SIGMA-IP: Sistema Inteligente de Gestão e Monitoramento de Ameaças para Redes IPs

Francisco V. J. Nobre[1], Yago M. da Costa[1], Davi O. Alves[1], Ramon S. Araujo[1], Lyedson S. Rodrigues[1], Antonio B. Neto[1], Lourenço A. Pereira Jr. [2], Rafael L. Gomes[1]

[1]Universidade Estadual do Ceará (UECE), Brasil.

{valderlan.nobre, yago.costa, dav.oliveira, ramon.araujo,

lyedson.silva, mozar.braga}@aluno.uece.br, rafa.lopes@uece.br

[2]Instituto Tecnológico de Aeronáutica (ITA), Brasil.

ljr@ita.br

Resumo. *Cada vez mais tem-se a necessidade de soluções de segurança dinâmicas e adaptativas, onde a abordagem de Inteligência sobre Ameaças Cibernéticas visa coletar, analisar e interpretar informações relevantes sobre ameaças digitais. Dentro deste contexto, este artigo apresenta o SIGMA-IP (Sistema Inteligente de Gestão e Monitoramento de Ameaças), uma solução que gerencia conexões em infraestruturas de rede de maneira autônoma e inteligente a partir de dados sobre ameaças. O SIGMA-IP monitora conexões, analisa informações coletadas de bases de ameaças públicas (tais como AbuseIPDB, VirusTotal, Pulsedive e IPVoid), e usa Aprendizado de Máquina (ML) para classificar as conexões em Blocklist, Allowlist ou Suspect (suspeitos). O comportamento dinâmico do SIGMA-IP habilita atualizações em tempo real e treinamento contínuo do ML, permite uma resposta rápida e robusta a ameaças cibernéticas. Experimentos realizados utilizando um ambiente de rede real indicam que o SIGMA-IP detecta ameaças de forma eficaz dentro de um tempo adequado para mitigar incidentes.*

Abstract. *An increasing need for dynamic and adaptive security solutions has emerged, where the approach of Cyber Threat Intelligence aims to collect, analyze, and interpret relevant information about digital threats. Within this context, this article presents SIGMA-IP (Intelligent Threat Management and Monitoring System), a solution that autonomously and intelligently manages connections in network infrastructures based on threat data. SIGMA-IP monitors connections, analyzes information collected from public threat databases (such as AbuseIPDB, VirusTotal, Pulsedive, and IPVoid), and uses Machine Learning (ML) to classify connections as Blocklist, Allowlist, or Suspect. The dynamic behavior of SIGMA-IP enables real-time updates, and continuous ML training allows for a rapid and robust response to cyber threats. Experiments conducted in a real network environment indicate that SIGMA-IP effectively detects threats within an appropriate timeframe to mitigate incidents.*

1. Introdução

No cenário digital em constante evolução de hoje, a segurança da informação não é apenas uma necessidade, mas uma prerrogativa para proteger dados sensíveis contra o

crescente volume e sofisticação dos ataques cibernéticos. Esta realidade exige soluções de segurança que sejam adaptáveis, resilientes e capazes de antecipar ameaças emergentes [Silveira et al. 2023, Souza et al. 2024], principalmente na tarefa de gerenciamento de conexões de rede. O gerenciamento de conexões de rede refere-se ao controle e supervisão do tráfego de rede entre dispositivos para garantir a segurança, desempenho e disponibilidade da rede, além de otimizar o uso dos recursos disponíveis [Portela et al. 2024b, Ferreira et al. 2024].

A integração de tecnologias avançadas e estratégias proativas é essencial para manter a integridade, a confidencialidade e a disponibilidade dos dados. A aplicação de um Firewall responsivo, que se adapta dinamicamente para bloquear endereços IP de baixa reputação com base em análises de reputação e comportamento, é uma resposta inovadora a essa necessidade, destacando uma evolução na forma como as redes podem ser protegidas. Outra abordagem para segurança da rede é Inteligência sobre Ameaças Cibernéticas (*Threat Intelligence*), que foca na coleta, análise e interpretação de informações sobre ameaças e atividades maliciosas no ambiente digital [Wagner et al. 2019, Portela et al. 2023]. Esta abordagem aplica registros de dados das conexões, que auxilia na identificação de tráfego malicioso. Essa perspectiva sobre as origens dos ataques, destacada pela referência [Komosny 2023], é essencial para implementar medidas de segurança mais eficazes e direcionadas, permitindo respostas mais precisas às ameaças.

Além disso, a capacidade de resposta automática a ameaças identificadas permite ações rápidas e eficazes em um ambiente onde cada segundo é crítico. Técnicas como análise de fluxos de logs em tempo real [Yadav and Mishra 2023] e sistemas de filtragem automática de pacotes [Rizkilina and Rosyid 2022] exemplificam como a tecnologia pode reforçar a segurança, permitindo que firewalls não apenas identifiquem, mas também reajam proativamente a ameaças emergentes. Essas estratégias utilizam o conceito de *blocklists* e a gestão da reputação de IPs são fundamentais para o fortalecimento da segurança de redes, onde as *blocklists* públicas, que catalogam endereços IP conhecidos por atividades maliciosas, são uma ferramenta valiosa para a pré-filtragem de tráfego suspeito.

Essa estratégia pode ser reforçada pela aplicação da técnica Tarpit [Walla and Rossow 2019], uma técnica defensiva que, em vez de bloquear completamente o acesso, degrada a conexão, desestimulando atividades maliciosas sem negar o acesso aos serviços. Esta técnica habilita um equilíbrio de análise, podendo mitigar a ocorrência de ataques ao mesmo tempo que não bloqueia usuários legítimos considerados ameaças de forma equivocada [Costa et al. 2024b]. Contudo, atualmente, estas abordagens conceituais mencionadas (Inteligência sobre Ameaças, Tarpit, resposta automatizada, etc) para gerenciamento de conexões são realizadas por equipes de cibersegurança, que analisam, de forma manual ou semi-automatizada, os dados coletados para identificar ameaças emergentes, vulnerabilidades exploradas, identidades de atores maliciosos e quaisquer outras informações relevantes [Afzaliseresht et al. 2020, Silva et al. 2023]. Desta forma, faz-se necessário desenvolver soluções de segurança que consigam automatizar o processo de análise de ameaças e automatização, incluindo aspectos de eficiência, escalabilidade e tempo de resposta.

Dentro deste contexto, este artigo apresenta o $SIGMA-IP$ (Sistema Inteligente de Gestão e Monitoramento de Ameaças), uma solução de segurança para gerenciar co-

nexões de rede com base em dados de Inteligência sobre Ameaças integrado com Aprendizado de Máquina (ML), visando identificar conexões que possam ser uma ameaça para os dispositivos e serviços na rede. O sistema combate ameaças de forma autônoma, atualizando *blocklists* em tempo real (Reputação de IP), aplicando técnicas de degradação (Tarpit) e classificando conexões com Aprendizado de Máquina. O $SIGMA - IP$ coleta dados das seguintes bases públicas: *AbuseIPDB*[1], *VirusTotal*[2], *Pulsedive*[3] e *IPVoid*[4]. O SIGMA-IP analisa conexões e as classifica como maliciosas (*blocklist*), confiáveis (*allowlist*) ou suspeitas (*suspect*) através de Aprendizado de Máquina (onde são analisadas as técnicas *Random Forest*, *SVM*, *Neural Network*, *Extra Trees*, *Decision Tree*, *K-nearest Neighbors* e *Convolutional neural network*).

Foram realizados experimentos utilizando um ambiente de rede real da Universidade Estadual do Ceará (UECE), onde o processo de monitoramento de rede e gerenciamento de conexões foi integrado com *switches* existentes, habilitando uma avaliação realística do desempenho do $SIGMA - IP$, bem como o seu impacto em uma rede de produção e, consequentemente, nos usuários da rede. Os resultados indicam que o $SIGMA - IP$ detecta ameaças de forma eficaz dentro de um tempo adequado para mitigar incidentes.

O restante deste artigo está organizado da seguinte forma. A Seção 2 detalha as soluções existentes para detecção de atividades maliciosas. A Seção 3 descreve o $SIGMA - IP$, enquanto a Seção 4 discute os experimentos realizados e os resultados obtidos. Finalmente, a Seção 5 conclui o artigo e apresenta trabalhos futuros.

2. Trabalhos Relacionados

Esta seção destaca trabalhos acadêmicos e pesquisas relevantes que exploraram diferentes técnicas e estratégias para identificar ameaças e atividades maliciosas. Além disso, faremos uma análise das abordagens existentes, identificando suas vantagens e limitações. Tosun et al. [Tosun et al. 2021] propõem um mecanismo de detecção para identificar fluxos de proxy IPs residenciais (*Residential IP* - RESIP) em dispositivos eletrônicos de consumo comprometidos, nos quais o software de proxy pode operar sem o conhecimento ou consentimento explícito do usuário. Assim, os autores investigam os provedores comerciais de serviços de proxy RESIP e analisam suas práticas de recrutamento de hosts, que frequentemente são suspeitas ou beiram a ilegalidade. Esta proposta é limitada a identificar ameças de IPs residenciais a partir dos fluxos que passam por um proxy.

Lazar et al. [Lazar et al. 2021] propõem o algoritmo IMDoC para detectar domínios maliciosos e associá-los a ações de malware em um ambiente de tráfego DNS em escala real. O método combina informações de arquivos de comunicação maliciosos, extraídos do *VirusTotal*, com padrões de solicitação de DNS observados em um ambiente de produção. Contudo, apesar de utilizar uma base de ameaças, a abordagem é limitada por depender exclusivamente de uma única fonte de dados, carecendo de adaptabilidade a diferentes contextos, restringindo-se à correlação de IPs com campanhas de malware.

Yang e Lim[Yang and Lim 2021] descrevem um método de detecção de tráfego

[1] abuseipdb.com
[2] virustotal.com
[3] pulsedive.com
[4] ipvoid.com

SSL malicioso, que recompõe registros SSL a partir de pacotes IP capturados e inspeciona as características desses registros SSL usando um método de aprendizado profundo. Após a recomposição de um registro SSL a partir de um ou vários pacotes IP, o método extrai o conteúdo não criptografado do registro recomposto e gera uma sequência de dados não criptografados a partir de registros SSL sucessivos para classificação baseada em aprendizado profundo. Similarmente, Wang et al. [Wang et al. 2020] apresentam um sistema de detecção de domínios chamado KSDom, o qual coleta uma grande quantidade de dados de tráfego DNS e dados externos ricos relacionados ao DNS e, em seguida, emprega o método K-means e SMOTE para lidar com os dados desequilibrados. Por fim, o KSDom utiliza o algoritmo de reforço categórico (CatBoost) para identificar domínios maliciosos. Ambas as propostas das referências [Yang and Lim 2021] e [Wang et al. 2020] não consideram bases de dados de ameaças a fim de implantar uma solução adaptável às ameaças. maliciosas.

A solução proposta neste artigo oferece uma abordagem alternativa às existentes, incorporando *Threat Intelligence* de forma integrada com múltiplas fontes de dados de monitoramento de rede. Além disso, ao utilizar técnicas de IA, a solução adapta-se a cenários dinâmicos e considera um espectro mais diversificado de características e comportamentos de IPs maliciosos. Isso difere das abordagens baseadas em *blocklist* e análises menos abrangentes, proporcionando detecção mais precisa e confiável.

3. Proposta

Este artigo apresenta o desenvolvimento e a solução $SIGMA - IP$ (Sistema Inteligente de Gestão e Monitoramento de Ameaças para IPs), uma solução de segurança e gerenciamento para redes corporativas, que evolui a partir da ferramenta $FIBRA$ [Costa et al. 2024a], desenvolvida anteriormente por parte dos autores. O $FIBRA$ é um firewall local automatizado com consultas restritas ao ***AbuseIPB***. O $SIGMA - IP$ diferencia-se ao combinar técnicas de Aprendizado de Máquina supervisionado, uma Interface de Programação de Aplicações (API) personalizada para interação com Firewalls e mecanismos de resposta automática baseados na análise de comportamento e reputação de endereços IP. Além disso, a proposta atual incorpora múltiplas fontes de dados para alimentar o modelo de IA, ampliando sua capacidade de detecção e precisão. O $SIGMA - IP$ representa um avanço significativo em relação aos Firewalls tradicionais, ao incorporar um fluxo automatizado de monitoramento, análise e mitigação de ameaças, garantindo proteção mais robusta e adaptável para as infraestruturas empresariais.

O $SIGMA - IP$ avança o estado da arte, pois as soluções existentes não consideram integrações abrangentes com bases de dados de ameaças combinadas a classificação automatizada baseada em aprendizado de máquina. Assim, as soluções existentes utilizam abordagens reativas e restritas a dados locais, reduzindo sua eficácia em cenários dinâmicos. Por outro lado, o $SIGMA - IP$ incorpora Inteligência sobre Ameaças de múltiplas fontes, oferecendo maior precisão na identificação de atividades maliciosas através de Aprendizado de Máquina.

3.1. Estrutura e Funcionamento do SIGMA-IP

O $SIGMA - IP$ foi projetado para funcionar de forma paralela à infraestrutura de rede, sem afetar seu funcionamento ou desempenho. Estruturalmente, ele utiliza contêineres,

APIs e ferramentas de *Cyber Threat Intelligence* para garantir a escalabilidade, portabilidade e eficiência na comunicação com o Firewall, coleta de dados e processamento do modelo de ML.

Para desenvolvimento e validação, foi utilizado um conjunto de 2.499 endereços IP, coletados e classificados a partir de quatro bases públicas citadas anteriormente. A maioria dos endereços é proveniente de listas de *blocklist* da base *AbuseIPDB*, complementados por endereços de *allowlist* de fontes confiáveis, como Google, Amazon e Microsoft. Esses dados, que compõem a base de *"ground truth"*, permitiram o treinamento e avaliação do sistema, com IPs unanimemente classificados garantindo confiabilidade.

O fluxo de execução inclui a detecção de conexões, armazenamento de dados na API, consulta de *blocklists* e scores de reputação à API do *AbuseIPDB*, *VirusTotal*, *Pulsedive* e *IPVoid*, análise e classificação de comportamento realizada pelo modelo de aprendizagem de máquina, e aplicação de degradação de conexão no gerenciamento do firewall enquanto o $SIGMA-IP$, simultaneamente, retorna uma resposta. Cada módulo foi projetado para operar de forma integrada, formando um sistema coeso, assim como ilustrado na Figura 1. Similarmente, o fluxo de execução é apresentado na Figura 2.

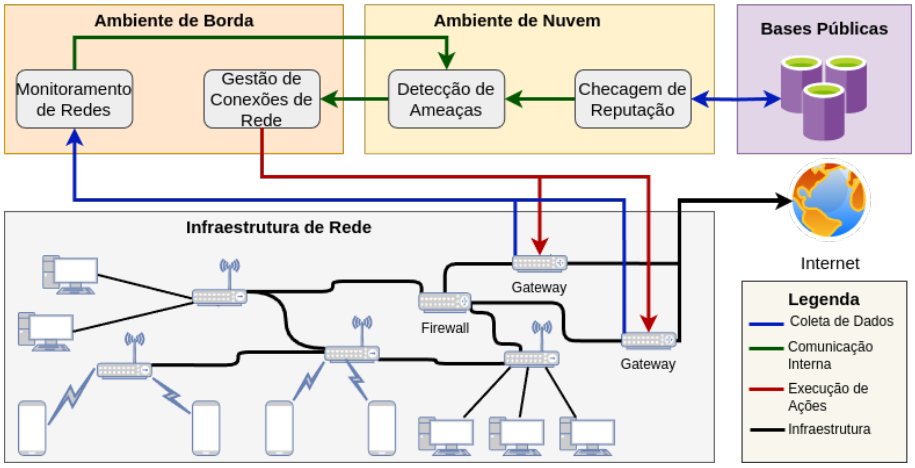


Figura 1. Visão Geral do $SIGMA-IP$.

Dentro desta organização, tem-se os seguintes módulos e funcionalidades:

- **Monitoramento de Rede**: O fluxo começa no momento em que uma nova conexão do tipo TCP sem flag SYN é detectada pelo servidor que hospeda o Firewall, ou seja, um IP abriu conexão com o servidor [Portela et al. 2024a]. Portanto, o módulo escuta e registra as conexões TCP do tipo SYN, que são então verificadas em relação à *allowlist* local do Firewall. Se o IP não estiver na *allowlist*, ele é encaminhado para o Módulo *Detecção de Ameaças* via API do $SIGMA-IP$ para uma análise mais detalhada.
- **Gestão de Conexões de Rede**: Com base nas determinações do Módulo *Detecção de Ameaças*, as tabelas internas de roteamento e as regras de Firewall dos equipamentos de rede são ajustadas. Além disso, duas outras ações são realizadas:
    1. Degradação Temporária de Conexões: Durante o processo de análise, conexões suspeitas sofrem degradação temporária, onde a conexão tem sua prioridade reduzida e sua velocidade limitada. Essa abordagem mantém a operação segura da rede enquanto mitiga possíveis impactos de ameaças.

2. Monitoramento Contínuo e Reavaliação Periódica: Paralelamente a coleta e avaliação por conexão, os IPs nas tabelas *Suspect*, *Allowlist* e *Blocklist* são periodicamente reavaliados com novas consultas aos repositórios de *Threat Intelligence* e nova classificação do modelo a cada: Três dias, semanalmente e mensalmente, respectivamente. Após a análise e classificação final, o $SIGMA-IP$ aplica as seguintes ações: *Blocklist*, IPs permanecem bloqueados por um período indefinido; *Allowlist*, IPs têm acesso liberado permanentemente; e, *Suspect*, IPs suspeitos continuam monitorados, com degradação temporária e reavaliações periódicas.

- **Detecção de Ameaças**: O $SIGMA-IP$ utiliza um modelo de aprendizado de máquina para classificar os IPs, um detalhamento do Modelo de IA usado será apresentado na Seção 3.2. Assim, este módulo executa as seguintes etapas: Coleta de características, dados relevantes das conexões, incluindo reputações externas, coletados pela API; e Classificação, o modelo retorna um dos seguintes resultados *Blocklist* (IP malicioso), *Allowlist* (IP confiável), e *Suspect* (IP suspeito que exige monitoramento contínuo até entrar em um das outras duas categorias). Esta informação é passada para o módulo *Gestão de Conexões de Rede*, o qual executará a ação correspondente (bloqueio de conexão ou Tarpit) nos equipamentos de rede.
- **Checagem de Reputação**: tem por objetivo consultar as bases públicas de ameaças a partir das tentativas de conexão capturadas. Assim, este emprega requisições para consultar endereços IP em *blocklists*, relatórios sobre IPs, geolocalização e pontuação de ações maliciosas. Estas informações irão servir de insumo para o modelo de ML a ser executado pelo Módulo *Detecção de Ameaças*. As bases públicas consultadas pelo $SIGMA - IP$ são: *AbuseIPDB*, Informações de *blocklists* e dados de reputação; *VirusTotal*, Dados sobre comportamentos maliciosos; *Pulsedive*, modelagem de ameaças avançadas; e, e *IPVoid*, detecção de ciberameaças. Essas informações são integradas à tabela *blocklist* da API para análise rápida e futura referência.

O fluxo de execução é apresentado na Figura 2, onde o fluxo de execução do $SIGMA - IP$ inicia com o monitoramento das conexões de rede. Ao detectar uma conexão, verifica se o endereço está bloqueado; se sim, o acesso é negado. Caso contrário, verifica a *allowlist* local: se listado, o acesso é liberado. Para endereços não listados na *allowlist*, o $SIGMA - IP$ registra o endereço na tabela Tarpit *Check* e consulta bases de dados públicas para coletar informações relevantes. Essas informações são então processadas por um modelo de inteligência artificial que classifica o endereço como pertencente a uma das três categorias: *Blocklist*, *Allowlist* ou *Suspect*. Endereços classificados como *Blocklist* têm o acesso negado, enquanto os da *Allowlist* têm o acesso liberado. Já os classificados como *Suspect* recebem um acesso restrito, permanecendo na tabela por 3 dias para monitoramento adicional. A classificação final é enviada ao Firewall, que aplica as regras de acesso em tempo real. Esse processo dinâmico e adaptativo permite ao $SIGMA - IP$ responder rapidamente a ameaças, garantindo maior segurança na rede.

O fluxo modelado do $SIGMA-IP$ é crucial para a segurança da rede, pois combina monitoramento ativo, análise em tempo real e inteligência artificial para identificar, classificar e responder a ameaças de forma automatizada e adaptativa. Ele consulta bases públicas para detectar riscos e aplica regras de acesso rapidamente, garantindo respostas precisas e minimizando falsos positivos. Essa precisão é alcançada por meio do uso de múltiplas fontes de dados e da classificação baseada em aprendizado de máquina, que

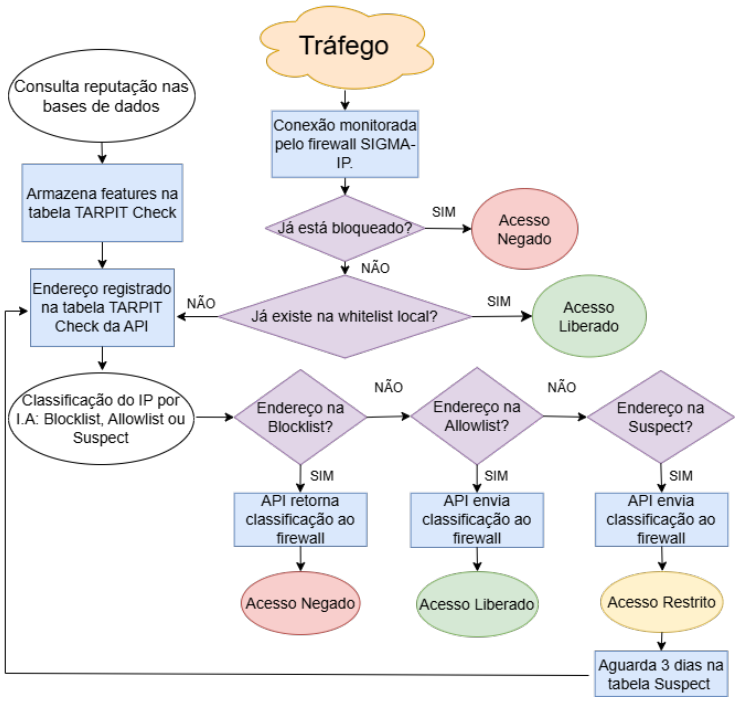


Figura 2. Fluxo de Execução do $SIGMA - IP$.

reduz erros de classificação ao considerar um amplo espectro de características dos IPs. Além disso, sua capacidade de aprendizado contínuo e escalabilidade permite enfrentar novas ameaças com eficiência, mitigando riscos em tempo real e assegurando um ambiente de rede resiliente e protegido.

3.2. Detecção de Ameaças usando Aprendizado de Máquina

O módulo de Detecção de Ameaças realiza a classificação de IPs por meio de aprendizado de máquina, avaliando diversos modelos de ML ao longo do desenvolvimento da proposta. Os modelos de ML foram treinados utilizado um conjunto de dados formado para a proposta que contém informações sobre IPs de diferentes bases de dados. O treinamento é feito na nuvem, porém a execução é realizada na borda. As características utilizadas para treinar os modelos apresentados nesse artigo foram:

- **abuseipdb_confidence_score** (0 a 100): Índice de confiança do AbuseIPDB que mede a probabilidade de um IP ser malicioso, baseado em denúncias anteriores.
- **abuseipdb_total_reports** (0 a 85000): Total de denúncias recebidas pelo IP no AbuseIPDB, indicando histórico de comportamento mal-intencionado.
- **abuseipdb_num_distinct_users** (0 a 2000): Número de usuários distintos que reportaram o IP, refletindo maior credibilidade das denúncias.
- **ipvoid_Detection_Count** (0 a 40): Quantidade de serviços no IPVoid que identificaram o IP como uma ameaça.
- **risk_recommended_pulsedive** (0 a 5): Risco geral consolidado do PulseDive.
- **virustotal_reputation** (-127 a 565): Reputação do IP no VirusTotal (pontuações negativas indicam má reputação).
- **virustotal_malicious** (0 a 20): Quantidade de marcações do IP como ”malicioso”.
- **virustotal_suspicious** (0 a 5): Quantidade de marcações do IP como ”suspeito”.

- **virustotal_undetected** (0 a 91): Quantidade de mecanismos que não detectaram ameaças associadas ao IP.
- **virustotal_harmless** (0 a 86): Mecanismos no VirusTotal que consideraram o IP "inofensivo".

A partir da coleta destes dados, é feito um processo de tratamento (eliminação de linhas que não continham informações de pelo menos uma das bases de dados). Em seguinda, os dados foram normalizados para o intervalo [0, 1] utilizando a técnica de normalização *min-max*, onde cada atributo foi escalado proporcionalmente ao seu valor mínimo e máximo observado. Categorização das características. Em seguida é efetuado o cálculo do ***Score***, que é uma métrica proposta que determinará em qual classe o IP se enquadrará. Quanto maior for o valor, maior será a probabilidade de um IP ser *blocklist*. Desta forma, o ***Score*** irá direcionar a rotulação dos dados para o treinamento dos modelos de ML. Assim, as Seções 3.2.1 e 3.2.2 irão definir o cálculo e o processo de definição do ***Score***.

3.2.1. Cálculo do *Score*

A definição do ***Score*** é baseada nas seguintes métricas em relação aos dados sobre ameaças: Variância, Correlação Média, Importância e a Normalização da Importância. Estas métricas serão descritas a seguir, enfatizando a relevância de cada uma no processo de análise dos dados sobre ameaças.

A Variância (V) serve como um indicador crucial para seleção de características, demonstrando que aquelas com maior variância possuem maior poder discriminativo [Guyon and Elisseeff 2003]. Essa abordagem ajuda a priorizar características que representam maior quantidade de informação útil e, ao mesmo tempo, filtrar características com variância muito baixa, que podem ser irrelevantes ou redundantes. A fórmula de calcular a variância é definida na Equação 1, onde x_i é valor normalizado da característica, $\bar{x}$ é média dos valores normalizados e n é número total de entradas no conjunto de dados.

A Correlação Média ($\bar{r}$) mede o grau de relacionamento entre uma característica i e as demais do conjunto de dados. Essa métrica é fundamental para identificar redundâncias [Hall 1999], já que características altamente correlacionadas entre si tendem a fornecer informações sobrepostas. Ao calcular a média das correlações de uma característica em relação as demais, é possível avaliar o quanto esta é alinhada ou complementar às demais. Essa abordagem contribui para uma análise mais robusta, priorizando características que não apenas possuem variância relevante, mas também trazem novas perspectivas ao conjunto de dados. A Correlação Média $\bar{r}_i$, de uma certa característica i, é definida de acordo com a Equação 2, onde $|R_{i,j}|$ é o valor absoluto do coeficiente de correlação entre as características i e j e m é o número total de características.

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (1) \qquad\qquad \bar{r}_i = \frac{1}{(m-1)}\sum_{j=1\ ;\ j\neq i}^{m}|R_{i,j}| \qquad (2)$$

A Importância é a média aritmética da Variância (V) e Correlação Média ($\bar{r}$), resultando em uma medida única de relevância para cada característica. Essa abordagem

equilibra a capacidade discriminativa da variância com a avaliação de complementaridade/redundância da correlação. Segundo Tang et al. [Tang et al. 2014], combinar diferentes métricas é crucial para uma avaliação mais abrangente das características, proporcionando uma visão balanceada e fundamentada em múltiplos aspectos estatísticos. Embora a variância não tenha limite superior teórico, enquanto a correlação varia entre 0 e 1, na prática, após a normalização dos dados, os valores de variância ficaram em uma faixa comparável à da correlação média. Essa proximidade de escalas justifica o uso da média aritmética para calcular a Importância, evitando que uma métrica domine o cálculo. A Importância é definida de acordo com a Equação 3, onde I_i é a Importância, V_i é a Variância e $\bar{r}_i$ é a correlação média da característica i.

$$I_i = \frac{V_i + \bar{r}_i}{2} \quad (3) \qquad N_i = \frac{I_i}{\sum_{j=1}^{m} I_j} \quad (4) \qquad Score = \sum_{i=1}^{m} n_i \cdot N_i \quad (5)$$

Similarmente, a Normalização da Importância de cada característica é obtido ao normalizar sua Importância em relação ao somatório do das Importâncias de todas as características, transformando-o em um valor proporcional entre 0 e 1, com soma igual a 1. Como métrica complementar, a Normalização da Importância fornece uma visão geral das perspectivas avaliadas e, aliado ao uso de múltiplas métricas [Chandrashekar and Sahin 2014], ajudando a capturar detalhes e aprimorar a seleção de variáveis. A Normalização da Importância é definida de acordo com a Equação 4, onde N_i é a Normalização da Importância da característica i, I_i é a importância da característica i, I_j é a contribuição (ou peso) de cada característica j e m o número total de características.

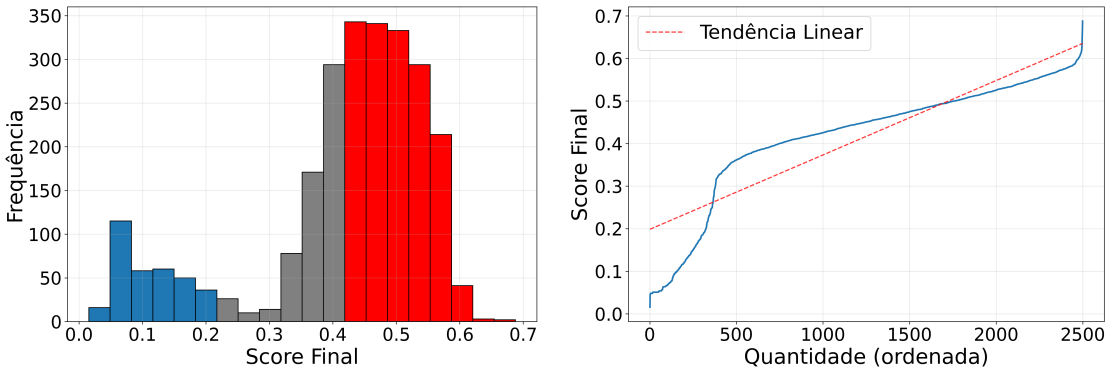
Por fim, o ***Score*** é a soma dos valores ponderados (valor normalizado multiplicado pela Normalização da Importância) de todas as características para uma amostra. Ele reflete a contribuição combinada de todas as características para a classificação. O ***Score*** é definido de acordo com a Equação 5, onde n_i é valor normalizado da característica i, N_i é a normalização da importânciada característica i e m é número total de características.

Essa abordagem de definição do ***Score*** se baseia em métricas estatísticas fundamentais, como variância e correlação, que capturam variabilidade e redundância das características. É escalável, adequado para conjunto de dados com muitas características e amostras, e de fácil interpretação, com pesos finais compreendidos como proporções relativas de importância. Além disso, é flexível, permitindo a inclusão de métricas específicas do domínio, ampliando sua aplicabilidade e confiabilidade.

3.2.2. *Score* para Classificações

Para definir as faixas do ***Score*** em cada classificação, analisou-se seu comportamento em IPs de *blocklist* e *allowlist*. A Figura 3(a) exibe um histograma bimodal, onde as cores indicam diferentes classificações: azul para *allowlist* (*scores* $< 0,2$), vermelho para *blocklist* (*scores* $> 0,4$) e cinza para *suspicious* (*scores* entre $0,2$ e $0,4$). Essas faixas foram definidas a partir da observação de que *scores* abaixo de $0,2$ correspondem a IPs consistentemente *allowlist*, enquanto *scores* acima de $0,4$ indicam IPs unanimemente maliciosos. A faixa intermediária reflete casos de divergência entre bases, com indícios tanto de confiabilidade quanto de risco.

A Figura 3(b),que mostra a distribuição ordenada dos ***Scores***. Observa-se que a curva tem um comportamento não linear, desviando-se da tendência representada pela linha vermelha tracejada. Inicialmente, há um crescimento acentuado até aproximadamente 0,34, seguido por uma região mais estável entre 0,35 e 0,57, correspondente à zona de suspeitos. Após esse intervalo (acima de 0,57), ocorre um novo aumento significativo, refletindo a classificação dos IPs maliciosos. Esta distribuição reflete um cenário típico de segurança, onde existe uma quantidade menor de IPs comprovadamente seguros, uma grande quantidade de IPs potencialmente maliciosos, e uma zona intermediária que requer análise adicional, demonstrando a eficácia do modelo em diferenciar os padrões de comportamento dos IPs analisados.



(a) Histograma do ***Score***. (b) Distribuição do ***Score***.

Figura 3. Análise do *Score*.

Ao final desse processo teremos um conjunto de dados pronto para ser submetido aos modelos de algoritmos de aprendizado para identificar comportamento de novos IPs baseados em suas caracteristícas extraídas das bases de dados de IPs maliciosos. Desse modo, para cada nova conexão que chegar, será realizada uma consulta nesse módulo para determinar em qual classe o IP irá se enquadrar. É válido ressaltar que neste artigo, foram considerados diversas técnicas de ML para analisar o comportamento dos dados : *Random Forest* (RF), *Extra Trees* (ET), *Convolutional neural network* (CNN) *Support Vector Machines* (SVM), *K-Nearest Neighbors* (KNN) e *Neural Networks* (NN). Estes algoritmos têm sido usados em soluções de cibersegurança usando IA [Tosun et al. 2021, Lazar et al. 2021, Costa et al. 2021].

4. Experimentos e Análise de Desempenho

Foram realizados dois experimentos com a finalidade de avaliar a viabilidade do funcionamento da solução proposta em um ambiente real: (A) Eficácia para classificar as ameaças, onde as métricas relacionadas aos modelos de ML são analisadas; e, (B) Aplicação da solução em ambientes de rede, analisando o impacto da ferramenta no desempenho da infraestrutura de rede. É válido ressaltar que o código desenvolvido, bem como os dados utilizados nos experimentos, estão disponíveis no repositório do projeto[5] e com as instruções necessárias para reprodutibilidade.

[5]https://github.com/valderlan/SIGMA-IP-SBRC-2025

4.1. Classificação de Ameaças

No módulo de Detecção de Ameaças é usado um modelo de ML, onde foram avaliadas diversas técnicas, as quais foram utilizados os hiperparâmetros mais apropriados (lista no repositório) para o balanceamento entre complexidade do modelo e capacidade de generalização através da técnica de *Grid Search* com validação cruzada.

Foram avaliadas as métricas: (A) Precisão, determina a capacidade de acertar quais das detecções positivas realmente são positivas; (B) Recall, a eficiência em detectar corretamente a entrada analisada ; (C) F1-Score, a média harmônica entre Precisão e Recall (ou seja, quanto maior a Precisão e Recall, maior será o F1-Score); (D) Acurácia, taxa de classificações corretas; e, (E) Média Harmônica, que representa um conjunto de dados por um único valor através da divisão entre quantidade de elementos e soma do inverso dos elementos do conjunto. Por fim, foi mensurado o tempo de classificação (em segundos), que é tempo gasto pelo modelo de ML para determinar pra qual das três classes (*blocklist*, *allowlist* ou *suspect*) o IP (entrada) em questão faz parte.

Os resultados dos experimentos realizados são apresentados na Tabela 1. Um aspecto interessante dos experimentos foi o desempenho satisfatório de todos os modelos de ML, onde todos os modelos analisados obtiveram uma eficiência superior a 95%. De forma mais específica, o ET destacou-se com a maior acurácia (99,57%), seguido pelo CNN (99,49%) e RF (99,40%). Notavelmente, o tempo de execução variou significativamente entre os modelos, com o RF requerendo $21,29s$, enquanto a DT completou em apenas $0,39s$. Esta variação reflete o *trade-off* entre complexidade computacional e desempenho. As métricas de F1-Score, Precisão e Recall mantiveram-se consistentemente altas em todos os modelos, com o ET apresentando o melhor equilíbrio geral, indicando robustez na classificação tanto de casos positivos quanto negativos.

Tabela 1. Desempenho dos Modelos de ML.

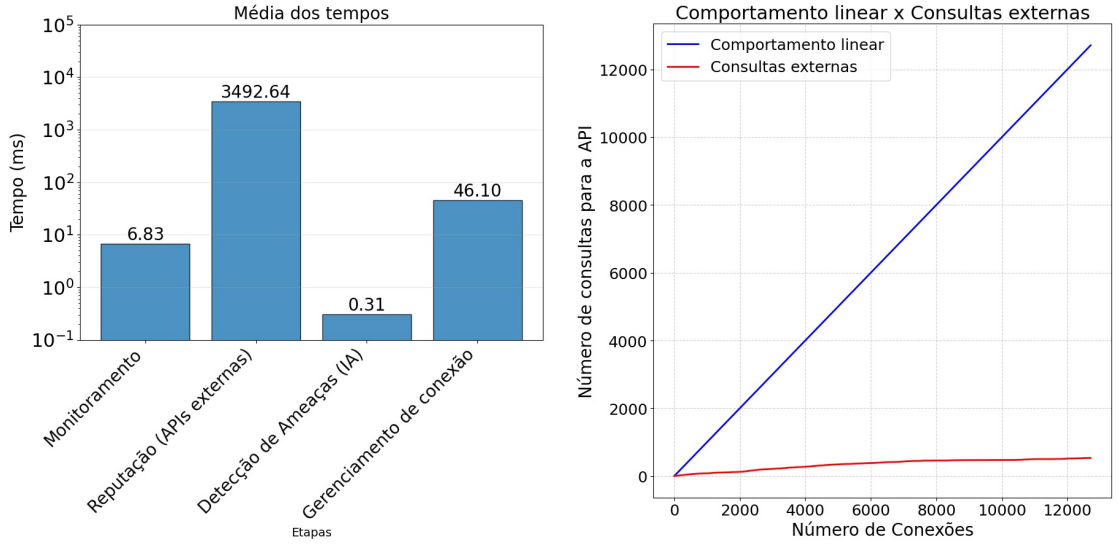
Modelo	Acurácia	F1-Score	Precisão	Recall	Med. Harm.	Tempo (s)
CNN	0,994911	0,994909	0,994915	0,994911	0,994913	5,6558
DT	0,959288	0,959323	0,959560	0,959288	0,959424	0,3944
ET	0,995759	0,995759	0,995779	0,995759	0,995769	2,4931
KNN	0,973707	0,973844	0,974428	0,973707	0,974067	1,1068
NN	0,980492	0,980487	0,981027	0,980492	0,980760	1,9226
RF	0,994063	0,994063	0,994117	0,994063	0,994090	21.2992
SVM	0,987277	0,987316	0,987750	0,987277	0,987513	10,0262

Com base nos experimentos, a ET obteve a melhor eficiência na Detecção de Ameaças. No entanto, seu tempo de execução ($2,49$s) impacta o desempenho da rede. Em cenários reais, onde tempo e eficiência são igualmente importantes, a DT é mais adequada, pois teve o menor tempo de execução ($0,39$s) e ainda alcançou 95,92% de acurácia. Apesar de um desempenho ligeiramente inferior às técnicas mais avançadas, mantém eficiência adequada no cenário que rapidez e eficácia são essenciais.

4.2. Aplicação na Infraestrutura de Rede

O experimento foi estruturado com base na coleta de tráfego do *gateway* (Switch TP-Link TL-SG105E) do LARCES/UECE, utilizando *Port mirroring* para o servidor de borda que executou os quatro módulos do $SIGMA-IP$ (i.e., Monitoramento, Gestão de Conexões,

Detecção de Ameaças e Checagem de Reputação). A seguir, as Figuras 4(a) e 4(b) apresentam, respectivamente, os resultados referentes ao tempo de execução das etapas do $SIGMA-IP$ (medindo o impacto da solução na rede) e o tempo de convergência (tempo necessário para o número de consultas às APIS externas sejam reduzido).



(a) Tempo de execução das etapas.

(b) Comportamento das Consultas externas.

Figura 4. Análise de Desempenho na Rede.

A partir dos dados apresentados na Figura 4(a), percebe-se que o tempo de execução da etapa de monitoramento, que é o realizado com alta frequência, é pequeno (cerca de $7ms$), onde o maior impacto está relacionado a consultas às bases de ameaças, tendo em vista que os IPs são verificados uma única vez, esta etapa não ocorre com alta frequência. Adicionalmente, a partir dos dados da Figura 4(b), nota-se que após o período de convergência (momento em que o uso das consultas externas estabiliza) as verificações ocorrem pontualmente, reduzindo o impacto para o gerenciamento de conexões.

Os resultados indicam que o $SIGMA-IP$ opera de forma eficaz sem comprometer significativamente o desempenho da rede. A integração do sistema com equipamentos de rede existentes foi projetada para minimizar latências e sobrecargas. O uso de técnicas como Tarpit garantiu que a operação e continuidade segura da rede fossem mantidas, mesmo diante de possíveis ameaças. Essa abordagem balanceada é crucial para redes corporativas.

5. Conclusão

Em um cenário de ameaças cibernéticas em constante evolução, a necessidade de soluções de segurança dinâmicas e adaptativas é crucial. Nesse contexto, o $SIGMA - IP$ foi concebido como uma solução inovadora para reforçar a segurança e auxiliar no gerenciamento em redes corporativas, integrando inteligência sobre ameaças cibernéticas, aprendizado de máquina e bases de reputação de endereços IP. Este sistema automatizado se destaca por sua capacidade de monitorar e classificar conexões de forma dinâmica, permitindo respostas rápidas e eficazes a comportamentos suspeitos. Como trabalhos futuros, tem-se a perspectiva de aprimorar a integração das técnicas de IA com os dados sobre ameaças para impulsionar o desempenho da classificação de comportamento de IPs.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (N^o 303877/2021-9 e N^o 405976/2022-4) pelo apoio financeiro e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

Afzaliseresht, N., Miao, Y., Michalska, S., Liu, Q., and Wang, H. (2020). From logs to stories: Human-centred data mining for cyber threat intelligence. *IEEE Access*, 8:19089–19099.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Comput. Electr. Eng.*, 40(1):16–28.

Costa, M., Costa, Y., Silva, D., Portela, A., and Gomes, R. (2024a). Gerenciamento de conexões usando firewall automatizado a partir de dados de inteligência sobre ameaças. In *Anais do XXIV Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, pages 815–821, Porto Alegre, RS, Brasil. SBC.

Costa, M. A., Costa, Y. M., Almeida, Y. O., Cardoso, F. J., and Gomes, R. L. (2024b). Connection management using automated firewall based on threat intelligence. In *Proceedings of the 2024 Latin America Networking Conference*, LANC '24, page 32–37, New York, NY, USA. Association for Computing Machinery.

Costa, W. L., Portela, A. L., and Gomes, R. L. (2021). Features-aware ddos detection in heterogeneous smart environments based on fog and cloud computing. *International Journal of Communication Networks and Information Security*, 13(3):491–498.

Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*. IFIP.

Guyon, I. M. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato.

Komosny, D. (2023). Evidential value of country location evidence obtained from ip address geolocation. *PeerJ Comput Sci.*

Lazar, D., Cohen, K., Freund, A., Bartik, A., and Ron, A. (2021). Imdoc: Identification of malicious domain campaigns via dns and communicating files. *IEEE Access*, 9:45242–45258.

Portela, A., Linhares, M. M., Nobre, F. V. J., Menezes, R., Mesquita, M., and Gomes, R. L. (2024a). The role of tcp congestion control in the throughput forecasting. In *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*, LADC '24, page 196–199, New York, NY, USA. Association for Computing Machinery.

Portela, A. L., Menezes, R. A., Costa, W. L., Silveira, M. M., Bittecnourt, L. F., and Gomes, R. L. (2023). Detection of iot devices and network anomalies based on anonymized network traffic. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6.

Portela, A. L. C., Ribeiro, S. E. S. B., Menezes, R. A., de Araujo, T., and Gomes, R. L. (2024b). T-for: An adaptable forecasting model for throughput performance. *IEEE Transactions on Network and Service Management*, pages 1–1.

Rizkilina, T. M. and Rosyid, N. R. (2022). Packet filtering automation system design based on data synchronization on ip profile database using python. *Journal of Internet and Software Engineering (JISE)*, 3:12–19.

Silva, M., Ribeiro, S., Carvalho, V., Cardoso, F., and Gomes, R. L. (2023). Scalable detection of sql injection in cyber physical systems. In *Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing*, LADC '23, page 220–225, New York, NY, USA. Association for Computing Machinery.

Silveira, M. M., Portela, A. L., Menezes, R. A., Souza, M. S., Silva, D. S., Mesquita, M. C., and Gomes, R. L. (2023). Data protection based on searchable encryption and anonymization techniques. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–5.

Souza, M. S., Ribeiro, S. E. S. B., Lima, V. C., Cardoso, F. J., and Gomes, R. L. (2024). Combining regular expressions and machine learning for sql injection detection in urban computing. *Journal of Internet Services and Applications*, 15(1):103–111.

Tang, J., Alelyani, S., and Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, pages 37–64.

Tosun, A., De Donno, M., Dragoni, N., and Fafoutis, X. (2021). Resip host detection: Identification of malicious residential ip proxy flows. In *2021 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6.

Wagner, T. D., Mahbub, K., Palomar, E., and Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers Security*, 87:101589.

Walla, S. and Rossow, C. (2019). Malpity: Automatic identification and exploitation of tarpit vulnerabilities in malware. In *2019 IEEE European Symposium on Security and Privacy (EuroSP)*, pages 590–605.

Wang, Q., Li, L., Jiang, B., Lu, Z., Liu, J., and Jian, S. (2020). Malicious domain detection based on k-means and smote. In *Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part II 20*, pages 468–481. Springer.

Yadav, M. and Mishra, D. S. (2023). Identification of network threats using live log stream analysis. In *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, pages 1–6.

Yang, J. and Lim, H. (2021). Deep learning approach for detecting malicious activities over encrypted secure channels. *IEEE Access*, 9:39229–39244.