Detecção Inteligente de IPs Maliciosos através do Monitoramento de Ameaças

Arthur C. Urbano[1], Yago M. Costa[1], Mariana C. de Paula[1], Ariel L. Portela[1], Ivo A. Pimenta[1], Rafael L. Gomes[1]

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

{arthur.urbano,yago.costa,mariana.cirino,ariel.portela,

ivo.aguiar}@aluno.uece.br,rafa.lopes@uece.br

Resumo. *A crescente sofisticação das ameaças cibernéticas exige que soluções avançadas para proteger a integridade e confidencialidade dos dados. Uma abordagem para lidar com este cenário é Inteligência sobre Ameaças que desempenha um papel crucial, permitindo que empresas e instituições coletem dados sobre possíveis ameaças e, a partir desses dados, possam lidar com incidentes de segurança. Dentro deste contexto, este artigo apresenta uma solução de Inteligência sobre Ameaças baseada em Inteligência Artificial (IA) para prevenção de ameaças cibernéticas através da detecção de Endereços IP maliciosos. O modelo de IA proposto é alimentado através da coleta de dados das bases sobre ameaças (VirusTotal, AbuseIPDB, Shodan, IBM X-Force e AlienVault). Esses dados usados no modelo de IA proposto oferecem indicativos valiosos sobre IPs e domínios suspeitos. Os resultados, utilizando esses dados reais, mostram que a solução consegue detectar ameaças de forma eficaz.*

Abstract. *The increasing sophistication of cyber threats to online services requires that advanced solutions to protect the integrity and confidentiality of data are deployed. One approach to deal with this scenario is Threat Intelligence which plays a crucial role, allowing companies and institutions to collect data on possible threats and, from this data, be able to deal with security incidents. Within this context, this article presents a Threat Intelligence solution based on Artificial Intelligence (AI) for preventing cyber threats through the detection of malicious IP addresses. The proposed AI model is fed through data collection from databases about new threats (VirusTotal, AbuseIPDB, Shodan, IBM X-Force, and AlienVault). These data used in the proposed AI model offer valuable indications about suspicious IPs and domains. The results, using this real data, show that the proposed solution can detect threats effectively.*

1. Introdução

O aumento alarmante de dos ataques cibernéticos às empresas destaca a importância de proteger os dados e serviços online. Uma abordagem para mitigar ataques cibernéticos é Inteligência sobre Ameaças Cibernéticas (*Threat Intelligence*) [Yang and Lim 2021], que desempenha um papel crucial na luta contra as ameaças digitais em constante evolução, permitindo que as organizações se antecipem a ataques, identifiquem suas origens e tomem medidas proativas para mitigar riscos [Moreira et al. 2021, Portela et al. 2024]. Um dos pontos mais importantes neste processo de mitigar riscos é identificar se um IP

que esta acessando os serviços ou dados é malicioso, ou seja, uma potencial ameaça [Tosun et al. 2021, Lazar et al. 2021].

Contudo, atualmente, estas tarefas relacionadas a *Threat Intelligence* são realizadas por equipes de inteligência de ameaças cibernéticas, 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, Portela et al. 2023]. Desta forma, faz-se necessário desenvolver soluções de segurança que consigam automatizar o processo de *Threat Intelligence*, incluindo aspectos de eficiência, escalabilidade e tempo de resposta[Silveira et al. 2023].

Dentro desse contexto, este artigo apresenta uma solução de *Threat Intelligence* intitulada Inteligência artificial para Monitoramento de IPs Maciliosos (IM-IP). IM-IP é uma solução que possui dois componentes: (1) Monitoramento de Bases, uma ferramenta de coleta de dados sobre relatórios de endereços IP reportados como maliciosas, onde as bases VirusTotal[1], AbuseIPDB[2], Shodan[3], IBM X-Force[4] e AlienVault[5] são monitoradas de forma automatizada; e, (2) Detecção de IPs maliciosos, um modelo de IA que analisa os dados coletados pelo monitoramento de bases para identificar potenciais ameaças ao serviços a serem protegidos.

O Monitoramento de Bases habilita o IM-IP a atualizar, de forma automatizada, sobre novas ameças que surgem, bem como desconsiderar endereços IP que foram reportados a um longo tempo ou tem baixo número de relatórios críticos recentes. Um aspecto importante sobre o Monitoramento de Bases é que, embora os relatórios dessas bases possam ter dados similares, as métricas de avaliação de um IP são distintas e exclusivas de cada base. Assim, a ferramenta de Monitoramento de Bases realiza a coleta, processamento e estruturação desses dados que serão usados como entrada para o treinamento dos modelos de IA. Com relação a Detecção de IPs maliciosos, o modelo de IA analisa e correlaciona os dados de maneira mais eficiente, o que possibilita a identificação de padrões e tendências que poderiam ser indetectáveis através de análises manuais. O uso de IA para *Threat Intelligence* traz benefícios como: capacidade de analisar um grande volume de dados, baixo tempo de resposta, alta precisão para detecção de padrões e adaptabilidade a mudanças ao longo do tempo.

O restante deste artigo está organizado da seguinte forma. A seção 2 detalha as soluções existentes para detecção de IPs maliciosos. A seção 3 descreve a solução proposta, enquanto a seção 4 discute os experimentos realizados e os resultados. Finalmente, a seção 5 conclui o artigo e apresenta trabalhos futuros.

2. Trabalhos Relacionados

Lazar et al. [Lazar et al. 2021] desenvolvem um algoritmo para detectar domínios maliciosos e relacioná-los a uma campanha de malware específica em um ambiente de tráfego DNS em escala real, denominado algoritmo de Identificação de Campanhas de Domínio

[1] virustotal.com
[2] abuseipdb.com
[3] shodan.io
[4] exchange.xforce.ibmcloud.com
[5] otx.alienvault.com

Malicioso (IMDoC). Esta proposta apresenta uma estrutura que combina a existência de arquivos de comunicação para os domínios observados e seus padrões de solicitação de DNS em um ambiente de produção real, onde foram usados arquivos de comunicação maliciosos extraídos do VirusTotal. Apesar de usar informações de uma base de ameaças, esta proposta se limita uma base e não apresenta uma solução adaptavél ao contexto, apenas correlacionando IPs com campanhas de malware.

Yang e Lim[Yang and Lim 2021] descrevem um método de detecção de tráfego 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. Desta forma, 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 as ameaças emergentes.

A solução proposta neste artigo oferece uma abordagem alternativa à aquelas existentes na literatura, incorporando *Threat Intelligence* de forma mais integrada com múltiplas fontes de dados. Além disso, ao utilizar técnicas de IA, a solução proposta tem a capacidade de se adaptar a cenários dinâmicos e levar em conta um espectro mais diversificado de características e comportamentos de IPs maliciosos. Isso se distingue das metodologias que dependem principalmente de listas negras e análises menos abrangentes, podendo proporcionar uma detecção mais precisa e confiável de entidades maliciosas.

3. Proposta

Este artigo apresenta o IM-IP, uma solução para *Threat Intelligence* sobre IPs maliciosos baseada em Inteligência artificial e Monitoramento de Bases de Ameaças. O IM-IP utiliza algoritmos de aprendizado de máquina que são integrados com dados de múltiplas fontes de relatórios sobre ameaças. Um dos aspectos inovadores desta abordagem é a ênfase na utilização de metadados dinâmicos dos IPs. Diferentemente das abordagens tradicionais que se baseiam em listas estáticas, essa proposta explora a natureza dinâmica dos dados, especialmente os metadados associados aos IPs.

A fim de atingir seu objetivo de detectar IPs maliciosos, o IM-IP esta estruturado em dois componentes, Monitoramento de Bases e Detecção de IPs Maliciosos, que estão diretamente relacionados, como ilustrado na Figura 1. O Componente Monitoramento de Bases coleta e processa os dados das diversas bases de ameaças, enquanto que o Componente Detecção de IPs Maliciosos é alimentado por esses dados para gerar o modelo de IA que irá fazer a detecção. Esta funcionalidade de detecção de IPs maliciosos pode ser usada por equipes de segurança ou ferramentas de detecção de intrusão. Os dois componentes do IM-IP possuem os seguintes módulos:

- Monitoramento de Bases
    1. Crawlers das Bases: Coleta periodicamente os dados sobre IP maliciosos (em *black list*) das bases de ameaças configuradas. É válido ressaltar que cada base de ameaças possui uma estruturação e conjunto de dados diferentes, assim faz-se necessário implementar um crawler para cada base de ameaça a ser monitorada.
    2. Processamento de Dados de Cada Base: Após a coleta de dados da base de ameaças, é feito um tratamento desses dados brutos, estruturando-os em um formato JSON para uma posterior extração das características.
    3. Extração de Características: O módulo recebe o arquivo JSON gerado a partir da coleta do Crawler e são extraídas as características referentes aquela base de ameaça, onde é aplicada uma abordagem de mapeamento e refinamento para derivar características sobre mudanças comportamentais.
    4. Junção de Características das Bases: Após extraídas as características das diversas bases de ameaças, estas são combinadas em um conjunto único de características que irá alimentar o conjunto de dados de treinamento do componente de Detecção de IPs Maliciosos.
- Detecção de IPs Maliciosos
    1. Conjunto de Dados de Treinamento: Armazena o conjunto de características coletado pelo componente Monitoramento de Bases, atuando como um ponto central de alimentação do treinamento do modelo de IA.
    2. Formação do Vetor de Características: Consulta o conjunto de dados de treinamento, seleciona características e realiza a formação do vetor de características a ser usado como entrada para o treinamento do modelo de IA.
    3. Treinamento do Modelo de IA: O treinamento do modelo de AI engloba a recepção do vetor de características e execução da técnica de IA. Cada técnica de IA aplica uma abordagem distinta para compreender os dados, resultando em uma eficiência singular de cada técnica.
    4. Detecção de IPs Maliciosos: Após o modelo treinado, este é capaz de detectar se um IP suspeito é considerado malicioso ou não. Assim, pode-se se formar um serviço de consulta a ser usado pela equipe de segurança da empresa ou por uma ferramenta de detecção de intrusão.

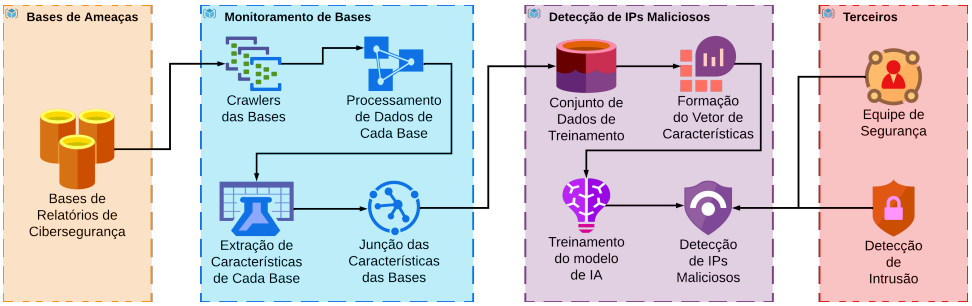


Figura 1. Visão Geral da Solução Proposta.

O IM-IP aplica uma metodologia que combina IA com a fusão de dados de várias fontes de inteligência de ameaças, com um foco especial na utilização de metadados dinâmicos. Esta metodologia visa alcançar uma detecção mais precisa e oportuna de IPs maliciosos, contribuindo assim para uma postura de segurança mais robusta e proativa para as empresas e instituições. A seguir, nas Seções 3.1 e 3.2 são detalhados os componentes Monitoramento de Bases e Detecção de IPs Maliciosos, respectivamente.

3.1. Monitoramento de Bases

Um dos aspectos inovadores desta abordagem é a ênfase na utilização de metadados dinâmicos dos IPs, bem como da fusão de elementos-chave de cinco bases de inteligência sobre ameaças (AbuseIPDB, Shodan, IBM X-Force, AlienVault e VirusTotal), para formar um vetor de 31 características. Diferentemente das abordagens tradicionais que se baseiam em listas estáticas, o IM-IP explora a natureza dinâmica dos dados, especialmente os metadados associados aos IPs, que são consultados via API.

Alguns campos que são comumente parte dos registros, como `abuseipdb_total_reports`, `abuseipdb_num_distinct_users`, e `abuseipdb_last_reported_at`, são mapeados e refinados para derivar características que detectam mudanças comportamentais. Um aspecto distintivo desta proposta é a introdução de características específicas, `delta_abuseipdb_reports`, `delta_abuseipdb_num_distinct_users`, e `delta_abuseipdb_last_reported_at`, para monitorar as mudanças no comportamento dos IPs com base no número de denúncias, possibilitando a captura da natureza dinâmica e as tendências comportamentais dos IPs maliciosos de forma mais eficaz. Ao empregar essas características dinâmicas, a proposta é capaz de analisar o comportamento dos IPs não apenas no momento atual, mas também como eles evoluem ao longo do tempo.

O campo *"abuseipdb_is_whitelisted"* indica se o endereço IP em questão está na lista branca (*whitelisted*), onde um IP na lista branca geralmente é considerado confiável. Contudo, é importante monitorar o valor do campo *"delta_abuseipdb_total_reports"* em conjunto, pois um IP que não está na lista branca e que possui um aumento na quantidade de relatórios pode indicar um comportamento suspeito. O campo *"abuseipdb_confidence_score"* representa a pontuação de confiança associada ao endereço IP, geralmente com base em relatórios e análises históricas. Uma pontuação alta pode indicar que o IP está envolvido em atividades maliciosas. Similarmente, os campos *"abuseipdb_country_code"* e *"virustotal_regional_internet_registry"* estão relacionados à localização geográfica do IP. O código do país e o registro de internet regional são importantes para identificar padrões geográficos em ataques de rede.

"abuseipdb_isp", *"virustotal_as_owner"*, *"ALIENVAULT_asn"* fornecem informações sobre o provedor de serviços de internet e o Sistema Autônomo (AS) ao qual o IP pertence. Essas informações podem ser valiosas para detectar atividades maliciosas associadas a ASNs específicos ou ISPs. Da mesma forma, o campo *"abuseipdb_domain"* indica o domínio associado ao IP, onde essa informação pode ser útil para rastrear atividades de um domínio específico. Os campos *"abuseipdb_total_reports"* e *"abuseipdb_num_distinct_users"* são o número total de relatórios e o número de usuários distintos que relataram o IP. Esses números fornecem uma visão geral da reputação do IP. Assim como os campos *"abuseipdb_last_reported_at"* e *"delta_abuseipdb_last_reported_at"* contêm a data e hora do último relatório e a diferença temporal desde o último relatório. São úteis para entender a frequência de relatórios associados ao IP.

"virustotal_last_analysis_stats" contém estatísticas sobre a última análise feita no IP, incluindo quantas verificações foram inofensivas, maliciosas, etc. Enquanto que os campos *"IBM_score"*, *"IBM_reason"*, *"IBM_reasonDescription"* são campos es-

pecíficos da IBM que fornecem uma pontuação e razão sobre a reputação do IP. Os campos *"SHODAN_asn"*, *"SHODAN_isp"* e *"SHODAN_vulns"* apresentam as informações fornecidas pela base de ameaças SHODAN sobre o IP, incluindo ASN, ISP e vulnerabilidades conhecidas. Na mesma linha, os campos *"ALIENVAULT_reputation"* e *"ALIENVAULT_false_positive"* são informações provenientes da ALIENVAULT sobre a reputação do IP e se ele é conhecido por disparar falsos positivos em sistemas de detecção.

Por fim, os campos *"delta_abuseipdb_total_reports"* e *"delta_abuseipdb_num_distinct_users"* representam a mudança na quantidade de relatórios e no número de usuários distintos que reportaram o IP. Um aumento nessas métricas pode ser um indicativo de atividade suspeita ou maliciosa. Além disso, é importante enfatizar que antes de alimentar os algoritmos de aprendizado de máquina com os dados, uma fase de "digestão"dos dados ou *data engineering* foi necessária. Isso inclui a transformação de dados categóricos em não categóricos. Como os algoritmos de aprendizado de máquina geralmente requerem entradas numéricas, os valores categóricos foram transformados em representações numéricas. Para isso, um conjunto de rótulos numéricos foi criado para cada valor categórico.

3.2. Detecção de IPs Maliciosos

As soluções de *Threat Intelligence* são essenciais na proteção contra ameaças cibernéticas, fornecendo informações sobre potenciais ameaças e ajudando as organizações a se prepararem para enfrentá-las. As bases sobre ameaças são dinamicamente contextualizadas, ou seja, as métricas individuais de avaliação variam de uma base para outra. Esta variação se deve ao fato de que diferentes comunidades e informações alimentam cada base de ameaça. Isso leva a discrepâncias nas informações fornecidas por diferentes fontes e torna mais desafiador para as equipes de segurança terem uma visão unificada e abrangente das possíveis de ameaças. Neste contexto, a aplicação de técnicas de IA sobre os dados das bases de ameaças habilita a correlação de informações de múltiplas fontes e descobrir padrões significativos. Por exemplo, identificar tendências geográficas, tais como o surgimento de ameaças de determinadas localidades, ou monitorar variações em ASN (*Autonomous System Numbers*) e ISPs (*Internet Service Providers*) que podem indicar comportamento malicioso.

No geral, o uso de IA para *Threat Intelligence* traz os seguintes benefícios: (a) Análise de grande volume de dados, o uso de IA pode lidar com grandes volumes de dados em tempo real, processando e analisando informações provenientes das diversas base de dados monitoradas, ação que é difícil para os analistas humanos detectarem manualmente; (b) Tempo de resposta, a partir de técnicas de IA é possível definir modelos para reconhecer indicadores de comprometimento e comportamentos suspeitos, permitindo a detecção antecipada de ameaças e a resposta rápida a incidentes de segurança cibernética, reduzindo o tempo de exposição a ataques; (c) Precisão da detecção, o modelo IA pode identificar ameaças ocultas e sofisticados que podem passar despercebidos pelos sistemas tradicionais de segurança ou agentes humanos, bem como ajudar a reduzir os falsos positivos, filtrando alertas e eventos de segurança que não representam ameaças reais; e, (d) Adaptação, o modelo de IA pode aprender com dados históricos e em tempo real, aprimorando suas habilidades de detecção à medida que são expostos a novos padrões de ameaças que são gerados pela ferramenta de Monitoramento de Bases proposta.

Neste artigo, foram considerados diversos algoritmos de IA para analisar o com-

portamento dos dados [Sarker et al. 2020]: *Random Forest* (RF), *Gradient Boosting* (GB), *Logistic Regression* (LR), *Support Vector Machines* (SVM), *K-Nearest Neighbors* (KNN) e *Neural Networks* (NN). Estes algoritmos tem sido usados em soluções de cibersegurança usando IA [Tosun et al. 2021, Lazar et al. 2021, Costa et al. 2021].

4. Experimentos Realizados

Esta seção apresenta os experimentos realizados para avaliar a solução proposta para detecção de IPs Maliciosos, bem como é crucial destacar a participação humana no fornecimento de *blacklists* e *whitelists* do conjunto de dados formado para o treinamento dos modelos de IA. As Seções 4.1 e 4.2 apresentam a configuração dos experimentos realizados e os resultados desses experimentos, respectivamente.

4.1. Configuração dos Experimentos

A fim de realizar os experimentos foi implementado um protótipo[6] usando diversas tecnologias (tais como FastAPI, Pandas, Numpy, Scikit-learn, MongoDB e Nginx). No experimento, é crucial destacar a participação humana no fornecimento de *blacklists* e *whitelists*, bem como a eficiência do IM-IP na detecção de IPs maliciosos. O protótipo foi configurado para executar requisições 4 vezes ao dia em um rodízio de 250 IPs por dois meses, gerando um conjunto de dados de aproximadamente 60 mil requisições ao longo do período. As *blacklists* de IPs foram oriundas do Tor Project [7], enquanto as *whitelists* consistiam de blocos de redes de empresas conhecidas mundialmente, tais como Google, Apple, e Microsoft. A proporção de IPs maliciosos no conjunto de dados é de cerca de 60%, onde para o treinamento dos modelos foram usados aproximadamente 70% dos registros, enquanto que o restante foi usado para testes.

Assim como descrito na Seção 3.2, foram considerados as seguintes técnicas de IA [Sarker et al. 2020]: RF, GB, LR, SVM, KNN e NN. 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 ; e, (C) F1-Score, a média harmônica entre Precisão e Recall (ou seja, quanto maior a Precisão e Recall, maior será o F1-Score).

4.2. Resultados

4.2.1. Análise dos Dados Coletados das Bases de Ameaças

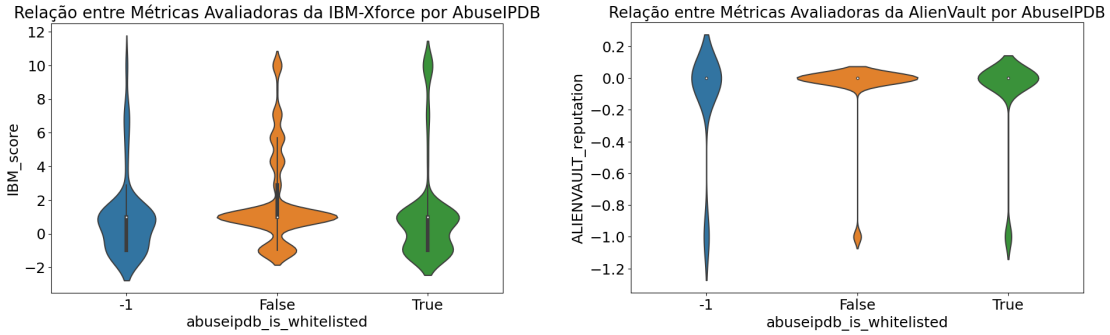
Uma observação crítica das bases de dados revela algumas divergências notáveis. Ao analisar os dados, é essencial estar ciente das divergências entre as bases de ameaça, pois estas divergências podem impactar as conclusões oriundas da análise dos dados. A integração de dados de diferentes fontes requer uma compreensão clara das características de cada conjunto de dados e uma análise cuidadosa para garantir que as conclusões sejam baseadas em informações precisas e confiáveis. Pode-se exemplificar a divergência entre o *IBM_score* e o atributo *abuseipdb_is_whitelisted*, conforme ilustrado na Figura 2(a).

Neste recorte, a variável *abuseipdb_is_whitelisted* parece ter um valor de -1 em certos casos (quando não há dados disponíveis), indicando que, para alguns registros,

[6]https://gitlab.com/ArthurCordeiro/cti-datasource

[7]https://www.torproject.org/about/history/

a base de dados *AbuseIPDB* não contém informações sobre se o endereço IP está ou não na lista branca. Ao mesmo tempo, o atributo *IBM_score* fornece uma pontuação de reputação para os endereços IP com base em critérios da IBM. A discrepância entre esses dois atributos pode ser explicada por diferenças nas metodologias de coleta de dados, critérios de avaliação e frequência de atualização das bases de dados.



(a) Comparação IBM-Xforce x AbuseIPDB. (b) Comparação ALIENVAULT x AbuseIPDB.

Figura 2. Comparações entre as bases de ameaças.

Prosseguindo com a análise das divergências entre as bases de dados, a Figura 2(b) nos permite observar a relação entre o atributo *ALIENVAULT_reputation* e *abuseipdb_is_whitelisted*. Observa-se uma inversão nos padrões quando comparado ao gráfico que avaliou *IBM_score*, sugerindo que as métricas usadas por ALIENVAULT e IBM para avaliar a reputação de endereços IP podem diferir significativamente. Por outro lado, o *IBM_score* pode basear-se em um conjunto de critérios e fontes de dados, *ALIENVAULT_reputation* provavelmente emprega uma metodologia diferente para calcular a reputação de um IP. Por exemplo, se vários IPs maliciosos estiverem concentrados em um Sistemas Autônomo (ASN) ou região geográfica específicos, isso pode indicar um padrão de comportamento que pode ser monitorado para fins de segurança cibernética. Essas tendências podem ser valiosas para organizações e profissionais de segurança que buscam entender as ameaças e proteger suas redes e sistemas. Ao combinar essas informações com os dados de diferentes fontes, é possível criar um perfil mais abrangente e informativo sobre os endereços IP, permitindo uma análise mais aprofundada e informada, que pode contribuir para a identificação e prevenção de ameaças cibernéticas.

Por exemplo, se vários IPs maliciosos estiverem concentrados em um Sistemas Autônomo (ASN) ou região geográfica específicos, isso pode indicar um padrão de comportamento que pode ser monitorado para fins de segurança cibernética. Essas tendências podem ser valiosas para organizações e profissionais de segurança que buscam entender as ameaças e proteger suas redes e sistemas. Ao combinar essas informações com os dados de diferentes fontes, é possível criar um perfil mais abrangente e informativo sobre os endereços IP, permitindo uma análise mais aprofundada e informada, que pode contribuir para a identificação e prevenção de ameaças cibernéticas.

Essas informações são cruciais para identificar endereços IP que podem estar envolvidos em campanhas de ataque em larga escala, onde é provável que afetem um grande número de vítimas. Além disso, estas informações são um indicativo de que a comunidade está ativamente monitorando e reportando tais IPs, o que pode auxiliar no desenvolvimento de estratégias de defesa mais eficazes.

4.2.2. Detecção de IPs Maliciosos

O valores de cada métrica de desempenho analisadas são apresentados na Tabela 1. Essas métricas permitem uma análise minuciosa do desempenho das técnicas de IA, auxiliando na seleção do modelo mais adequado para a tarefa de detecção de IPs maliciosos e fornecendo informações para futuras melhorias e otimizações. Observa-se um desempenho interessante do IM-IP, com uma Precisão variando de 92% a 99%. Esses resultados demonstram que o IM-IP alcançou um alto nível de precisão na tarefa de classificação. A técnica LR mostrou-se eficiente, convergindo após 211 iterações. Por outro lado, a técnica NN teve um processo de convergência mais demorado, com aproximadamente 407 iterações necessárias para atingir a convergência.

Tabela 1. Resultados de Eficiência na Detecção de IPs Maliciosos do IM-IP

Métrica	LR	RF	GB	SVM	KNN	NN
Precisão (%)	0.92	0.96	0.94	0.95	0.98	0.99
Recall (%)	0.89	0.95	0.95	0.88	0.86	0.95
F1-Score (%)	0.92	0.95	0.95	0.91	0.90	0.97

Os resultados obtidos mostram que a técnica NN se apresentou a mais adequada para lidar com o problema de detecção de IPs maliciosos. Os demais algoritmos, RF, GB, SVM e KNN apresentarem valores de Precisão, Recall e F1-Score interessantes, mas consideram as três métricas juntas tiveram um desempenho abaixo que a técnica NN. A análise detalhada das métricas de desempenho, combinada com os números de convergência das técnicas, fornece uma compreensão abrangente da eficácia desses modelos e orienta o desenvolvimento de soluções mais precisas e eficientes no campo de Inteligência sobre Ameaças. Assim, as experimentos realizados neste trabalho proporcionaram uma série de percepções e conclusões importantes. Ao analisar os dados provenientes de diversas bases de ameaças, foi possível identificar tendências e padrões nos comportamentos dos IPs maliciosos. Através da aplicação de técnicas de IA, foi possível construir modelos capazes de classificar com alta precisão se um IP é malicioso ou não.

Foi observado que a combinação de métricas e metadados provenientes das diferentes bases de ameaças contribuiu significativamente para a identificação de IPs maliciosos, onde o uso de informações geográficas, ASNs e provedores de serviços de internet permitiu uma melhor compreensão da origem e distribuição desses IPs.

5. Conclusão

Este artigo descreveu o IM-IP, uma solução de Inteligência sobre Ameaças baseada em IA para detecção de Endereços IP maliciosos, onde o modelo de IA proposto é alimentado através de uma ferramenta de coleta de dados das bases de ameaças. Esta ferramenta habilita a solução a se atualizar de forma automatizada sobre novas ameças que surgem. Esses dados usados no modelo de IA proposto oferecem indicativos valiosos sobre IPs e domínios suspeitos. Os resultados obtidos evidenciaram a eficácia (atingindo uma Precisão de 99%) do IM-IP na detecção de IPs maliciosos com base nas informações disponíveis nas bases de ameaças. Como trabalho futuro, pretende-se a expandir a análise para incluir essas entidades, buscando identificar padrões de comportamento malicioso em diferentes níveis da infraestrutura de rede.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) do Brasil (N^o 303877/2021-9) pelo apoio financeiro.

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.

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.

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.

Moreira, D. A. B., Marques, H. P., Costa, W. L., Celestino, J., Gomes, R. L., and Nogueira, M. (2021). Anomaly detection in smart environments using ai over fog and cloud computing. In *2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC)*, pages 1–2.

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. (2024). T-for: An adaptable forecasting model for throughput performance. *IEEE Transactions on Network and Service Management*, pages 1–1.

Sarker, I. H., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., and Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7:1–29.

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.

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.

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.

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