LABORATÓRIO DE CIBERSEGURANÇA DEFENSIVA E INTELIGÊNCIA ARTIFICIAL

THIAGO JOSÉ LUCAS[1]
[1]*Fatec Ourinhos - Segurança da Informação*
thiago.lucas01@fatec.sp.gov.br

Defensive Cybersecurity and Artificial Intelligence Laboratory

Eixo Tecnológico: *Infraestrutura*

Resumo

A Detecção de Intrusão pela utilização de algoritmos de Inteligência Artificial e Aprendizagem de Máquina é a temática central do Laboratório de Cibersegurança Defensiva. Justifica-se tal tema pelo crescimento constante da ocorrência de ataques a empresas, corporações e governos, denotando a evolução em número e complexidade das ações de *hackers/crackers*. O Objetivo Geral é o desenvolvimento de Sistemas de Detecção de Intrusão que apresentem uma boa relação custo/benefício em termos de custo computacional e acurácia na detecção de ataques, com a menor incidência possível de erros de classificação (falsos-positivo e falsos-negativo). Os Objetivos Específicos consistem em: (a) estabelecer um método de revisão constante e sistemática da literatura para alinhas os detectores de intrusão desenvolvidos com as melhores práticas científicas e mantendo as propostas na busca pela evolução do estado-da-arte; (b) treinar modelos tendo como base algoritmos de Inteligência Artificial; (c) testar modelos com métodos estatísticos robustos e (d) publicar os modelos/métodos/resultados em congressos e periódicos com alto impacto científico. O método consistirá em treinar os Detectores de Intrusão utilizando *datasets* relevantes tendo como base as linguagens Python e R e sua bibliotecas de Aprendizagem de Máquina. Os resultados esperados são publicações dos detectores documentando os modelos de cibersegurança defensiva em congressos e periódicos com alto impacto científico. Alguns resultados parciais já obtidos podem ser observados na Seção de Resultados.

Palavras-chave: *Cibersegurança Defensiva, Sistemas de Detecção de Intrusão, Machine Learning, Inteligência Artificial, Detecção de Ataques.*

Abstract

Intrusion Detection using Artificial Intelligence and Machine Learning algorithms is the central theme of the Defensive Cybersecurity Laboratory. This theme is justified by the constant growth in the occurrence of attacks on companies, corporations and governments, denoting the evolution in number and complexity of hacker/cracker actions. The General Objective is to develop Intrusion Detection Systems that present a good cost/benefit ratio in terms of computational cost and accuracy in detecting attacks, with the lowest possible incidence of classification errors (false positives and false negatives). The Specific Objectives consist of: (a) establishing a method of constant and systematic review of the literature to align the developed intrusion detectors with the best scientific practices and maintaining the proposals in the search for the evolution of the state-of-the-art; (b) training models based on Artificial Intelligence algorithms; (c) test models with robust statistical methods and (d) publish the models/methods/results in conferences and journals with high scientific impact. The method will consist of training the Intrusion Detectors using relevant datasets based on the Python and R languages and their Machine Learning libraries. The expected results are publications of the detectors documenting the defensive cybersecurity models in conferences and journals with high scientific impact. Some partial results already obtained can be seen in the "Resultados" Section.

Key-words: *Defensive Cybersecurity, Intrusion Detection Systems, Machine Learning, Artificial Intelligence, Attack Detection.*

1. Introdução

A segurança da informação se tornou um pilar fundamental na sociedade digital, com a crescente dependência de sistemas computacionais em todos os setores. As redes de computadores e os sistemas de informação armazenam dados confidenciais e críticos, tornando-

os alvos frequentes de ataques cibernéticos. Diante dessa crescente ameaça, os sistemas tradicionais de detecção de intrusão (IDS) baseados em assinaturas não acompanham a sofisticação e a adaptabilidade dos ataques modernos. Neste sentido, os sistemas de detecção de intrusão baseados em *machine learning* são uma importante alternativa para superar as limitações dos métodos tradicionais, dada a sua capacidade de aprender com grandes volumes de dados, identificar padrões complexos e se adaptar a novas ameaças em tempo real. Diversos trabalhos do proponente deste projeto provam a eficiência e a eficácia dos modelos de detecção de intrusão baseados em *machine learning* [6], [7], [8] e [9].

O relatório anual da Cisco [3] de 2023 revela um aumento de 9% no volume de ataques cibernéticos em 2022, com um custo médio global de US$ 4,24 milhões por ataque. Corroborando tais dados, um estudo da Gartner [4] prevê que até 2023, 70% dos ataques bem-sucedidos não serão detectados por IDS baseados em assinaturas.

Faz-se válido destacar as vantagens do Machine Learning (ML) para Detecção de Intrusão:

- Detecção Anomalias: O ML pode identificar comportamentos anormais na rede, mesmo que não sejam conhecidos anteriormente.
- Adaptabilidade: Os modelos de ML podem ser atualizados continuamente com novos dados, aprimorando sua capacidade de detectar novas ameaças.
- Precisão Aprimorada: O ML pode reduzir a taxa de falsos positivos, evitando alertas desnecessários que sobrecarregam os analistas de segurança.

Pesquisas em IDS baseados em ML tem apresentado avanços significativos nos últimos anos [1], [2], [5], [10] e [11]. Diversas técnicas de ML, como redes neurais artificiais, algoritmos de ensemble learning e aprendizado por reforço têm sido aplicadas com sucesso na detecção de intrusão [2], [10] e [11].

Destacam-se, neste ponto, alguns avanços recentes:

- Desenvolvimento de Modelos Eficazes: Modelos de ML mais eficazes e precisos estão sendo desenvolvidos, como redes neurais convolucionais profundas (CNNs) e redes neurais recorrentes (RNNs), que demonstram alta performance na detecção de ataques complexos [10] e [11];
- Integração com Ferramentas de Segurança: A integração de IDS baseados em ML com outras ferramentas de segurança, como firewalls e sistemas de prevenção de intrusão (IPS), permite uma resposta mais rápida e eficaz a ataques cibernéticos [1].

Faz-se necessário também destacar os gargalos e desafios da área:

- Falta de Padronização: Ainda não existe um padrão definido para implementação de IDS baseados em ML, o que dificulta a interoperabilidade e a adoção em larga escala [2];
- Interpretabilidade dos Modelos: A interpretabilidade dos modelos de ML utilizados em IDS é crucial para entender as decisões tomadas pelo sistema e garantir a confiabilidade das detecções [5];
- Atualização de Dados: A atualização frequente dos modelos de ML com dados relevantes e atualizados é essencial para manter a efetividade do sistema contra as últimas ameaças [2].

Quanto à problemática, destaca-se que o aumento da frequência e da sofisticação dos ataques cibernéticos, a ineficiência dos métodos tradicionais de IDS e a necessidade de soluções mais precisas e adaptáveis impulsionam a demanda por sistemas de detecção de intrusão baseados em machine learning. Assim sendo, apresentam-se os Requisitos do Problema:

- Detecção Precisa de Ameaças: O sistema deve ser capaz de detectar com alta precisão diversos tipos de ataques cibernéticos, incluindo ataques conhecidos e desconhecidos.
- Baixa Taxa de Falsos Positivos: O sistema deve minimizar a geração de alertas desnecessários que sobrecarregam os analistas de segurança.

- Adaptabilidade a Novas Ameaças: O sistema deve ser capaz de se adaptar a novas ameaças e ataques emergentes em tempo real.

O Objetivo Geral deste Projeto de RJI é desenvolver sistemas de detecção de intrusão baseados em inteligência artificial/aprendizagem de máquina. Para tal, são definidas três etapas que se completam:

- Etapa 1 (1º ano): desenvolvimento de IDS com métodos individuais de ML;
- Etapa 2 (2º ano): desenvolvimento de IDS com métodos agrupados de ML;
- Etapa 3 (3º ano): desenvolvimento de IDS após seleção de características de ajustes de dados;

Os Objetivos Específicos que se apresentam são:

- Etapa 1: (a) realizar revisões sistemáticas na literatura em relação ao uso de algoritmos individuais de machine learning aplicados à cibersegurança para que sejam extraídas variáveis relevantes de forma a orientar a metodologia específica das pesquisas; (b) implementar os sistemas de detecção de intrusão à luz das variáveis de relevância observada na etapa "a"; (c) testar os sistemas implementados por meio de avaliação específica de modelos de IA; (d) coletar os resultados e documentá-los de forma a produzir elementos científicos que levem o nome do Centro Paula Souza aos principais congressos e revistas científicas da área da ciência da computação.
- Etapa 2: (a) realizar revisões sistemáticas na literatura em relação ao uso de algoritmos agrupados (ensemble learning) de *machine learning* aplicados à cibersegurança para que sejam extraídas variáveis relevantes de forma a orientar a metodologia específica das pesquisas; (b) implementar os sistemas de ensemble learning à luz das variáveis de relevância observada na etapa "a"; (c) testar os sistemas implementados por meio de avaliação específica de modelos de IA; (d) coletar os resultados e documentá-los de forma a produzir elementos científicos que levem o nome do Centro Paula Souza aos principais congressos e revistas científicas da área da ciência da computação.
- Etapa 3: (a) realizar revisões sistemáticas na literatura em relação ao uso de algoritmos de seleção de características (*feature selection*) e de ajuste de dados (*over/undersampling*) aplicados à cibersegurança para que sejam extraídas variáveis relevantes de forma a orientar a metodologia específica das pesquisas; (b) implementar os IDS obtidos após as seleções de características ou ajuste dos dados à luz das variáveis de relevância observada na etapa "a"; (c) testar os sistemas implementados por meio de avaliação específica de modelos de IA; (d) coletar os resultados e documentá-los de forma a produzir elementos científicos que levem o nome do Centro Paula Souza aos principais congressos e revistas científicas da área da ciência da computação.

Pretende-se, por meio do Laboratório objeto deste Projeto, fomentar na comunidade científica a importância da utilização de métodos inteligentes para detecção de intrusão em sistemas computacionais e redes de computadores. Este documento apresenta na Seção 2 detalhes acerca de metodologia; na Seção 3 os resultados esperados e alguns resultados parciais já obtidos e na Seção 4 as considerações finais.

2. Metodologia

A Metodologia deste projeto de pesquisa visa implementar e avaliar sistemas de detecção de intrusão baseados em machine learning para redes de computadores e sistemas computacionais. A metodologia será composta por sete etapas principais, abrangendo desde a revisão sistemática da literatura até a documentação dos resultados para submissão em congressos e revistas de alto impacto científico.

A primeira etapa consistirá na realização de uma revisão sistemática da literatura para identificar o estado da arte em IDS baseados em ML. Os artigos selecionados serão analisados criticamente para extrair as principais informações sobre as técnicas de ML utilizadas, os *datasets* empregados, os resultados obtidos e os desafios enfrentados.

Na segunda etapa, serão extraídas as variáveis principais dos artigos selecionados na revisão sistemática da literatura. As variáveis extraídas incluirão:

- Tipo de ataque: O tipo de ataque que o sistema IDS visa detectar (por exemplo, ataques DDoS, ataques de varredura, ataques de injeção de SQL).
- Técnica de ML: A técnica de ML utilizada para implementar o sistema IDS (por exemplo, redes neurais artificiais, algoritmos de aprendizado de conjunto, florestas aleatórias).
- *Dataset*: O *dataset* utilizado para treinar e avaliar o sistema IDS.
- Métricas de desempenho: As métricas de desempenho utilizadas para avaliar o sistema IDS (por exemplo, precisão, taxa de detecção verdadeira, taxa de falsos positivos).

A partir da análise das variáveis extraídas na etapa 3, serão observadas as tendências de pesquisa em IDS baseados em ML. Essa análise permitirá identificar as técnicas de ML mais utilizadas, os datasets mais populares e as métricas de desempenho mais relevantes para a área.

Na quinta etapa, serão obtidos e pré-processados os *datasets* para treinamento dos modelos IDS. Os *datasets* serão selecionados com base nas tendências de pesquisa observadas na etapa 4 e nas características das redes de computadores que serão protegidas pelos sistemas IDS. O pré-processamento dos *datasets* envolverá a limpeza dos dados, a normalização e a transformação dos dados em um formato adequado para o treinamento dos modelos ML.

A sexta etapa consistirá no treinamento dos modelos IDS utilizando diversos algoritmos de machine learning. As linguagens de programação R e Python serão utilizadas, juntamente com as bibliotecas “mlxtend”, “scikit-learn”, “pandas”, “numpy” e “tensorflow”. Diversos algoritmos de ML serão testados, como redes neurais artificiais, algoritmos de aprendizado de conjunto e florestas aleatórias.

Na sétima etapa, os modelos IDS treinados na etapa 6 serão testados em um *dataset* de teste independente. Os resultados dos testes serão avaliados utilizando métricas de desempenho como precisão, taxa de detecção verdadeira e taxa de falsos positivos. A análise estatística dos resultados será realizada para determinar se os modelos IDS apresentam um bom desempenho na detecção de intrusões.

Na última etapa, os resultados da pesquisa serão documentados em um artigo científico para submissão a congressos e revistas de alto impacto científico. O artigo descreverá a metodologia utilizada, os resultados obtidos e a análise dos resultados. O artigo também discutirá as contribuições da pesquisa para a área de IDS baseados em ML e as implicações práticas dos resultados. Faz-se importante destacar aqui a necessidade do apoio financeiro por parte do empregador com as custas referentes às despesas de publicação ou de participação em congressos.

3. Resultados e Discussão

O projeto de pesquisa visa desenvolver e avaliar sistemas de detecção de intrusão em redes de computadores e sistemas computacionais utilizando machine learning. Os resultados esperados para cada etapa do projeto são descritos a seguir:

3.1. Etapa 1: Desenvolvimento de IDS com Métodos Individuais de ML

- Espera-se desenvolver IDS utilizando técnicas de ML individuais, como redes neurais artificiais (RNA), algoritmos de aprendizado de conjunto (ensemble) e florestas aleatórias.
- Os IDS individuais serão avaliados utilizando métricas de desempenho como precisão, taxa de detecção verdadeira (TDR) e taxa de falsos positivos (FPR).
- Com base nos resultados da avaliação, serão selecionadas as melhores técnicas de ML individuais para serem utilizadas nas etapas subsequentes do projeto.

3.2. Etapa 2: Desenvolvimento de IDS com Métodos Agrupados (Ensemble) de ML
- Espera-se desenvolver IDS utilizando métodos de ensemble de ML, como bagging, boosting e stacking.
- Os métodos de ensemble combinarão as melhores técnicas de ML individuais selecionadas na etapa 1.
- Os IDS com ensemble serão avaliados utilizando as mesmas métricas de desempenho da etapa 1.
- O desempenho dos IDS individuais será comparado com o desempenho dos IDS com ensemble para determinar se os métodos de ensemble oferecem uma melhor performance na detecção de intrusões.

3.3. Etapa 3: Desenvolvimento de IDS após Seleção de Características e Ajustes de Dados
- Espera-se realizar a seleção de características relevantes para os datasets utilizados no treinamento dos IDS.
- As técnicas de ajuste de dados, como oversampling e undersampling, serão aplicadas para lidar com problemas de desequilíbrio de classes nos datasets.
- Espera-se desenvolver IDS utilizando as características selecionadas e os dados ajustados.
- Os IDS com características selecionadas e dados ajustados serão avaliados utilizando as mesmas métricas de desempenho das etapas 1 e 2.
- Análise do desempenho dos IDS com diferentes configurações (sem seleção de características e ajuste de dados, com seleção de características, com ajuste de dados, com seleção de características e ajuste de dados) será comparado para determinar o impacto dessas técnicas no desempenho dos IDS.

Em resumo, os Resultados Gerais Esperados são:
- Espera-se desenvolver IDS eficazes para detecção de intrusões em redes de computadores e sistemas computacionais utilizando ML.
- Espera-se que os IDS desenvolvidos neste projeto apresentem melhor precisão e eficiência na detecção de intrusões em comparação com os métodos tradicionais de IDS.
- Espera-se que os resultados deste projeto contribuam para a área de segurança da informação, fornecendo novas abordagens e metodologias para o desenvolvimento de IDS mais eficazes.

Alguns resultados parciais (já obtidos) podem ser observados na sequência:

1. Atuação como revisor do artigo “Robust Framework for Detecting and Classifying Network-based Attacks using Ensemble Machine Learning Approach” para o *International Journal of Computer Theory and Engineering;*
2. Submissão do artigo “Undersampling aplicado a Detecção de Ransonwares: Uma Análise dos Efeitos do Nearmiss e do Random Undersampling” para o XXV Simpósio Brasileiro de Cibersegurança (SBSeg 2025);
3. Submissão do artigo “Large Language Models for Intrusion Detection: Tokenization Impacts on DDoS Flows” para o XXV Simpósio Brasileiro de Cibersegurança (SBSeg 2025);

4. Submissão do artigo “Avaliação de Técnicas de Ensemble Learning na Detecção de Ataques SQL Injection” para o periódico RETEC (Revista de Tecnologias - ISSN 1806-0323);
5. Submissão do artigo “Análise do Desempenho de Classificadores Supervisionados na Detecção de Fraude em Transações por Cartões de Crédito” para o periódico RETEC (Revista de Tecnologias - ISSN 1806-0323);
6. Submissão do artigo “Detection of Obfuscation Malware: A Federated Transfer Learning-based Approach with Hybrid Neural Networks” para o periódico “IEEE Latin America Transactions”;
7. Submissão do artigo “Evaluation of Machine Learning Algorithms for Attack Detection in SCADA Systems” para o periódico “International Journal of
8. Computer Information Systems and Industrial Management Applications”;
9. Desenvolvimento e publicação do website do laboratório, disponível em https://detectai.fatecourinhos.edu.br/
10. Orientação em andamento de 12 Trabalhos de Conclusão de Curso com temática relacionada aos desafios do laboratório;
11. Co-orientação de Mestrado na Unesp do discente “Carlos de Jesus Reis” com temática relacionada aos desafios do laboratório.

4. Considerações finais

O presente projeto de pesquisa tem como objetivo o desenvolvimento e a avaliação de Sistemas de Detecção de Intrusão (IDS) utilizando técnicas de Machine Learning, visando aprimorar a precisão e a eficiência na identificação de ameaças em redes de computadores e sistemas computacionais.

Estão em desenvolvimento IDSs baseados em diferentes abordagens de aprendizado de máquina, desde métodos individuais até técnicas de *ensemble*, além da aplicação de estratégias de seleção de características e ajustes de dados. As avaliações iniciais indicam que a combinação dessas técnicas pode resultar em melhorias significativas na detecção de ataques e na redução da taxa de falsos positivos, um dos principais desafios em sistemas de segurança cibernética.

Além do avanço técnico, o projeto já resultou em importantes produções acadêmicas, como a submissão de artigos para periódicos especializados, o desenvolvimento de um website para divulgação dos trabalhos do laboratório e a orientação de diversas pesquisas de graduação (Fatec Ourinhos) e pós-graduação (Unesp). Esses resultados reforçam a relevância da investigação na área de segurança da informação e o impacto da pesquisa na formação de novos profissionais.

Cabe ressaltar que o projeto ainda está em andamento e novas etapas serão conduzidas para aprofundar a análise dos modelos desenvolvidos e validar suas aplicações em cenários reais. Espera-se que os próximos estudos permitam consolidar as descobertas iniciais e contribuir ainda mais para o avanço dos sistemas de detecção de intrusão baseados em aprendizado de máquina.

Referências

[1] AL-BATAINEH, M.; KHALIL, I.; AL-FALAKHI, I. A novel hybrid intrusion detection system based on machine learning and fuzzy logic for the internet of things. Sensors, v. 23, n. 3, p. 3072, 2023.

[2] BUCALA, C.; POKORNY, D.; RUŽICKA, M. A survey of intrusion detection systems based on machine learning for the internet of things. Wireless Networks, v. 27, n. 8, p. 2799-2821, 2021.

[3] CISCO. Cybersecurity Reports. Disponível em: https://tinyurl.com/8sdcnfxz. Acesso em: 25 mar. 2025.

[4] GARTNER. Security & Risk Management Summit. Disponível em: https://tinyurl.com/bh6wcfka. Acesso em: 25 mar. 2025.

[5] LASHKARI, A.; MAHMOUDI, M.; DEHGHANI, S. A survey on machine learning-based anomaly detection for network intrusion detection systems. Journal of Network and Computer Applications, v. 178, p. 103063, 2023.

[6] LUCAS, Thiago José et al. A comprehensive survey on ensemble learning-based intrusion detection approaches in computer networks. IEEE Access, v. 11, p. 122638-122676, 2023.

[7] LUCAS, Thiago José et al. An ensemble pruning approach to optimize intrusion detection systems performance. In: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2022. p. 1173-1179.

[8] LUCAS, Thiago José et al. Ensemble Diversity Pruning on Cybersecurity: Optimizing Intrusion Detection Systems. In: 2024 31st International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, 2024. p. 1-6.

[9] LUCAS, Thiago José. et al. Stacking-based committees para detecçao de ataques em redes de computadores-uma abordagem por exaustao. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC). SBC, 2021. p. 644-657.

[10] MOUSTAFA, N.; AL-SARHAN, A.; AL-ALI, A.; TARI, Z. A comprehensive survey on machine learning-based intrusion detection systems for the internet of things. Wireless Networks, v. 28, n. 8, p. 3625-3661, 2022.

[11] WANG, Y.; LIN, Z.; LI, X.; REN, Y. A survey of machine learning for network intrusion detection. Neurocomputing, v. 521, p. 1-30, 2022.