# Engenharia de Sinais Precoces de Alerta Para a Predição de Ataques DDoS

**Anderson B. de Neira[1], Ligia F. Borges[2], Alex M. de Araújo[1], Michele Nogueira[1,2]**

[1]Departamento de Informática - Universidade Federal do Paraná

[2]Departamento de Ciência da Computação - Universidade Federal de Minas Gerais

{abneira,amaraujo}@inf.ufpr.br, {ligia.borges,michele}@dcc.ufmg.br

***Resumo.*** *A rápida escalada dos ataques de negação de serviço distribuído (DDoS) requer sua predição. Grandes volumes de tráfego de ataques limitam o tempo disponível para detê-los. Os atacantes escondem suas ações da detecção com o objetivo de aumentar seus danos. Portanto, é essencial predizer os sinais da preparação dos ataques. Este trabalho apresenta a abordagem **E**ngenharia de **S**inais **P**recoces de **A**lerta (ESPA). ESPA gera características que possam indicar a preparação do ataque no tráfego de rede baseado na teoria dos sinais precoces de alerta. A partir da indicação é possível detê-los antecipadamente. A abordagem ESPA permite a predição de um ataque DDoS 30 minutos antes do seu início efetivo.*

***Abstract.*** *The quick escalation of Distributed Denial of Service attacks (DDoS) requires their prediction. The large volume of attack traffic constraints the time available to prevent them. Attackers hide their actions to delay attack detection and increase their damage. Therefore, it is essential to identify the signs of attack preparation. This work presents an early warning signs engineering approach (ESPA). ESPA generates features based on the theory of early warning signals to enhance attack prevention. With this identification, it is possible to stop them in advance. The ESPA approach allows predicting a DDoS attack 30 minutes before its start.*

## 1. Introdução

As ameaças cibernéticas evoluem continuamente, dentre elas, o ataque de negação de serviço distribuído (do inglês, *Distributed Denial of Service* — DDoS) é um dos mais comprometedores [Jyoti and Behal 2021]. Após o atacante lançar efetivamente o ataque, os administradores de rede são surpreendidos pelo rápido crescimento no consumo de recursos que gera uma negação de serviço e prejudica os usuários reais. No primeiro semestre de 2022, os especialistas identificaram 6.019.888 ataques DDoS oriundos de 190 países diferentes [Netscout 2022]. Esses ataques alcançam grandes volumes de dados em uma velocidade nunca vista antes. A Microsoft Corporation foi alvo de um ataque DDoS que, em apenas um minuto, atingiu 3,47 terabits por segundo [Toh et al. 2022].

A literatura ressalta a importância de coibir os ataques DDoS antes que eles efetivamente iniciem [Gupta and Badve 2017]. Contudo, a tarefa de identificar indícios da preparação dos ataques é desafiadora, pois os atacantes evitam conduzir ações que impactam nas características do tráfego de rede usadas pelos sistemas tradicionais de detecção de ataque DDoS. Assim, essas características são inadequadas para o combate antecipado dos ataques DDoS. Além disso, devido ao grande desbalanceamento dos dados, a preparação dos ataques é frequentemente confundida com o tráfego normal da rede, pois a fase de preparação produz um volume de tráfego de rede ínfimo comparado ao tráfego do ataque. Isto posterga a detecção do ataque para quando este se encontra em uma fase avançada, sendo consequentemente mais difícil de lidar.

Definir características representativas para identificar grandes mudanças de comportamento não é um problema particular dos ataques DDoS. Mesmo sem mudanças claras no comportamento, é possível observar sinais precoces de alerta através de certas características e utilizá-las para predizer ocorrências futuras. Por exemplo, Proverbio *et al.* (2022) avaliaram os sinais precoces de alerta para predizer novas ondas do novo coronavírus. Os autores realizaram a predição por meio do cálculo da variância, da autocorrelação do *lag* 1 (AC-1), o coeficiente de variação (CV) e o *skewness* sobre o total de casos ativos. Tasa *et al.* (2022) usaram o *skewness* e a *kurtosis* para categorizar as vibrações de terremotos. Os autores extraíram o valor do *skewness* e a *kurtosis* e treinaram uma Rede Neural Artificial. Apesar do potencial dos sinais precoces de alerta, é incipiente a investigação deles para a identificação de indícios da preparação dos ataques e, particularmente, dos ataques DDoS.

Este trabalho apresenta uma abordagem denominada de **E**ngenharia de **S**inais **P**recoces de **A**lerta (ESPA) a fim de sistematizar a predição de ataques DDoS embasada na teoria dos sinais precoces de alerta. As técnicas de predição produzem indícios de ataques futuros (*i.e.*, ainda não lançados) [Abdlhamed et al. 2017]. A abordagem ESPA processa os dados do tráfego de rede e gera sinais precoces de alerta. Os sinais precoces de alerta são métricas estatísticas que indicam futuras mudanças no comportamento do sistema [Scheffer 2009]. O objetivo da abordagem ESPA é criar características do tráfego de rede que suportem a predição eficiente dos ataques DDoS.

Este artigo avalia a abordagem ESPA seguindo três experimentos. Dois experimentos usam capturas disponibilizada pelo *Czech Technical University dataset* (CTU-13) referente à rede local da universidade [Garcia et al. 2014] e um usou a captura *DDoS Evaluation dataset* (CIC-DDoS2019) [Sharafaldin et al. 2019] onde a vítima está conectada aos atacantes pela Internet. No primeiro experimento, a abordagem apoiou a predição do ataque que ocorreu três minutos e 56 segundos antes do seu lançamento e, no segundo, a abordagem auxiliou a predição que ocorreu 30 minutos antes do atacante lançar efetivamente o ataque. A acurácia proporcionada pela abordagem para a predição foi de, respectivamente, 92,9% e 84,6%. No terceiro experimento, a proposta proporcionou a predição do ataque DDoS 15 minutos e 1 segundo antes do início do ataque.

A abordagem ESPA gerou novas características permitindo a predição dos ataques. Os resultados viabilizam o uso da teoria dos sinais de alerta precoce para predizer ataques DDoS, além do aumento do tempo de predição em relação à literatura. A literatura prediz ataques DDoS com cinco minutos e 41 segundos de antecedência [Rahal et al. 2020], enquanto a abordagem ESPA proporcionou a predição do ataque com 30 minutos de antecedência. A abordagem proposta facilita a adoção em ambientes reais, pois evita o uso de dados rotulados para a predição de ataques DDoS. A explicabilidade também é uma contribuição deste trabalho, pois o resultado da abordagem ESPA é expresso visualmente, permitindo a compreensão da predição dos ataques DDoS.

Este artigo prossegue como segue. A Seção 2 apresenta a literatura dos sinais precoces de alerta e da predição de ataques. A Seção 3 detalha a abordagem ESPA. A Seção 4 apresenta os resultados deste trabalho. Por fim, a Seção 5 conclui este artigo.

## 2. Trabalhos Relacionados

A literatura apresenta o uso de diferentes sinais precoces de alerta para prever ocorrências futuras em diferentes áreas, com maior ênfase na previsão do tempo, mas não apenas. Takimoto (2009) utilizou os sinais precoces de alerta *skewness*, o desvio padrão e a taxa de retorno para predizer variações da população de espécies invasoras. Takimoto (2009) aplica os sinais precoces de alerta sobre a variação da população invasora para predizer uma explosão populacional repentina. É importante prever mudanças populacionais, pois os invasores podem superar as espécies nativas, encaminhando-as para a extinção (ou seja, uma transição crítica). Boers e Rypdal (2021) avaliaram a os sinais precoces de alerta (variância e a AC-1) para a predição do derretimento da camada de gelo da Groenlândia. Os autores avaliaram os sinais precoces sobre o derretimento do gelo e sobre as flutuações da altura do manto de gelo. Em ambos os casos, os autores identificaram aumentos nos sinais precoces de alerta, sugerindo que, no futuro próximo, o derretimento de gelo pode sofrer uma mudança permanente de comportamento (transição crítica).

Quando aplicados à predição de ataques DDoS, a literatura sobre os sinais precoces de alerta é limitada. Santos *et al.* (2013) propuseram uma solução para monitorar textos encontrados na Internet. O objetivo de monitorar textos de mídias sociais é centralizar o processamento das postagens e usá-las para identificar potenciais alertas. Assim, é possível utilizar esses alertas para antecipar problemas em redes de computadores. Apesar de não ser específico para ataques DDoS, a solução filtra postagens relacionadas com ataques DDoS, desde que os atacantes sinalizem o ataque. Salemi *et al.* (2021) propuseram uma solução para predizer ataques DDoS usando a *Recurrent Neural Echo State Network* (SCESN). A solução processa o tráfego da rede e projeta o comportamento do tráfego da rede para os próximos segundos. Com a projeção do tráfego futuro, a solução prevê o erro dessa projeção utilizando a SCESN. A SCESN é pré-treinada com base em dados históricos da rede. A solução analisa esse erro de predição, usando o expoente de Lyapunov para predizer o ataque. O melhor resultado foi predizer ataques com antecipação de 20 segundos. Entretanto, utilizar dados históricos para o treinamento limita a solução, pois ataques diferentes dos treinados podem não ser preditos.

Kivalov e Strelkovskaya (2022) criaram uma solução para prever ataques DDoS com base na extrapolação *spline* do tráfego de rede. A extrapolação de *spline* prevê o pico de tráfego com base no tráfego autossemelhante. Os autores utilizaram as extrapolações lineares e cúbicas para prever ataques DDoS. A solução utiliza o tráfego de rede antes, durante e depois de um ataque DDoS para prever ataques semelhantes. As *splines* cúbicas apresentaram os melhores resultados para prever o ataque nos próximos segundos. Apesar de os métodos de extrapolação de *spline* serem simples de implementar, a solução só pode prever ataques DDoS que correspondam aos imputados. Assim, generalizar a solução para outros ataques DDoS e diminuir o uso de dados rotulados é necessário.

Em geral, a literatura foca na predição de ataques usando características do tráfego de rede convencionais. Sabendo disso, os atacantes escondem suas ações gerando a menor quantidade de tráfego possível. Além disso, a literatura usa comumente dados rotulados para treinar soluções. Isso dificulta a generalização dos resultados e a solução pode não ser eficaz contra ataques DDoS diferentes dos treinados, incluindo os ataques denominados de *zero-day*. Os poucos trabalhos que lidam com a predição de ataques DDoS, como [Rahal et al. 2020], apresentam resultados pouco compreensíveis, são complexos para configurar ou não são customizáveis. Assim, este trabalho evolui a literatura com uma abordagem de engenharia de sinais para suportar a predição de ataques DDoS que gera novas características representativas do tráfego de rede.

## 3. Engenharia de sinais precoces de alerta

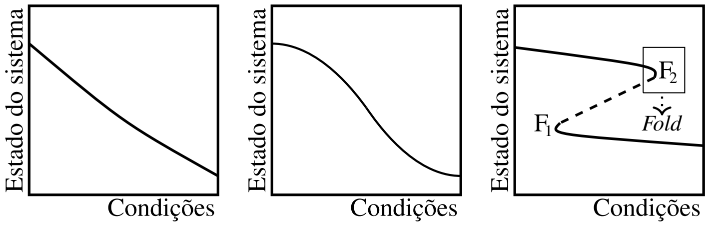
Esta seção descreve a proposta de engenharia de sinais precoces de alerta (ESPA) para a predição de ataques DDoS. Esta abordagem analisa o tráfego de rede sobre a perspectiva dos sinais precoces e gera novas características para identificar mudanças no comportamento da rede e antecipar possíveis ataques DDoS. A Subseção 3.1 detalha a teoria que embasa a abordagem ESPA, e a Subseção 3.2 apresenta o funcionamento da abordagem.

### 3.1. A teoria dos sinais precoces de alerta

A abordagem ESPA utiliza a teoria dos sinais precoces de alerta. O estudo de sinais precoces de alerta baseia-se em conceitos como sistemas em equilíbrio, sistemas dinâmicos e na análise de séries temporais. Os sistemas naturais podem estar em equilíbrio, ou seja, mesmo que as condições (*i.e.*, parâmetros do sistema) variem, o sistema tende a compensar essa variação e retomar o equilíbrio. Alguns sistemas podem transitar de um ponto de equilíbrio para outro ponto de equilíbrio conforme mudanças nas condições. Por exemplo, um sistema equilibrado no estado A pode transitar para o estado B conforme as condições variem. O novo ponto de equilíbrio pode apresentar características similares ou diferentes às observadas no ponto A. Prever e entender como ocorrem as mudanças auxiliam no seu gerenciamento, evitando catástrofes e extinções, por exemplo [Scheffer 2009].

Um sistema pode transitar entre os estados de equilíbrio de diferentes formas frente às mudanças. Um sistema em equilíbrio pode transitar suavemente (Figura 1(a)) ou abruptamente (Figura 1(b)) para outro estado de equilíbrio. Alguns sistemas não conseguem transitar suavemente ou abruptamente entre os estados de equilíbrio (Figura 1(c)). Quando as condições mudam, o sistema é forçado a transitar criticamente entre estados de equilíbrio. A parte superior da curva na Figura 1(c) apresenta o comportamento inicial do sistema. Com a evolução das condições observadas, o sistema varia criando uma dobra na representação gráfica do sistema. Quando as condições observadas variam o suficiente, o sistema supera o ponto $F_2$ e encaminha-se para o novo comportamento do sistema, iniciado em $F_1$. Devido à dobra, o sistema não transita suavemente para a parte inferior da curva quando as condições observadas mudam. A transição dos pontos $F_2$ para $F_1$ é marcada por instabilidades e é denominada de transição crítica. Durante a transição crítica, um sistema pode apresentar instabilidades, por exemplo, uma espécie ser extinta. Outra característica das transições críticas é que mesmo retornando as condições para o patamar anterior, o sistema pode não retornar ao antigo ponto de equilíbrio. Deste modo, recuperar sistemas que passam por transições críticas não é uma tarefa trivial [Scheffer 2009].

A literatura evoluiu a ponto de identificar evidências capazes de representar a ocorrência de transições críticas durante as observações dos sistemas. As evidências representam mudanças nos valores calculados sobre os dados observados a partir de métricas estatísticas antes da transição crítica. Portanto, é possível utilizar métricas estatísticas para identificar sinais precoces de alerta e antecipar transições críticas [Bury et al. 2020]. A variância, *kurtosis*, *skewness*, AC-1, CV e o *power spectrum* são exemplos de métricas estatísticas capazes de produzir sinais precoces de alerta [Dakos et al. 2012]. Em geral, as métricas estatísticas são genéricas. Isso significa que eles se apoiam em características comuns das transições críticas para calcular os sinais precoces de alerta. Isso é benéfico pois eles podem ser utilizados em diferentes aplicações. Assim, a abordagem proposta neste trabalho toma como base as métricas estatísticas para predizer os ataques DDoS.




(a) Transição suave (b) Transição abrupta (c) Transição crítica

**Figura 1. Exemplos de transição dos sistemas [Scheffer 2009]**

Para proporcionar alto nível de explicabilidade, este trabalho avaliou três sinais precoces de alerta. O primeiro deles é o ***Kurtosis*** (Eq. 1). O termo $N$ representa a quantidade total de itens observados em uma série temporal. Para calcular o *Kurtosis* é necessário o $\hat{y}$ (Eq. 2). O termo $x_t$ refere-se a cada item observado na série temporal com seu índice. E o termo $\bar{x}$ refere-se a média simples de toda a série temporal. Apesar das diferentes interpretações para o *Kurtosis* [Oja 1981], uma das mais difundidas e controversas relaciona o valor do *Kurtosis* com grau de achatamento da curva da série temporal. O *Kurtosis* pode ser utilizado como sinal precoce de alerta pois Biggs *et al.* (2009) e Dakos *et al.* (2012) verificaram que o valor do *Kurtosis* pode variar ou apresentar picos próximo de transições críticas. Por exemplo, a distribuição de dados do sistema pode apresentar um padrão plano longe da transição crítica e um padrão com pico próximo à transição crítica. Essas variações indicam mudanças na distribuição de dados. Essas mudanças podem antecipar transições críticas [Dakos et al. 2012], permitindo a predição de ataques DDoS.

$$Kurtosis = \frac{(N-1)}{(N-2)(N-3)}(N-1)\hat{y} + 6 \quad (1) \qquad \hat{y} = \frac{N\sum(x_t - \bar{x})^4}{[\sum(x_t - \bar{x})^2]^2} \quad (2)$$

O segundo sinal precoce de alerta utilizado neste trabalho é o ***Skewness*** (Eq. 3). O termo $x_t$ refere-se a cada item observado na série temporal. O $N$ representa a quantidade total de itens observados. O $\bar{x}$ refere-se a média aritmética simples do conjunto e $\sigma$ representa o desvio padrão dos dados (Eq. 4). O *Skewness* mensura o grau de assimetria das observações de uma série temporal. Guttal e Jayaprakash (2008) verificaram que aumentos na assimetria da distribuição podem indicar um alerta precoce confiável.

$$Skewness = \frac{N\sum_{t=1}^{N}(x_t - \bar{x})^3}{(N-1)(N-2)\sigma^3} \quad (3) \qquad \sigma = \sqrt{\frac{\sum x - \bar{x}^2}{(N-1)}} \quad (4)$$

O **coeficiente de variação** (CV) é o terceiro sinal precoce de alerta utilizado por este trabalho. Para obter o CV basta dividir o desvio padrão dos dados analisados ($\sigma$) pela média dos dados ($\bar{x}$), onde $\bar{x} \neq 0$ [Bedeian and Mossholder 2000]. O CV é utilizado como indicador de diversidade em relação à média dos conjuntos de dados analisados. Assim, o limite inferior de CV ($CV = 0$) indica uniformidade completa do conjunto de dados [Bedeian and Mossholder 2000]. A literatura identificou que aumentos no CV indicam a ocorrência de uma transição crítica. Portanto, o CV pode ser utilizado como um sinal precoce de alerta [Dakos et al. 2012].

### 3.2. A abordagem ESPA

A abordagem ESPA consiste na aplicação dos sinais precoces de alerta sobre os dados do tráfego de rede para a criação de novas características que suportem a eficiente detecção de ataques DDoS. O objetivo de criar características é ressaltar a preparação dos ataques e proporcionar uma abordagem para predição de ataque DDoS explicável, customizável e sem depender de dados rotulados. A Figura 2 apresenta o funcionamento da proposta.

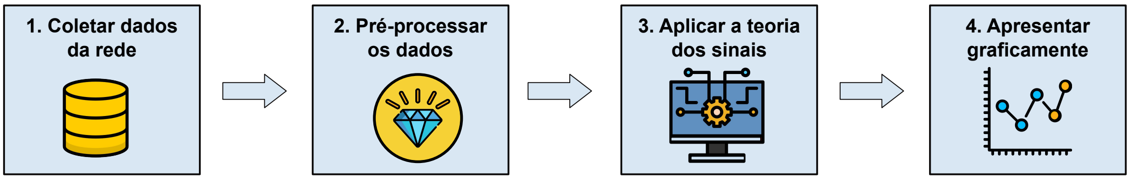



**Figura 2. Visão geral da ESPA**

A Etapa 1 define a coleta do tráfego de rede. Existem algumas formas de coletar o tráfego de rede. Em uma coleta centralizada, o roteador encaminha uma cópia do cabeçalho dos pacotes para um dispositivo que irá executar a abordagem. Dependendo do volume de dados trafegados, o processamento centralizado requer muitos recursos. Assim, o administrador de rede pode processar o tráfego de rede de modo distribuído. O processamento distribuído, quando possível, é preferível para evitar entraves no processamento dos dados. A abordagem ESPA é independente do modo de coleta de dados. Assim, a Etapa 1 indica que a coleta do tráfego de rede deve ser feita. O administrador de rede define o modo que se adeque aos seus objetivos e recursos disponíveis.

O administrador de rede define as características do tráfego de rede e executa a coleta de dados sobre elas. O imperativo para que a abordagem funcione é que as características do tráfego de rede utilizadas possam ter impacto diante de uma preparação do ataque. Por exemplo, Jaber *et al.* (2017) citam que atacantes podem realizar testes antes do ataque. Assim, características como a quantidade de dispositivos trocando pacotes ou quantidade de pacotes enviados/recebidos podem ser influenciados pela preparação dos ataques [Jaber et al. 2017]. Feng *et al.* (2018) apresentam 40 atributos representativos para a detecção de comunicação de comando e controle (C&C). Como a comunicação de C&C ocorre antes do ataque, esses atributos podem ser usados para a predição de ataques DDoS. Assim, neste momento, este trabalho não realiza a seleção de atributos.

Dados faltantes e escalas diferentes degradam o desempenho das técnicas de AM. A abordagem ESPA pré-processa os dados faltantes de dois modos (Etapa 2). A abordagem pode remover as características com entradas faltantes, adicionar um valor baseado na moda, mediana, média ou ainda adicionar uma constante, por exemplo “-1”. Para lidar com dados em escalas diferentes, a abordagem utiliza a padronização. A padronização dimensiona os atributos removendo a média das observações dividido pelo desvio padrão delas. A abordagem ESPA também pode utilizar a estratégia Min-Max para dimensionamento dos dados. Esta estratégia dimensiona os valores das características entre um intervalo pré-definido. O valor do intervalo pode ser entre zero e um, ou de forma absoluta utilizando os valores reais da base [de Neira et al. 2022]. O uso da Etapa 2 é opcional e customizável pois não existem dados faltantes nos experimentos realizados neste trabalho. Além disso, é o intuito deste avaliar como a solução vai ser portar frente a dados não padronizados. Contudo é recomendado usá-la para obter predições mais acuradas.

Para que a abordagem opere corretamente o tráfego de rede coletado na forma de características deve estar ordenado em relação ao tempo da coleta. O administrador pode definir o intervalo de tempo da coleta. O tráfego de rede coletado é armazenado na forma de um vetor em que cada item (índice) do vetor é referente ao intervalo de tempo da coleta. Por exemplo, o valor presente em cada item do vetor representa a quantidade de pacotes recebidos por minuto. O primeiro item do vetor armazena a quantidade de pacotes recebidos pela rede no primeiro minuto da coleta e assim por diante.

As novas características são criadas quando a abordagem ESPA processa a coleta do tráfego de rede aplicando a teoria dos sinais precoces de alerta sobre os dados coletados na rede e pré-processados (Etapa 3). Por exemplo, a abordagem calcula o *kurtosis*, o *skewness* e CV para o vetor com quantidade de pacotes recebidos por minuto. Assim, a abordagem cria três novas características, o *kurtosis*, o *skewness* e o CV da quantidade de pacotes recebidos. Esse processo é repetido a uma nova coleta de dados. O administrador de rede pode escolher qual sinal precoce de alerta será aplicado em cada característica. Isso diminui a quantidade de novas características criadas pela abordagem e simplifica a interpretação dos resultados. Para eliminar tendências errôneas e usar a solução proposta ao longo do tempo, este trabalho aplica a ESPA utilizando o conceito de janela deslizante de tamanho fixo [Bury et al. 2020]. Na Etapa 4, as características são usadas para a predição dos ataques DDoS. A Seção 4 apresenta a predição de ataques DDoS utilizando as novas características e o K-means, um algoritmo de AM não supervisionado. Além disso, as características são apresentadas graficamente para melhor compreensão dos resultados.

## 4. Avaliação

Este artigo avaliou a proposta em três experimentos. Utilizou-se as capturas 51 e 52 do *dataset* CTU-13, correspondentes a rede local, e o *dataset* CIC-DDoS2019, onde a Internet conecta a vítima aos atacantes. Esses *datasets* rotulam o início de um ataque DDoS e os *bots*. A definição do início do ataque auxilia na avaliação que utiliza o tráfego de rede anterior ao início do ataque. Através do rótulo dos *bots*, pode-se verificar se a abordagem ESPA identificou indícios da preparação do ataque antes do seu lançamento. Por fim, os *datasets* apresentam algum registro de comunicação entre os *bots* ou alguma ação dos atacantes anteriores ao ataque, como a infecção do *bots* ou testes de ataque.

A primeira característica utilizada por este trabalho é a quantidade de pacotes por segundo, uma das características mais relevantes segundo Feng *et al.* (2018). Além disso, é razoável afirmar que se os atacantes testarem os ataques [Jaber et al. 2017] a quantidade de pacotes trafegados pode variar. Para contabilizar essa característica, contou-se quantos pacotes a rede trafegou em cada segundo. A quantidade de endereços de Protocolo da Internet (do inglês, *Internet Protocol* - IP) na origem e no destino dos pacotes representam as outras duas características do tráfego de rede usadas. Para definir a quantidade de endereços IPs de origem, contaram-se quantos endereços únicos enviaram pacotes através do campo endereço de origem do pacote IP. A quantidade de endereços de destino baseia-se na contagem de endereços IPs únicos existentes no campo de destino do pacote IP. Este trabalho escolheu essas características pois a falsificação de endereços IP é uma prática comum em ataques DDoS [Jyoti and Behal 2021]. Então, a quantidade de endereços IPs que enviam pacotes antes do ataque apresenta potencial para ser utilizada pela abordagem ESPA, pois a preparação do ataque pode causar variações nessa característica. É possível afirmar que a varredura de dispositivos vulneráveis pode impactar na quantidade de endereços IP de destino. Este trabalho utilizou apenas características presentes no cabeçalho dos pacotes para preservar a privacidade. Para ressaltar a relevância dos resultados, as avaliações da ESPA não utilizaram estratégia de pré-processamento dos dados definidos na Subseção 3.2. Contudo utilizá-las pode melhorar os resultados das predições.

O tráfego de rede coletado na forma da contagem dos atributos foi agrupado usando o intervalo de tempo de um segundo em todos os três experimentos. Assim, a cada segundo, a abordagem extrai as características do tráfego de rede e as armazena em um vetor dedicado para cada uma delas. A quantidade de tráfego malicioso antes do início do ataque presente nos *datasets* seria inexpressiva para realizar a predição de ataque DDoS caso o intervalo de tempo fosse exacerbado. Além disso, grandes intervalos de tempo iriam comprometer a avaliação da captura 52 da CTU-13. Por exemplo, caso o intervalo de tempo fosse um minuto, o experimento não teria dados suficientes para avaliar essa captura. Pois os vetores de características do tráfego de rede possuiriam 17 itens. Assim, analisar apenas 17 intervalos não apresentaria um resultado conclusivo.

Após a definição dos *datasets* e das características do tráfego de rede, este trabalho executou a abordagem ESPA. A execução da abordagem consiste em calcular os sinais precoces de alerta *Kurtosis*, *Skewness* e CV (Subseção 3.1) utilizando uma janela móvel de tamanho fixo. Com o intuito de melhorar a explicabilidade dos resultados, calculou-se apenas um sinal precoce de alerta para cada característica do tráfego de rede. Portanto, o *Kurtosis* foi calculado para a quantidade de endereços IP de origem, o *Skewness* foi calculado para a quantidade de endereços IP de destino e o CV foi calculado para o total de pacotes. Essa definição foi realizada empiricamente e, no futuro, receberá novos trabalhos para automatizar esse processo. Todos os resultados da abordagem ESPA estão disponíveis online[1]. O tamanho janela móvel de tamanho compreende 5% de cada *dataset*. A literatura ainda não é unânime sobre esse valor, por exemplo, Bury *et al.* (2020) utilizaram 40% do *dataset*. Dessa forma, o tamanho da janela móvel deslizante ainda carece de investigações. Isso não apenas na área de redes, mas também em outras. Este trabalho definiu 5% com o intuito de maximizar o tempo de predição dos ataques DDoS, pois ao utilizar um valor grande a predição poderia ser atrasada. Contudo, um valor demasiadamente pequeno pode não ser significativo devido à pouca informação disponível.

Como a abordagem ESPA aplica os sinais precoces de alerta para distinguir a preparação dos ataques DDoS criando novas características, utilizou-se o algoritmo K-means para gerar as alertas de ataques futuros. O K-means é um algoritmo de AM não supervisionado que distribui os intervalos de tempo em um espaço com a quantidade de dimensões igual às novas características: *Kurtosis* do total de endereços IPs de origem, *Skewness* do total de endereços IPs de destino e o CV do total de pacotes. Após a distribuição, o algoritmo define K pontos que serão considerados os centros dos agrupamentos. O valor de K é dois pois o objetivo é separar os sinais da preparação do ataque do tráfego normal. Os intervalos com menor distância Euclidiana para cada centro são agrupados formando os dois grupos. Após todos os intervalos serem agrupados, um novo centro é calculado

[1] `https://github.com/andersonneira/wgrs-data-2023`

utilizando a média de todos os intervalos pertencentes a cada agrupamento. Esse processo é repetido até que os centros não mudem. O K-means foi escolhido pois não usa dados rotulados para predizer ataques, simplificando a adoção proposta em ambientes reais.

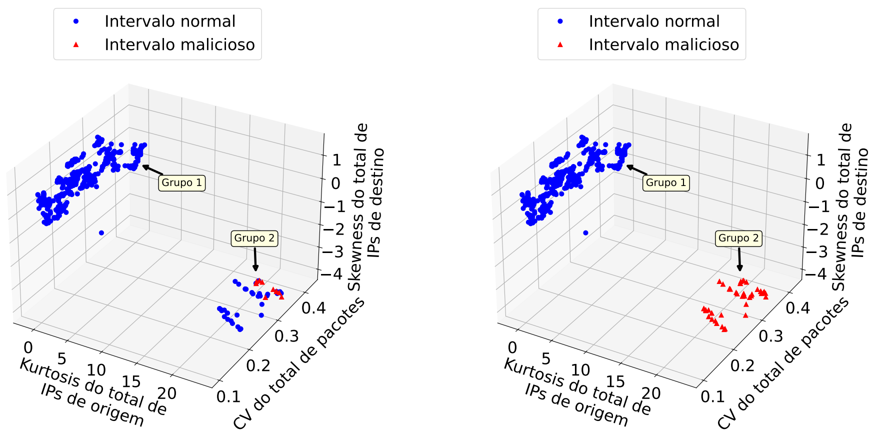
A avaliação dos resultados utiliza a acurácia, precisão e o *recall*. Para defini-los, é necessário medir o número total de verdadeiros positivos ($VP$) e verdadeiros negativos ($VN$), o total de falsos positivos ($FP$) e falsos negativos ($FN$), e o total de observações ($N$). A acurácia avalia as classificações do sistema (Eq. 5). No entanto, existem poucos sinais de preparação para o ataque, pois os invasores ocultam suas ações. A precisão e o *recall* são usados para complementar a análise dos resultados. A precisão indica a relação entre as observações rotuladas pelo sistema para um tipo específico (classes positiva ou negativa) e quantas eram do tipo assumido (Eq. 6). O *recall* apresenta a relação entre todas as observações esperadas do tipo específico e quantas observações desse tipo o sistema classificou corretamente (Eq. 7). Como a quantidade de amostras nas classes varia muito devido ao desbalanceamento inerente ao ataque DDoS e a possibilidade de obter a precisão e o *recall* para as classes positivas e negativas, este trabalho usa a precisão e o *recall* médios ponderados pela quantidade de amostras de cada classe.

$$Acurácia = \frac{VP + VN}{N} \quad (5) \qquad Precisão = \frac{VP}{VP + FP} \quad (6) \qquad Recall = \frac{VP}{VP + FN} \quad (7)$$

### 4.1. Experimento 1

O Experimento 1 usou o tráfego de rede da captura 52 do *dataset* CTU-13. A captura possui 972 segundos, 555 MB, 6.336.398 pacotes, um ataque do tipo inundação de *Internet Control Message Protocol* (ICMP) e três *bots*. O ataque foi conduzido por pesquisadores, porém eles também capturaram dados reais da universidade Tcheca. Os pesquisadores infectaram os *bots* no segundo 527 e lançaram o ataque no segundo 778 da captura. A Figura 3(a) apresenta o resultado da aplicação da abordagem usando a captura 52 no intervalo entre início da captura e o segundo 542. Essa figura apresenta a variação do *Kurtosis* para a quantidade de IPs de origem, o *Skewness* para a quantidade de IPs de destino e o CV para o total de pacotes medida a cada intervalo de um segundo. A abordagem ESPA identificou mudança no comportamento dos sinais precoces de alerta e isso acarretou a existência de dois grupos na visualização dos dados. O Grupo 1 possui apenas intervalos de tempo normais, onde não existem *bots* enviando pacotes. O Grupo 2 possui 48 intervalos onde 38 são normais e 10 maliciosos, onde existe comunicação dos *bots*. Por fim, destaca-se que o Grupo 2 começou a ser formado no fim do processo de inicialização das máquinas virtuais que hospedavam os *bots*.

O K-means aplicado sobre os resultados da abordagem ESPA (Figura 3(b)) identificou dois grupos. O Grupo 2 possui a menor quantidade de intervalos e os mais recentes (próximos ao segundo 542). Considerando o Grupo 2 como intervalos maliciosos, a abordagem ESPA distingue sinais da preparação de um ataque DDoS e consequentemente o K-means pode automatizar a predição do ataque. Considerando os intervalos do Grupo 1 como normais, a proposta classifica corretamente 494 intervalos. Tomando os intervalos do Grupo 2 como malicioso, a proposta identifica corretamente 10 intervalos maliciosos e rotula incorretamente 38 intervalos normais. A acurácia foi de 92,9%, a precisão média ponderada de 98,54% e o *recall* médio ponderado de 92,99%.




(a) Rótulos reais (b) Predição da proposta

**Figura 3. Predição de ataques DDoS no Experimento 1**

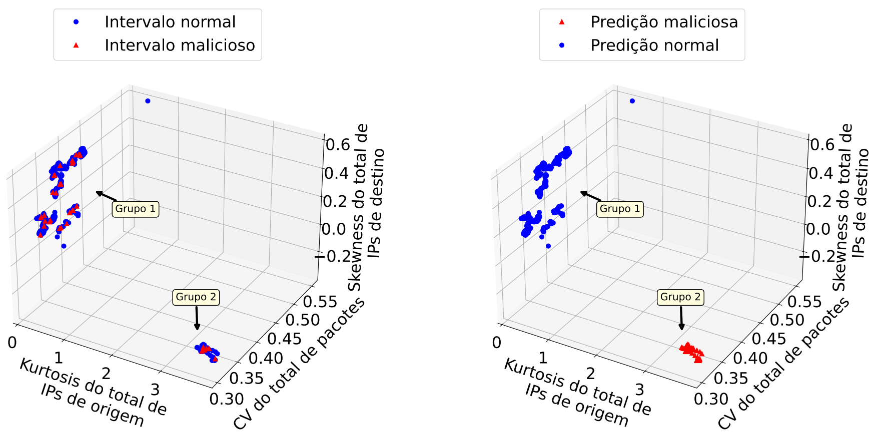
### 4.2. Experimento 2

O Experimento 2 usou o tráfego de rede coletado na captura 51 disponibilizado pelo *dataset* CTU-13. A captura possui 8803 segundos, 41 GB, 46.997.342 pacotes, ataques do tipo inundação de ICMP e User Datagram Protocol (UDP) e 10 *bots*. Os ataques apresentados nesta captura foram conduzidos por pesquisadores, porém eles também capturaram dados reais da universidade Tcheca. Os pesquisadores infectaram os *bots* no segundo 2643 e eles lançaram os ataques no segundo 5632 da captura. A Figura 4(a) apresenta o resultado da abordagem ESPA entre os segundos 3294 e 3794 do *dataset*. Como no Experimento 1, a figura apresenta a existência de dois grupos de intervalos com resultados de abordagem ESPA diferentes. Contudo, diferentemente do experimento anterior, ambos os grupos possuem intervalos normais, ou seja, onde há tráfego oriundo dos *bots* pois os *bots* estão mais ativos próximos ao início do ataque.

Aplicando a mesma metodologia do Experimento 1, o resultado da clusterização do K-means corrobora a existência de dois grupos de intervalos. A Figura 4(b) apresenta os resultados do K-means considerando o grupo com menos intervalos e mais recente como tráfego malicioso. O Grupo 1 possui 413 intervalos normais corretamente identificados e 32 intervalos maliciosos erroneamente considerados normais. O Grupo 2 agrega 10 intervalos maliciosos corretamente identificados como maliciosos e 45 intervalos normais erroneamente considerados maliciosos. Esses resultados indicam uma acurácia de 84,6%, uma precisão média ponderada de 86,54% e um *recall* médio ponderado de 84,60%.

### 4.3. Experimento 3

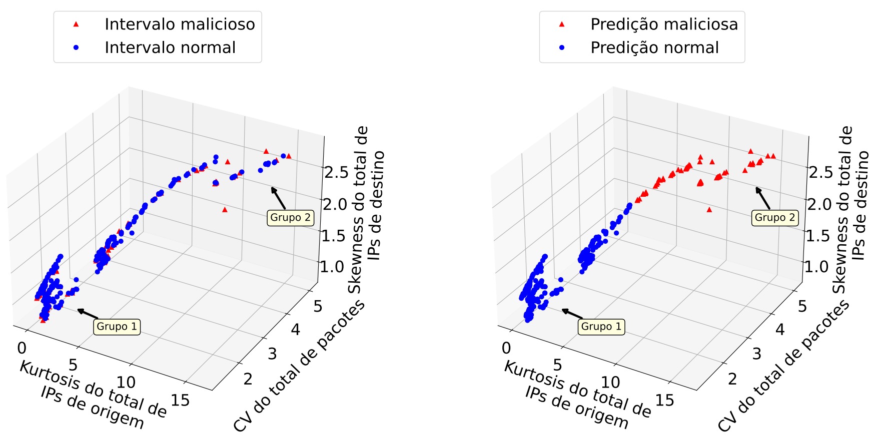
O Experimento 3 utilizou o *dataset* CIC-DDoS2019 que possui 19 ataques DDoS lançados pelos pesquisadores em dois dias. O *dataset* possui 27 GB de dados referente aos ataques e dados reais e 61.407.883 de pacotes. A documentação indica o momento de início e fim dos ataques. A avaliação da abordagem concentrou-se no primeiro ataque DDoS realizado no primeiro dia da captura. O ataque começou no segundo 1484 da captura e durou 540 segundos. A Figura 5(a) apresenta o resultado da abordagem ESPA aplicada sobre intervalos capturados no *dataset* até o segundo 582. Apesar de não ser tão claro quanto nos experimentos anteriores, o resultado da abordagem ESPA apresenta o mesmo




(a) Rótulos reais (b) Predição da proposta

**Figura 4. Predição de ataques DDoS no Experimento 2**

comportamento dos experimentos anteriores. Existem dois grupos onde alguns intervalos normais e maliciosos se combinam. O Grupo 2 da Figura 5(a) possui grande concentração de intervalos maliciosos e os dados mais recentes (próximo do segundo 582). O K-means aplicado sobre os resultados (Figura 5(b)) da abordagem ESPA reforça a separação dos dados. Como nos experimentos anteriores, considerando o grupo menor e mais recente como malicioso, a acurácia é de 69,2%, a precisão média ponderada é de 63,35% e o *recall* médio ponderado é de 69,24%.




(a) Rótulos reais (b) Predição da proposta

**Figura 5. Predição de ataques DDoS no Experimento 3**

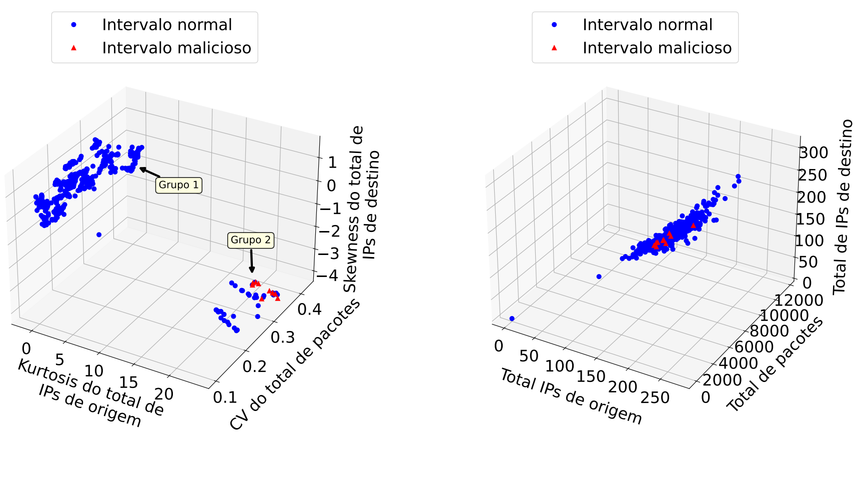
## 4.4. Discussão dos resultados

No Experimento 1, a proposta possibilitou a predição do ataque DDoS 3 minutos e 56 segundos antes do início do ataque com acurácia de 98,54%. Esse resultado é relevante pois a base possui cerca de 16 minutos de captura. A proposta ESPA obteve esses resultados apenas 15 segundos após o início da infecção. Como os atacantes escondem a preparação dos ataques, a quantidade de pacotes originada pelos *bots* é menor do que a quantidade de pacotes oriundos dos dispositivos normais da rede. Assim, a proposta proporcionou a predição do ataque em um cenário desbalanceado. O Experimento 2 também apresentou resultados relevantes, pois em um cenário com mais tráfego de rede e mais tempo de captura e com ataques DDoS diferentes, a proposta proporcionou a predição do primeiro ataque. A documentação indica que a infecção dos *bots* ocorreu 49 minutos e 40 segundos antes do início do ataque. A abordagem ESPA identificou sinais da preparação do ataque 30 minutos antes do seu início e 19 minutos após as infecções. O Experimento 3 apresentou resultados inferiores comparados aos anteriores. Contudo, a abordagem ESPA identificou o mesmo comportamento nos três experimentos. Antes do início do ataque, os sinais precoces de alerta processados pela abordagem apresentam uma mudança simbolizada pela existência de dois grupos. Nos três experimentos, um grupo é majoritariamente composto por intervalos de dados onde os *bots* não atuam, e o outro é formado por segundos com e sem a atuação dos *bots*. Mesmo que a divisão dos dois grupos não fosse tão clara quanto nos outros, a proposta deste trabalho proporcionou a predição do ataque DDoS com 15 minutos antes de seu lançamento. Por fim, a Tabela 1 resume os resultados.

**Tabela 1. Resultados dos experimentos.**

| Experimento | Acurácia | Precisão | *Recall* | Tempo de predição |
|---|---|---|---|---|
| Experimento 1 | **92,9%** | **98,54%** | **92,99%** | 3 minutos e 56 segundos |
| Experimento 2 | 84,6% | 86,54% | 84,60% | **30 minutos e 00 segundos** |
| Experimento 3 | 69,2% | 63,35% | 69,24% | 15 minutos e 00 segundos |

A Figura 6 apresenta a comparação do resultado da abordagem contra os dados originais. Na Figura 6(a), o resultado da abordagem ESPA separa o conjunto de intervalos com pacotes originados por *bots* da maioria dos segundos sem tráfego malicioso. Enquanto na Figura 6(b), os atacantes conseguem esconder seu comportamento com o tráfego de rede normal. Aplicar o K-means nos dados originais não separa a preparação do ataque do tráfego normal. Neste caso, nenhum intervalo malicioso é identificado, enquanto a proposta deste trabalho identificou 10. O mesmo acontece nos outros experimentos. A proposta identificou corretamente nos Experimentos 2 e 3 respectivamente 10 e 25 intervalos maliciosos que foram utilizados para predizer os ataques. Enquanto sem a abordagem proposta, o resultado seria 3 e 4 respectivamente para os Experimentos 2 e 3. Assim, os resultados apresentados neste trabalho demonstram o mérito da proposta. Também é importante citar que o CV é o sinal precoce de alerta que menos impacta nos resultados. Contudo, o CV contribui no aspecto visual da geração dos dados. Isso favorece a interpretação dos resultados e é benéfico para a gerência de redes.




(a) Resultados com a abordagem ESPA (b) Resultados sem a abordagem ESPA

**Figura 6. Comparação da abordagem com os dados originais no Experimento 1**

A abordagem ESPA evolui a literatura de três formas. Primeiramente, ela aumenta o tempo de predição em relação à literatura. Em Rahal *et al.* (2020), os autores predizem o ataque DDoS na captura 51 do CTU-13 com 5 minutos e 41 segundos de antecedência, enquanto a proposta ESPA melhora os resultados predizendo o mesmo ataque com 30 minutos de antecedência. A presente proposta também evolui trabalhos anteriores melhorando o tempo de predição. Na captura 52 da CTU-13, o tempo de predição de Neira *et al.* (2022) foi de 3 minutos e 22 segundos, enquanto a atual proposta proporcionou a predição do ataque com 3 minutos e 56 segundos de antecedência. Além disso, a abordagem ESPA atingiu os resultados sem a utilização de AM supervisionado, ao contrário de *Neira et al.* (2022). O AM não-supervisionado não utiliza rótulos para atingir os resultados apresentados. Isto simplifica a aplicação da proposta em ambientes reais. Por fim, a explicabilidade dos resultados é uma importante evolução em relação à literatura. As figuras em três dimensões apresentam o tráfego da rede evoluindo e formando um novo grupo. Intuitivamente, o administrador de rede pode considerar que o novo grupo provém de um comportamento diferente do usual. Portanto, esse novo grupo merece atenção.

## 5. Conclusão

A detecção de ataques DDoS não é suficiente para evitar os prejuízos causados pelos ataques DDoS. Este trabalho apresenta a predição de ataques DDoS baseados na abordagem ESPA. A abordagem processa o tráfego de rede em busca de sinais da preparação dos ataques. Isso proporciona a predição de ataques DDoS sem a utilização de rótulos. Os resultados da avaliação indicam que a proposta deste trabalho prediz diferentes ataques com até 30 minutos antes do início e com uma acurácia máxima de 92,99%. Os trabalhos futuros avaliarão algoritmos como o DBSCAN, o BIRCH e o uso de métricas como a coesão e separação para minimizar os erros de predição. Além disso, trabalhos futuros focarão em melhorias na abordagem ESPA, como a seleção de características, seleção de sinais precoces de alerta para melhorar os resultados obtidos sem prejudicar a explicabilidade.


## Agradecimentos

Este trabalho foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), bolsas #309129/2017-6 e #432204/2018-0, pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), bolsas #2018/23098-0 e #2022/06840-0, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), bolsas #88887.501287/2020-00 e #88887.509309/2020-00.


## Referências


Abdlhamed, M., Kifayat, K., Shi, Q., and Hurst, W. (2017). *Intrusion Prediction Systems*, pages 155–174. Springer International Publishing, Cham.

Bedeian, A. G. and Mossholder, K. W. (2000). On the use of the coefficient of variation as a measure of diversity. *ORM*, 3(3):285–297.

Biggs, R., Carpenter, S. R., and Brock, W. A. (2009). Turning back from the brink: Detecting an impending regime shift in time to avert it. *PNAS*, 106(3):826–831.

Boers, N. and Rypdal, M. (2021). Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point. *PNAS*, 118(21).

Bury, T. M., Bauch, C. T., and Anand, M. (2020). Detecting and distinguishing tipping points using spectral early warning signals. *J. R. Soc.*, 17(170).

Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., and Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLOS ONE*, 7(7):1–20.

de Neira, A. B., de Araujo, A. M., and Nogueira, M. (2022). An intelligent system for DDoS attack prediction based on early warning signals. *IEEE TNSM*, -(-):1–13.

Feng, Y., Akiyama, H., Lu, L., and Sakurai, K. (2018). Feature selection for machine learning-based early detection of distributed cyber attacks. In *DASC*, pages 173–180, Greece. IEEE.

Garcia, S., Grill, M., Stiborek, J., and Zunino, A. (2014). An empirical comparison of botnet detection methods. *Computers & Security*, 45:100–123.

Gupta, B. B. and Badve, O. P. (2017). Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a cloud computing environment. *NCA*, 28(12):3655–3682.

Guttal, V. and Jayaprakash, C. (2008). Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology Letters*, 11(5):450–460.

Jaber, A. N., Zolkipli, M. F., Majid, M. A., and Anwar, S. (2017). Methods for preventing distributed denial of service attacks in cloud computing. *ASL*, 23(6):5282–5285.

Jyoti, N. and Behal, S. (2021). A meta-evaluation of machine learning techniques for detection of DDoS attacks. In *INDIACom*, pages 522–526, New Delhi, India. IEEE.

Kivalov, S. and Strelkovskaya, I. (2022). Detection and prediction of DDoS cyber attacks using spline functions. In *TCSET*, pages 710–713, Ukraine.

Netscout (2022). Issue 9: Findings from 1st half 2022. [Acesso em: 11/22]. www.netscout.com/threatreport/global-highlights.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Stat.*, 8(3):154–168.

Proverbio, D., Kemp, F., Magni, S., and Gonçalves, J. (2022). Performance of early warning signals for disease re-emergence: A case study on COVID-19 data. *PLOS Computational Biology*, 18(3):e1009958.

Rahal, B. M., Santos, A., and Nogueira, M. (2020). A distributed architecture for DDoS prediction and bot detection. *IEEE Access*, 8:159756–159772.

Salemi, H., Rostami, H., Talatian-Azad, S., and Khosravi, M. R. (2021). Leaesn: Predicting DDoS attack in healthcare systems based on lyapunov exponent analysis and echo state neural networks. *MTA*, -(-):1–22.

Santos, L. A. F., Campiolo, R., Gerosa, M. A., and Batista, D. M. (2013). Extração de alertas de segurança postados em mensagens de redes sociais. In *SBRC*, pages 791–804, Brasil.

Scheffer, M. (2009). *Critical Transitions in Nature and Society*. PUP.

Sharafaldin, I., Lashkari, A. H., Hakak, S., and Ghorbani, A. A. (2019). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *ICCST*.

Takimoto, G. (2009). Early warning signals of demographic regime shifts in invading populations. *Population Ecology*, 51(3):419–426.

Tasa, F. A., Istiqomah, Murti, M. A., and Alinursafa, I. (2022). Classification of earthquake vibrations using the ANN (Artificial Neural Network) algorithm. In *IAICT*, pages 102–107.

Toh, A., Vij, A., and Pasha, S. (2022). Azure DDoS protection—2021 Q3 and Q4 DDoS attack trends. azure.microsoft.com/en-us/blog/azure-ddos-protection-2021-q3-and-q4-ddos-attack-trends/ ([Acesso em: 01/22]).