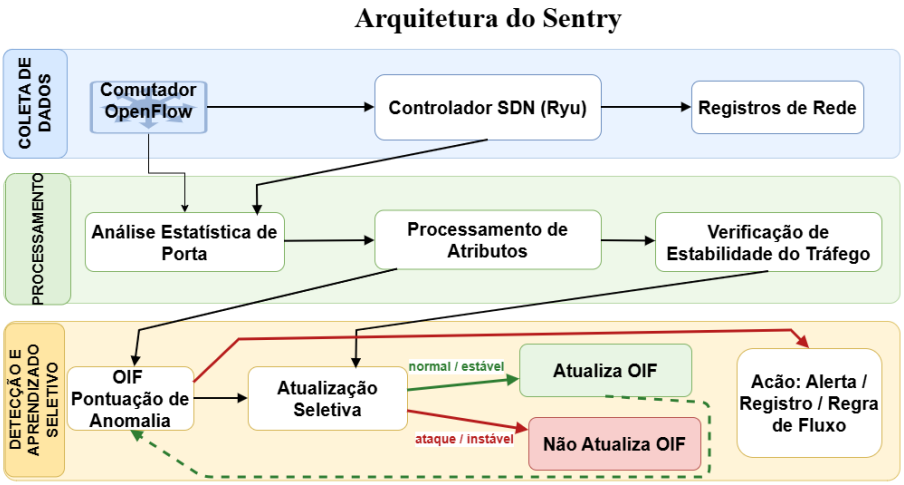

Figura 1. Visão Geral da Arquitetura do Sentry.

`OFPFlowStatsRequest` habilitam estatísticas por fluxo, permitindo a discriminação do tráfego por protocolo, como TCP e *User Datagram Protocol* (UDP).

As métricas de tráfego coletadas são processadas calculando-se a diferença entre amostras consecutivas e convertendo-as em taxas (bytes por segundo e pacotes por segundo). A conversão para taxas temporais elimina efeitos cumulativos e permite a análise dinâmica do tráfego. Os valores das amostras de tráfego obtidas são armazenados em estruturas temporárias e organizados por comutadores e portas, possibilitando a construção de séries temporais por meio de janelas deslizantes de tamanho W. Nesse estágio de organização de dados, o Sentry estrutura métricas essenciais para a identificação de assimetrias, como o *Port Network Flow* (PNF) e o *Paired Port Network Flow* ($PPNF$), que relacionam os fluxos de entrada e saída de portas individuais e de seus pares adjacentes.

3.2. Processamento de Atributos

A etapa de processamento é responsável por transformar os dados brutos coletados em uma representação estruturada e informativa do comportamento da rede. Características estatísticas que descrevem o comportamento do tráfego sob diferentes perspectivas são extraídas das séries temporais. As seguintes características do tráfego de rede são extraídas: medidas de tendência central, como as médias de tráfego μ_{TCP} e μ_{UDP}; medidas de dispersão, como os coeficientes de variação CV_{TCP} e CV_{UDP}; métricas baseadas na distribuição, como a entropia H; métricas estruturais, como a razão entre o tráfego TCP e o tráfego total $ratio_{TCP}$ e os indicadores de fluxo PNF e $PPNF$.

As características extraídas são definidas matematicamente da seguinte forma:

$$\mu_{UDP} = \frac{1}{n}\sum_{i=1}^{n} b_i^{UDP}, \quad \mu_{TCP} = \frac{1}{n}\sum_{i=1}^{n} b_i^{TCP} \tag{1}$$

$$CV_{UDP} = \frac{\sigma_{UDP}}{\mu_{UDP}}, \quad CV_{TCP} = \frac{\sigma_{TCP}}{\mu_{TCP}} \tag{2}$$

$$H = -\sum_{i=1}^{n} p_i \ln(p_i) \tag{3}$$

$$ratio_{TCP} = \frac{\mu_{TCP}}{\mu_{TCP} + \mu_{UDP}} \tag{4}$$

$$PNF = \frac{v_{in}}{v_{out}}, \quad PPNF = \frac{v_{in}^{paired}}{v_{out}^{paired}} \tag{5}$$

$$\mu_{PNF} = \frac{1}{n}\sum_{i=1}^{n} PNF_i, \quad \mu_{PPNF} = \frac{1}{n}\sum_{i=1}^{n} PPNF_i \tag{6}$$

Na Equação (1), b_i^{TCP} e b_i^{UDP} representam as taxas de bytes do tráfego TCP e UDP no instante i, calculadas a partir de estatísticas de fluxo em intervalos de amostragem consecutivos. Os coeficientes de variação definidos na Equação (2) quantificam a variabilidade temporal do tráfego.

A entropia, definida na Equação (3), é calculada a partir da distribuição normalizada das taxas de pacotes dentro da janela de observação, onde p_i representa a proporção da contagem de pacotes no instante i em relação ao número total de pacotes acumulados em todas as W amostras da janela de observação. Assim, a entropia captura a variabilidade e a incerteza da dinâmica da taxa de pacotes ao longo do tempo.

A proporção entre o tráfego TCP e o tráfego total, dada pela Equação (4), fornece informações sobre a composição do tráfego. Finalmente, as métricas estruturais PNF e $PPNF$, definidas na Equação (5), representam a assimetria e a correlação instantâneas do fluxo entre pares de portas. Para o modelo de aprendizado, suas médias temporais, definidas na Equação (6), capturam o comportamento estrutural estável ao longo do tempo.

Durante a fase de aquecimento, o Sentry opera por um período T_{warmup}, no qual coleta apenas dados de tráfego legítimos, sem realizar detecção. Ao final desse período, os dados coletados são utilizados para treinar um normalizador (`StandardScaler`), responsável pela padronização do vetor de características, e também para inicializar o modelo de detecção. Paralelamente, são definidos valores de referência correspondentes às medianas da taxa de transferência TCP e da proporção de tráfego TCP, utilizados posteriormente como critérios de validação na fase operacional. Esse processo de validação garante que o sistema construa uma representação robusta do comportamento normal da rede antes de entrar em modo operacional.

As definições matemáticas apresentadas nas Equações (1) a (6), bem como a estratégia de análise centrada no comportamento das portas dos comutadores, foram baseadas na metodologia proposta pela contramedida Trident [Tang et al. 2025]. O Sentry expande essa abordagem ao integrar tais métricas a um modelo de ML *online* com mecanismos de atualização seletiva. Para viabilizar essa seletividade, o Sentry introduz a métrica $ratio_{TCP}$ (Equação 4), utilizada para validar a integridade do tráfego e garantir a resiliência do aprendizado contra a deriva de conceito em ambientes dinâmicos.

3.3. Detecção e Aprendizado Seletivo

O processo de detecção e aprendizado seletivo do Sentry é dividido em três estágios sucessivos: triagem estatística por porta, classificação e atualização seletiva.

3.3.1. Triagem Estatística por Porta

Inicialmente, realiza-se uma triagem baseada na série temporal da métrica PNF, onde anomalias estatísticas são identificadas por desvios que excedem k desvios padrão ($\mu \pm k\sigma$). Para cada janela deslizante, a proporção de amostras que excedem esse limite é calculada, representando o grau de instabilidade no comportamento da porta. A proporção de amostras da janela deslizante é comparada a um limiar de sensibilidade τ, baseado na probabilidade teórica de ocorrência de valores discrepantes em uma distribuição gaussiana sob a regra de 3σ. De acordo com o modelo de probabilidade normal, aproximadamente 99,73% dos dados devem estar contidos nesse intervalo [Montgomery and Runger 2018]. Uma porta é classificada como anormal quando a fração de valores discrepantes ultrapassa τ, caso contrário, o comportamento é considerado normal. Este critério permite a captura de variações de tráfego sutis, porém persistentes, tornando-o particularmente eficaz na detecção de ataques LDoS, nos quais pequenas perturbações se acumulam ao longo do tempo sem causar mudanças abruptas no volume de tráfego.

3.3.2. Classificação

Caso uma anormalidade seja detectada na triagem, o vetor de atributos estatísticos é submetido ao modelo para a classificação final do evento. O Sentry gera uma pontuação de anomalia que, ao exceder o limite definido, confirma a ocorrência de um ataque e aciona ações de mitigação, como geração de alertas, registro de logs e instalação de regras de fluxo nos comutadores. Essa abordagem híbrida garante que o modelo de ML seja consultado apenas quando há indícios reais de instabilidade, otimizando o processamento.

3.3.3. Aprendizado Seletivo

A atualização do Sentry é condicionada à validação do comportamento do tráfego TCP, verificando se a taxa de transferência e a composição do fluxo TCP permanecem próximas aos valores de referência definidos no período de aquecimento (Equação 7).

$$\begin{aligned} \mu_{TCP} &\geq \alpha \cdot \mu_{base} \\ ratio_{TCP} &\geq \beta \cdot ratio_{base} \end{aligned} \tag{7}$$

onde α e β representam os fatores de tolerância configurados para o Sentry. Esses parâmetros são fundamentais para conferir resiliência ao Sentry, permitindo que flutuações naturais e ruídos de baixa intensidade no tráfego não interrompam o processo de atualização. Os parâmetros μ_{base} e $ratio_{base}$ referem-se aos valores medianos de vazão e proporção TCP, respectivamente, estabelecidos durante a fase de aquecimento da rede. As condições de análise do comportamento do tráfego TCP garantem que o aprendizado ocorra exclusivamente sob comportamento legítimo.

A atualização do modelo ocorre seletivamente, sendo permitida somente quando a porta apresenta comportamento normal por um mínimo de N_{consec} amostras consecutivas e quando as métricas de tráfego TCP permanecem alinhadas com os valores de referência definidos durante o período de aquecimento. Se as condições relacionadas ao comportamento do tráfego de rede não forem atendidas, a atualização é temporariamente suspensa, impedindo a incorporação de dados potencialmente contaminados por tráfego malicioso.

O Sentry adota uma estratégia de processamento em mini-lotes para atualizações, na qual amostras válidas são acumuladas em filas e processadas, reduzindo o custo computacional e aumentando a estabilidade do aprendizado. Essa abordagem adaptativa garante que o Sentry mantenha sua capacidade de detecção mesmo em cenários dinâmicos e adversos, como ataques LDoS com características variáveis, evitando a degradação de desempenho associada à deriva de conceito.

Conforme ilustrado na Fig. 1, o Sentry separa o processo de inferência do processo de aprendizado, garantindo que a detecção de anomalias possa acionar ações de resposta a ataques, enquanto as atualizações do modelo são realizadas somente em condições de rede estáveis e confiáveis. O design adotado no Sentry impede a incorporação de padrões maliciosos durante as atualizações adaptativas, abordando uma limitação fundamental das contramedidas existentes baseadas em aprendizado *online*.

4. Metodologia

Para avaliar o desempenho da arquitetura proposta, adota-se uma abordagem via emulação. Os conjuntos de parâmetros de ataque LDoS foram extraídos da contramedida Trident [Tang et al. 2025], a qual serve de base para avaliação das contramedidas analisadas: XGBoost baseado em *Gradient Boosting* [Chen and Guestrin 2016], OIF e o Sentry.

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 estrutura de rede é baseada na topologia da Rede Nacional de Ensino e Pesquisa (RNP) do Topology Zoo [Knight et al. 2011], ilustrada na Figura 2. A topologia é composta por cinco comutadores (Brasília, Porto Alegre, Florianópolis, São Paulo e Rio de Janeiro) e quatro estações conectadas aos comutadores de Brasília, Porto Alegre, Florianópolis e Rio de Janeiro, com um enlace de gargalo controlável.

Todos os enlaces de acesso e interconexão entre estações e comutadores possuem largura de banda de 100 Mb/s, com exceção do enlace de gargalo, localizado entre São Paulo-Rio de Janeiro, que 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 LDoS 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 Porto Alegre e Florianópolis em direção a Estação Rio de Janeiro. A Estação Porto Alegre envia um fluxo TCP (Porta 5202), enquanto a Estação Florianópolis envia um fluxo UDP (Porta 5203).
- Tráfego de Ataque: A Estação Brasília atua como o atacante, enviando rajadas UDP (Porta 5201) em direção a Estação Rio de Janeiro, visando exaurir a banda do enlace de gargalo e causar descarte de pacotes de fluxos legítimos TCP.

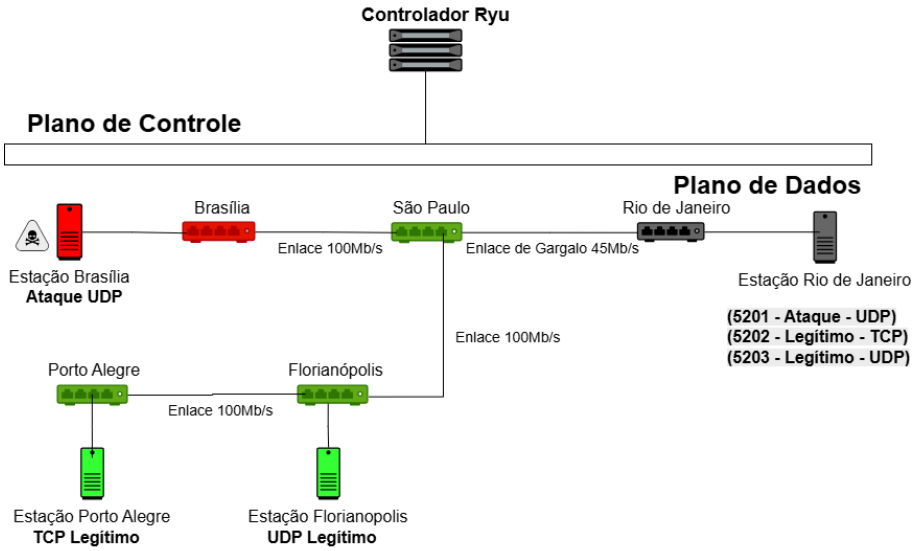

Figura 2. A Topologia de Rede do Ambiente de Avaliação.

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: taxa de ataque (R), período (T) e duração do pulso (L). Para simular um tráfego com deriva de conceito, são conduzidos experimentos variando esses três parâmetros em relação à capacidade do enlace.

No intuito de avaliar as contramedidas reproduzidas, este estudo segrega os testes em três cenários distintos. No primeiro cenário, o XGBoost, como representante do modelo supervisionado é treinado com um conjunto de dados contendo 3.883 amostras, das quais 1.662 (42,80%) pertencem à classe normal e 2.221 (57,20%) à classe ataque, resultando em um conjunto balanceado. Esse cenário envolve tráfego normal e de ataques LDoS de 45 e 55 Mb/s. No segundo cenário, as contramedidas (XGBoost, OIF e Sentry) são submetidas a um conjunto de testes com as taxas de ataques próximas às de treinamento do XGBoost, com valores de 50 e 60 Mb/s. As taxas de ataque usadas para treinamento e teste, bem como a duração dos pulsos, seguem os parâmetros utilizados na contramedida Trident [Tang et al. 2025]. Por fim, no terceiro cenário, são incluídas as taxas de ataque de 30, 70 e 80 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 supervisionada na fase de treinamento. Dessa forma, avalia-se a capacidade das contramedidas aqui reproduzidas quanto a generalização de detecção em diversos parâmetros de ataque.

- XGBoost: cada execução tem duração total de 6 minutos, divididos em 180 segundos de somente tráfego legítimo iniciados em $t_1 = 0$, seguido de mais 180 segundos de tráfego legítimo com ataque LDoS, iniciados em $t_2 = 180$.
- OIF e Sentry: cada execução tem duração total de 9 minutos, estruturada da seguinte forma: 360 segundos de somente tráfego legítimo, sendo o instante $t_1 = 0$ referente a fase de aquecimento, $t_2 = 180$ ao início do modo operação, seguido de mais 180 segundos de tráfego legítimo com ataque LDoS no instante $t_3 = 360$. A Tabela 1 resume os parâmetros utilizados na geração dos fluxos de tráfego.

Tabela 1. Parâmetros de Configuração do Experimento.

Parâmetro	**XGBoost**	**OIF e Sentry**
Duração dos Cenários	6 minutos	9 minutos
Fase de Aquecimento	–	$t_1 = 0$: Tráfego legítimo
Tráfego Legítimo	$t_1 = 0$: Tráfego legítimo	$t_2 = 180$: Tráfego legítimo
Ataque LDoS	$t_2 = 180$: Início do ataque	$t_3 = 360$: Início do ataque
Cenários de ataque	R = 50 Mb/s, T = 2,5 s, L = 0,4 s	
	R = 60 Mb/s, T = 1,5 s, L = 0,2 s	
	R = 60 Mb/s, T = 2,5 s, L = 0,2 s	
	R = 30 Mb/s, T = 1,0 s, L = 0,4 s	
	R = 70 Mb/s, T = 1,4 s, L = 0,4 s	
	R = 80 Mb/s, T = 1,1 s, L = 0,3 s	

4.3. Parametrização e Configurações de Controle

As definições operacionais do Sentry, incluindo os limiares de segurança para o aprendizado seletivo e os parâmetros do detector estatístico, estão consolidadas na Tabela 2. Esses parâmetros foram definidos por meio de testes exploratórios durante o desenvolvimento da contramedida, mantidos fixos durante toda a avaliação experimental e visam garantir que a contramedida seja atualizada apenas sob condições de tráfego legítimo, permitindo uma variação natural de até 5% em relação ao baseline (α e β) e utilizando o critério estatístico de 3σ (k) para a filtragem de valores discrepantes.

Tabela 2. Parâmetros de Configuração e Controle do Sentry.

Símbolo	**Descrição**	**Valor**
W	Tamanho da Janela Deslizante	30
T_{warmup}	Período de aquecimento do tráfego	180 s
α	Fator de tolerância para vazão média TCP	0,95
β	Fator de tolerância para proporção TCP	0,95
k	Multiplicador de desvio padrão	3
τ	Limiar de proporção de anomalias estatísticas	0,0026
T_{int}	Intervalo de consulta	0,5 s
N_{consec}	Amostras consecutivas normais	25

4.4. Validação Estatística dos Resultados

Cada cenário de teste é submetido a 30 execuções independentes utilizando a configuração apresentada na Tabela 2. A confiabilidade das métricas é aferida através de intervalos de confiança de 99%. Essa metodologia assegura a representatividade do comportamento médio da contramedida perante a natureza estocástica da rede, sendo crucial para distinguir degradações reais de flutuações aleatórias.

5. Análise de Desempenho

Nesta seção, avalia-se o desempenho das diferentes contramedidas consideradas neste trabalho. Inicialmente, investiga-se a perda de desempenho da contramedida XGBoost

quando exposto a variações dinâmicas na intensidade e na duração dos pulsos de ataque, evidenciando os efeitos da deriva de conceito. Adicionalmente, avalia-se o comportamento do OIF sem mecanismos de controle de atualização, destacando sua suscetibilidade à contaminação por tráfego malicioso. Por fim, analisa-se a robustez do Sentry, demonstrando sua capacidade de manter desempenho consistente mesmo sob diferentes configurações de ataques LDoS.

5.1. Análise Comparativa das Contramedidas

A análise comparativa concentra-se na capacidade de cada contramedida em manter desempenho consistente diante de variações nos parâmetros do ataque, caracterizando cenários de deriva de conceito. A Figura 3 apresenta o comparativo de desempenho das três contramedidas, considerando as métricas de Revocação e Pontuação F1 para os seis cenários de ataque definidos na Tabela 1. Os cenários são projetados para simular variações nas taxas de ataque, incluindo configurações próximas ao conjunto de treinamento e outras significativamente distintas, a fim de avaliar a capacidade de generalização das contramedidas. Ressalta-se que a avaliação das métricas de Revocação e Pontuação F1 é realizada exclusivamente sobre os comutadores (São Paulo e Rio de Janeiro) diretamente impactados pelo tráfego de ataque LDoS, uma vez que apenas esses refletem efetivamente o comportamento das contramedidas frente às condições adversas impostas. Adicionalmente, os 20 segundos iniciais e finais de cada execução são descartados para mitigar ruídos transientes de estabilização da rede e variações na sincronização dos registros.

A fim de garantir uma comparação isonômica, as três contramedidas operam sobre a mesma base de extração de métricas. Isso significa que o processo de normalização por porta PNF e a extração dos 10 atributos estatísticos derivados das métricas apresentadas nas Equações (1) a (6) foram idênticos para todos os testes, submetendo cada contramedida exatamente aos mesmos sinais de rede e indicadores de anomalia.

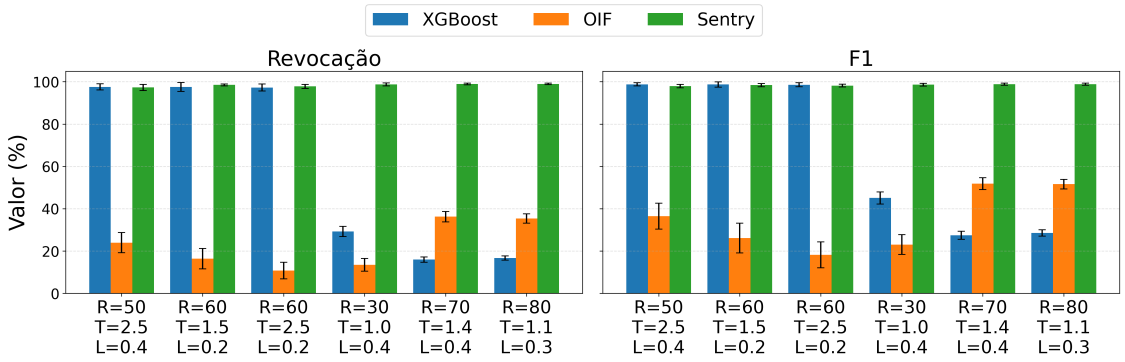

Figura 3. Métricas de desempenho das contramedidas XGBoost, OIF e Sentry para os diferentes cenários de ataques LDoS (Tabela 1) com variação dos parâmetros de período T e largura de pulso L durante o ataque.

Conforme pode ser observado, o XGBoost mantém altas métricas de desempenho quando os parâmetros de ataque são próximos aos parâmetros em que é treinado (45 e 55 Mb/s). Entretanto, sofre uma queda acentuada em suas métricas quando submetido a padrões distantes do que foi visto, como nos cenários de 30, 70 e 80 Mb/s, nos quais a métrica de revocação cai para 30,99%, 17,95% e 19,07%, respectivamente. Esse com-

portamento evidencia a limitação do modelo em lidar com mudanças na distribuição dos dados, resultando na não detecção de aproximadamente 69% a 82% do tráfego malicioso.

Por sua vez, o OIF, na ausência de mecanismos de controle sobre o processo de atualização, passa a incorporar gradualmente padrões de tráfego malicioso em sua base de conhecimento. Essa contaminação do OIF resulta na redução progressiva da capacidade de detecção. Experimentalmente, observa-se que, após períodos prolongados de ataque, o OIF apresenta valores de revocação significativamente reduzidos, variando aproximadamente entre 10,50% e 31,90%, além de baixas pontuações de F1, situadas entre 17,87% e 46,95%, indicando aumento de falsos negativos e perda de sensibilidade ao ataque.

Em contraste, o Sentry demonstra maior estabilidade e robustez ao longo de todos os cenários avaliados. Ao restringir o processo de atualização do modelo apenas a condições específicas de tráfego legítimo, como estabilidade da vazão TCP, consistência da composição do tráfego e sequência de estados normais, o Sentry evita a incorporação de padrões de ataque durante sua fase adaptativa. Como resultado, o Sentry mantém altos níveis de desempenho mesmo sob variações significativas nos parâmetros de ataque. O Sentry atinge consistentemente uma taxa de revocação de até 98,89%, juntamente com uma pontuação F1 de até 98,53%, demonstrando sua capacidade de adaptação sem comprometer a precisão da detecção.

5.2. Impacto da Mitigação na Vazão Legítima

A Figura 4 apresenta a evolução da vazão TCP legítima no cenário de ataque com $R = 70$ Mb/s, $T = 1{,}4$ s e $L = 0{,}4$ s. Constata-se que, na ausência de mitigação, a vazão sofre uma queda após o início do ataque em $T = 360\ s$, reduzindo-se para aproximadamente 10 Mb/s. A contramedida XGBoost, por sua vez, também não preserva a vazão do TCP, indicando falha de detecção sob deriva de conceito e atingindo níveis de vazão igualmente próximos de 10 Mb/s. O OIF apresenta oscilações após o início do ataque, com a vazão do TCP sofrendo degradação e alcançando aproximadamente 25 Mb/s, refletindo sua instabilidade decorrente da atualização irrestrita. Em contraste, o Sentry mantém a vazão próxima ao comportamento observado antes do ataque, evidenciando que a mitigação acionada pelos comutadores impactados preserva o tráfego legítimo.

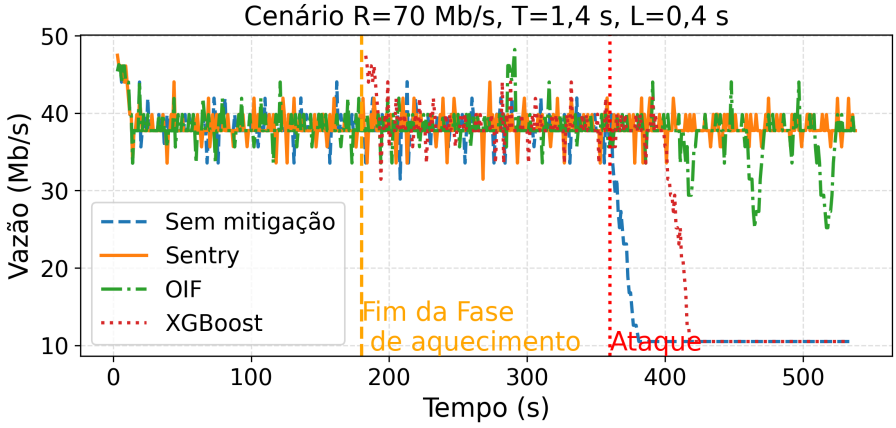

Figura 4. Evolução temporal da vazão TCP legítima no cenário com taxa *R* = 70 Mb/s, período *T* = 1,4 s e duração do pulso *L* = 0,4 s.

As evidências demonstradas confirmam que contramedidas baseadas na premissa de estacionariedade tornam-se vulneráveis a ataques LDoS com parâmetros dinâmicos, uma vez que alterações na intensidade e periodicidade do tráfego deslocam o comportamento do ataque para regiões não observadas durante o treinamento. De forma complementar, observa-se que abordagens adaptativas irrestritas, como o OIF, sofrem degradação ao longo do tempo devido à incorporação progressiva de padrões maliciosos durante o processo de atualização. Nesse contexto, o Sentry demonstra-se mais adequado para cenários dinâmicos, ao incorporar mecanismos de controle no processo de atualização. Ao restringir o aprendizado a condições específicas de estabilidade do tráfego legítimo, o Sentry evita a contaminação do modelo e preserva sua capacidade de detecção ao longo do tempo. Dessa forma, elimina-se a necessidade de retreinamento manual frequente, reduzindo o custo operacional e permitindo uma resposta mais compatível com a velocidade de adaptação dos ataques modernos. Os resultados reforçam a importância de estratégias de aprendizado adaptativo controlado como um caminho promissor para o desenvolvimento de contramedidas robustas em ambientes SDN.

6. Conclusão e Trabalhos Futuros

Este trabalho investigou a fragilidade estrutural de contramedidas baseadas em ML frente a ataques LDoS dinâmicos, com foco em três abordagens distintas: as contramedidas XGBoost, OIF e Sentry. Diferentemente da literatura predominante, que prioriza a maximização de métricas em cenários controlados, a avaliação conduzida neste artigo enfatizou a análise de vulnerabilidade decorrente da deriva de conceito induzida pela variabilidade dos ataques ao longo do tempo.

Os resultados obtidos permitem traçar três conclusões principais. Primeiramente, demonstrou-se que o XGBoost, quando treinado de forma estática, sofre degradação de desempenho ao ser exposto a cenários distintos daqueles observados durante o treinamento, com a revocação e pontuação F1 caindo para níveis próximos de 18% e 30%, respectivamente. Em segundo lugar, observou-se que, embora o OIF apresente capacidade adaptativa, sua ausência de mecanismos de controle no processo de atualização resulta na contaminação do modelo e, consequentemente, na perda progressiva de capacidade de detecção, com a revocação e pontuação F1 chegando a valores próximos de 10% e 18%, respectivamente. Finalmente, o Sentry demonstrou maior robustez frente à variabilidade dos ataques ao incorporar critérios seletivos no processo de atualização. Ao restringir o aprendizado a condições que representam efetivamente o tráfego legítimo, o Sentry evita a contaminação e mantém um desempenho consistente mesmo em cenários dinâmicos, com valores de revocação e pontuação F1 atingindo até 98,89% e 98,53%, respectivamente.

Conclui-se, portanto, que a premissa de estacionariedade adotada em contramedidas estáticas representa uma vulnerabilidade crítica, enquanto abordagens adaptativas sem controle introduzem um novo vetor de fragilidade relacionado à contaminação do modelo. Nesse sentido, a proposta de aprendizado seletivo apresentada neste trabalho aponta para uma mudança de paradigma na defesa contra ataques LDoS, ao equilibrar adaptabilidade e confiabilidade na detecção.

Como trabalhos futuros, pretende-se investigar a ampliação do Sentry para cenários mais complexos, incluindo múltiplos vetores de ataque e ambientes com maior heterogeneidade de tráfego. Além disso, busca-se explorar estratégias adicionais de controle

do aprendizado. O objetivo é avançar no desenvolvimento de contramedidas capazes de operar de forma autônoma e escalável em ambientes SDN sujeitos a ataques dinâmicos.

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq (408255/2023-4 e 407304/2025-8), CAPES (88887.987121/2024-00 e 88881.987122/2024-01), COFECUB (Ma1074/25), 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/Countermeasure-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).

Alashhab, A. A., Zahid, M. S. M., Alashhab, M., and Alashhab, S. (2023). Online machine learning approach to detect and mitigate low-rate ddos attacks in sdn-based networks. In *2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pages 1–6, Kota Kinabalu, Malaysia. IEEE.

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.

dos Santos, B. M., Bastos, I. V., and Moraes, I. M. (2026). Degradação de desempenho de contramedidas estáticas a ataques de negação de serviço de baixo volume. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, Brasil. Sociedade Brasileira de Computação (SBC).

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.

Knight, S., Nguyen, H. X., Falkner, N., Bowden, R., and Roughan, M. (2011). Internet topology zoo. *IEEE Journal on Selected Areas in Communications*, 29(9):1765–1775.

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.

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.

Mathew, R. R. and Vidhate, A. (2024). Adaptive dos attack detection in sdn. In *Proceedings of the 2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE.

Montgomery, D. C. and Runger, G. C. (2018). *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 7th edition.

Pandiyakumari S, H. and Suganya R, D. (2025). Adaptive detection of low-rate denial-of-service attacks: A machine learning and reinforcement learning perspective. In *Proceedings of the 2025 6th International Conference on Communication, Computing & Industry 6.0 (C2I6)*. IEEE.

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.

Rong, X., Wang, S., Guo, C., and Tao, X. (2024). Adaptive weight xgboost: Detecting and mitigating low-rate dos attack in network slicing. In *Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE.

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.

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/`.