UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

BRUNO FERREIRA LEAL

AVALIAÇÃO DE SISTEMAS DE DETECÇÃO DE INTRUSÃO BASEADOS EM FLUXO UTILIZANDO SEGMENTAÇÃO DE REDE

São José do Rio Preto
2021

BRUNO FERREIRA LEAL

AVALIAÇÃO DE SISTEMAS DE DETECÇÃO DE INTRUSÃO BASEADOS EM FLUXO UTILIZANDO SEGMENTAÇÃO DE REDE

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Orientador:
Prof. Dr. Adriano Mauro Cansian

São José do Rio Preto
2021

L435a

Leal, Bruno Ferreira

Avaliação de Sistemas de Detecção de Intrusão baseados em Fluxo utilizando Segmentação de Rede / Bruno Ferreira Leal. -- São José do Rio Preto, 2021

78 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Adriano Mauro Cansian

1. Sistemas de Detecção de Intrusão. 2. Segmentação de Rede. 3. MITRE ATT&CK. 4. Fluxo de Rede. 5. Aprendizado de Máquina. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

BRUNO FERREIRA LEAL

AVALIAÇÃO DE SISTEMAS DE DETECÇÃO DE INTRUSÃO BASEADOS EM FLUXO UTILIZANDO SEGMENTAÇÃO DE REDE

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. Adriano Mauro Cansian
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Leandro Alves Neves
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Robson de Oliveira Albuquerque
UnB – Universidade de Brasília

São José do Rio Preto
09 de setembro de 2021

Àqueles que me auxiliaram e deram apoio e acreditaram que eu seria capaz de atingir meus objetivos.

AGRADECIMENTOS

Primeiramente agradeço a minha família por todo suporte e confiança ao longo destes anos. Agradeço ao meu pai, minha mãe, e aos meus irmãos, pelo incentivo aos estudos, e acreditarem em mim desde muito tempo atrás. Agradeço minha esposa, pela paciência, apoio e incentivo para eu pudesse chegar até aqui. Agradeço também aos meus tios e tias, primos e primas, amigos, e minha avô, que jamais esquecerei.

Agradeço ao meu orientador Prof. Dr. Adriano Mauro Cansian, pelo apoio, pelos ensinamentos fornecidos e pela contribuição para minha formação acadêmica e pessoal. Também agradeço aos atuais membros do laboratório ACME!, com destaque para Amanda e o João, pelos conhecimentos compartilhados e auxílio no desenvolvimento deste projeto, assim como agradeço aos outros membros, atuais e que já se foram, do laboratório, mas contribuíram igualmente.

Agradeço, também, ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, campus de São José do Rio Preto, pelo curso de pós-graduação ali ministrado por professores de excelência. E, por fim, também agradeço a Prof. Dr. Adriana Barbosa que por meio de seu auxílio prestado aos alunos da permanência estudantil durante a graduação, permitiu que eu e outros amigos e colegas tivessem condições um pouco melhores para concluir seus estudos.

“Nunca olhe alguém de cima para baixo, a menos que esteja ajudando-o a se levantar.”

Jesse Jackson (2011)

RESUMO

Os Sistemas de Detecção de Intrusão (IDS) são um dos mecanismos primários e principais de segurança adotados para identificar e monitorar ataques à rede. Embora trabalhos tenham sido desenvolvidos com o objetivo de melhorar a capacidade de detecção destes sistemas, para IDSs baseados em fluxo, desafios em torno da obtenção de melhores resultados ainda persistem. Além disso, como parte importante do monitoramento de segurança, o nível de observabilidade empregado pelo sistema é uma das características que mais agregam confiança aos resultados por ele obtido. Com o objetivo de proporcionar uma abordagem de avaliação de IDSs baseados em fluxo que considere tal aspecto, este trabalho aplica algoritmos de aprendizagem de máquina não supervisionados, DBSCAN e *K-Means*, para automatizar a segmentação de rede e demonstrar como esta estratégia proporciona ganhos na taxa de acurácia de IDSs, quando aplicada sobre dados de fluxo presentes nos segmentos formados. Para os diferentes modelos de IDS avaliados, representados neste trabalho por meio dos algoritmos de aprendizagem de máquina KNN, *Naive Bayes*, XGBoost e TPOT, as análises foram realizadas observando técnicas de ataque mapeadas a partir do *framework* MITRE ATT&CK. Os resultados obtidos por meio da abordagem proposta chegaram a 97,67% na taxa de acurácia de detecção dos eventos de interesse.

Palavras-Chave: Sistemas de Detecção de Intrusão. Segmentação de Rede. MITRE ATT&CK. Fluxo de Rede. Aprendizado de Máquina.

ABSTRACT

Intrusion Detection Systems (IDS) are one of the primary and main security tools adopted to identify and monitor network attacks. Although works have been developed with the aim to improve the detection capability of these systems, for flow-based IDSs, challenges around obtaining better results still persist. Furthermore, as an important part of security monitoring, the level of observability used by the system is one of the characteristics that most add confidence to the results obtained by it. In order to provide a flow-based IDS evaluation approach that considers this aspect, this work applies unsupervised machine learning algorithms, DBSCAN and K-Means, to automate network segmentation and demonstrate how this strategy provides gains in accuracy rate of IDSs, when applied on flow data present in the formed segments. For the different IDS models evaluated, represented in this work through the machine learning algorithms KNN, Naive Bayes, XGBoost and TPOT, the analyzes were carried out observing attack techniques mapped from the MITRE ATT&CK framework. The results obtained through the proposed approach reached 97.67% in the detection accuracy rate of the events of interest.

Keywords: Intrusion Detection Systems. Network Segmentation. MITRE ATT&CK. Network Flow. Machine Learning.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

A: Acurácia

AGC: Acurácia Genérica do Conjunto

AMS: Acurácia Média dos Segmentos

APT: *Advanced Persistent Threat*

ATT&CK: *Adversarial Tactics, Techniques and Common Knowledge*

BCDR: *Business Continuity and Disaster Recovery*

CERT.br: Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil

CNN: *Convolutional Neural Network*

CPS: *Cyber-Physical Systems*

CPU: *Central Process Unit*

CTI: *Cyber Threat Intelligence*

DeTT&CT: Detect Tactics, Techniques & Combat Threats

E: Especificidade

ELK: *Elasticsearch, Logstash and Kibana*

ETL: *Extract, Transform, Load*

FN: Falso Negativo

FP: Falso Positivo

F: *F1-Score*

FTP: *File Transfer Protocol*

HIDS: *Host-Based Intrusion Detection System*

HTTP: *Hypertext Transfer Protocol*

HTTPS: *Hypertext Transfer Protocol Secure*

ICMP: *Internet Control Message Protocol*

ID: Identificador

IDS: *Intrusion Detection System*

IETF: *Internet Engineering Task Force*

IoT: *Internet of Things*

IPS: *Intrusion Prevention System*

IPv4: *Internet Protocol version* 4

IPv6: *Internet Protocol version* 6

NIDS: *Network-Based Intrusion Detection System*

P: Precisão

PCA: *Principal Component Analysis*

RFC: *Request for Comments*

SC: *Silhouette Score*

SDN: *Software Defined Networking*

SMB: *Server Message Block*

S: Sensibilidade

SSH: *Secure Shell*

STIX: *Structured Threat Information Expression*

TAXII: *Trusted Automated Exchange of Intelligence Information*

TCP: *Transmission Control Protocol*

TFP: Taxa de Falso Positivo

TOMATO: *Threat Observability and Monitoring Assessment Tool*

UDP: *User Datagram Protocol*

UNESP: Universidade Estadual Paulista “Júlio de Mesquita Filho”

VN: Verdadeiro Negativo

VP: Verdadeiro Positivo

SUMÁRIO

CAPÍTULO 1 - Introdução

1.1 Considerações Iniciais

A adoção e implantação de Sistemas de Detecção de Intrusão (IDS, sigla em inglês) para identificação de ataques em redes de computadores é uma das principais estratégias adotadas para realizar o supervisionamento e monitoramento de ameaças nestas redes (HINDY et al., 2018). A utilização adequada deste tipo de sistema, como quando implementado em conjunto com uma política de segurança consistente, e que englobe aspectos como Continuidade do Negócio e Resposta a Incidentes (BCDR, sigla em inglês), possibilita que equipes de segurança sejam capazes de, tão logo quanto possível, identificar ataques ocorrendo na rede, e adotar contramedidas que mitiguem ou interrompam o ataque, bem como os impactos decorrentes do mesmo. Métodos que exploram a eficiência e otimização de IDSs têm sido propostos em trabalhos como os de Aljawarneh, Aldwairi e Yassein (2018), Ali et al. (2018), e Vinayakumar et al. (2019).

A natureza das análises realizadas por IDSs está diretamente relacionada ao processamento e busca de padrões, ou comportamentos anômalos na rede, os quais, por sua vez, caracterizam os eventos de interesse, podendo ser maliciosos ou não. Ao considerar o volume crescente de tráfego de dados nas redes de computadores, e o recurso computacional necessário para realizar a detecção de eventos sobre o tráfego gerado, a não utilização de todo o conteúdo trafegado, também conhecido como *payload*, para identificar e avaliar tais padrões e comportamentos, ocasionou, ao

longo do tempo, o estudo e desenvolvimento de IDSs que utilizam apenas informações presentes no fluxo da rede para realizar a detecção destes eventos. No Brasil, o volume de dados trafegados entre os Sistemas Autônomos (ASs) que compõem a Internet Brasileira, registrados pelo IX.br, projeto do Comitê Gestor da Internet no Brasil (CGIbr), chegou em 2021 à marca de 18,83 Tbps (Tera *bits* por segundo), um aumento de cerca de 3.138,33% no volume de dados trafegados entre estes ASs quando comparado aos cerca de 0,6 Tbps de dados trafegados no início do ano de 2015 (IX.br, 2021). Os metadados que representam o fluxo da rede e que são utilizados por IDSs baseados em fluxo para realizar a detecção dos eventos de interesse são descritos por meio de padrões bem conhecidos, como *Netflow* (CLAISE, 2004), *Openflow* (MCKEOWN et al., 2008) e *Sflow* (PANCHEN; PHAAL; MCKEE, 2001).

Além disso, devido ao aumento do número de ameaças presentes na rede e da quantidade de dados a serem analisados, surgiu a necessidade de se automatizar o processo de detecção de IDSs baseados em fluxo, e, posteriormente, melhorar a taxa de detecção desses sistemas considerando tais aspectos. Com isso, trabalhos que aplicam técnicas de aprendizado de máquina para identificar comportamentos e atividades maliciosas na rede por meio da análise do fluxo foram propostos por (TERZI; TERZI; SAGIROGLU, 2017), (KAKIHATA et al., 2017) e (GONÇALVES, 2019).

Também, dentre os diferentes modelos de IDSs baseados em fluxo propostos na literatura, nota-se a utilização de diferentes métricas de avaliação para os mesmos, dentre as quais se destacam a Taxa de Falso Positivo (TFP) e a Acurácia (A), sendo essas as métricas observadas com maior frequência nos trabalhos relacionados, ora descritos no CAPÍTULO 3 - Trabalhos Relacionados. Observando-se as métricas citadas, de modo geral, a avaliação da qualidade de um IDS é baseada na eficácia com que o mesmo identifica eventos na rede, sendo os eventos maliciosos, ou ataques, as ocorrências de interesse a serem identificadas corretamente por IDSs. No entanto, em diversos cenários, não se considera para tal análise a qualidade e eficácia da detecção desses sistemas ao considerar o nível de observabilidade dos IDSs ao identificar eventos na rede, ou seja, não são observados quão capazes tais sistemas são de identificar ameaças na rede como um todo (KHRAISAT et. al., 2019).

Ademais, ao considerar a abrangência do sistema com relação a sua capacidade de observação do ambiente, as métricas utilizadas para validar sua eficácia demonstram maior completeza (JOY; MHAMDI; MITSOS, 2020). Para ambientes de rede que geram dados heterogêneos e classificados como *Big Data*, como é o caso de redes de médio e grande porte, ou mesmo redes de pequeno porte com características de uso intensas e também heterogêneas, a obtenção de altas taxas de acurácia para essas redes acabam por não representarem efetivamente o estado e nível de segurança do ambiente sendo monitorado, visto que, em diversos cenários, tais análises não são capazes de considerar as especificidades de todos os dispositivos envolvidos no tráfego observado (BAO, H.; HE, H.; LIU, Z; LIU, Z, 2019).

Por fim, a observação e avaliação da capacidade de detecção de ameaças por IDSs exige que os mesmos sejam capazes de identificar ameaças em ambientes reais (HALVORSEN; WAITE; HAHN, 2019), tais como as ameaças definidas por meio das técnicas de ataques descritas no *framework* de detecção e mapeamento de ameaças MITRE ATT&CK (MITRE, 2021).

1.2 Justificativa

O desenvolvimento e aperfeiçoamento de IDSs baseados em fluxo tem se mostrado um campo de estudo promissor, haja vista que, diferentemente de IDSs que realizam análises considerando o *payload*, as análises realizadas sobre o fluxo não estão sujeitas à obscuridade de informações, como é o caso quando o tráfego a ser analisado está criptografado - característica presente em redes e sistemas que prezam cada vez mais por segurança.

Outro fator que favorece a utilização de IDSs baseados em fluxo é a não violação da privacidade de acesso. Diferentemente de IDSs que realizam a análise do conteúdo sendo trafegado e, consequentemente, acarretam na violação da privacidade do usuário, IDSs baseados em fluxo analisam apenas metadados que descrevem o tráfego entre cliente e servidor, assim, evadindo a necessidade de se conhecer o que está sendo consumido pelo usuário, e indo ao encontro de legislações que prezam pela privacidade no uso da rede, como a Lei do Marco Civil da Internet no Brasil

(LEI Nº 12.965, 2014) e a Lei Geral de Proteção de Dados (LEI Nº 13.709, 2018) brasileira.

Durante o levantamento de estudos relacionados a este trabalho, ora direcionados ao desenvolvimento de IDSs baseados em fluxo, notou-se certa escassez na exploração de métodos que sejam capazes de avaliar a eficiência desses sistemas no que tange a completude das análises realizadas em redes heterogêneas. Abordagens para este tipo de análise são apresentadas em (SCHEERES; ALFRIEND; FRUEH, 2018), (MILLER; ALIASGARI, 2018) e (HALVORSEN; WAITE; HAHN, 2019). Contudo, os estudos citados fazem uso de dados diferentes daqueles intrínsecos ao fluxo da rede e, por sua vez, ampliam o espectro de informações analisadas ao incluir registros de atividades provenientes de cada dispositivo na rede, informações de sessão, e outros dados extraídos do *payload*, ou seja, dados que não compõem estritamente o fluxo da rede. Deste modo, embora tenham apresentados resultados considerados bons, com acurácia acima de noventa por cento, os trabalhos mencionados realizam um processo de treinamento e validação de IDSs baseados em fluxo, embora não utilizem informações única e estritamente presentes no mesmo.

Portanto, o presente trabalho se justifica diante da abertura para o estudo de uma abordagem de avaliação de IDSs que, além de considerar apenas dados presentes no fluxo de rede para avaliar sua eficiência, propõe a utilização da estratégia de segmentação de rede como meio para obter uma melhor observabilidade dos elementos que a compõem e, deste modo, obter melhorias na detecção de ataques conhecidos.

1.3 Objetivo

O objetivo principal deste trabalho é desenvolver de uma abordagem de avaliação de IDSs baseados em fluxo que faz uso do processo de segmentação de rede para identificar de maneira mais assertiva eventos maliciosos na rede, com foco em eventos presentes em cenários reais e mapeados a partir do *framework* MITRE ATT&CK. A abordagem proposta tem como principal característica a capacidade de, utilizando dados exclusivamente provenientes do fluxo, realizar a detecção de

eventos maliciosos em toda a rede observada, por meio da análise de seus segmentos, com foco direcionado a eventos oriundos dos MITRE ATT&CK. Para cumprir este objetivo, os seguintes objetivos específicos foram atingidos:

- Por meio da análise das técnicas e subtécnicas de ataque descritas no *framework* MITRE ATT&CK, foram enumeradas aquelas cujo processo de detecção é realizado utilizando dados oriundos do fluxo da rede;
- Foi realizado o mapeamento dos principais atributos mencionados na literatura para a identificação dos ataques mapeados a partir do *framework*;
- Foi desenvolvida uma base de fluxos, ACME'21, a partir dos dados exportados da rede da UNESP por meio do protocolo *Netflow* V9, que contém a classificação dos ataques mapeados a partir do MITRE ATT&CK;
- Foi realizada a análise da eficiência da abordagem proposta para a avaliação de IDSs baseados em fluxo ao compará-la a estratégias de avaliação consideradas usuais, ou seja, que não aplicam a técnica de segmentação de rede em sua abordagem de detecção. A análise comparativa foi realizada sobre as bases de fluxos UNSW-NB15 e ACME'21.

1.4 Organização da Dissertação

Este trabalho está dividido em seis capítulos principais, incluindo o atual. No Capítulo 2, a fim de contextualizar o leitor sobre os conceitos e tecnologias utilizadas para o desenvolvimento deste trabalho, é realizada uma revisão bibliográfica sobre as principais ferramentas e tecnologias utilizadas. Já no Capítulo 3 são apresentados os trabalhos encontrados na literatura que são correlatos e contribuíram para o desenvolvimento da abordagem proposta neste trabalho. No Capítulo 4 é apresentada a metodologia utilizada para o cumprimento dos objetivos propostos, a qual contém a descrição da arquitetura proposta e o detalhamento de seus componentes. No Capítulo 5 são apresentados os resultados obtidos a partir da aplicação do método

proposto, assim como a discussão dos mesmos. Por fim, no Capítulo 6 são apresentadas as conclusões obtidas a partir dos resultados, as dificuldades encontradas no desenvolvimento do trabalho, e propostas de atividades futuras que possam estender a aplicação do mesmo.

CAPÍTULO 2 - Fundamentação Teórica

Este capítulo tem por objetivo apresentar a fundamentação teórica necessária para o entendimento do trabalho, introduzindo os conceitos e tecnologias utilizadas para o desenvolvimento do mesmo.

2.1 Fluxo de Rede

A extração de informações sintetizadas a partir de um conjunto de dados, também conhecidas como metadados, reduz consideravelmente a quantidade de informação a ser analisada, independentemente do objetivo traçado para tal análise (GARTNER, 2016). Em redes de computadores, toda a troca de dados entre dispositivos interconectados é conhecida como tráfego (KUROSE; ROSS, 2016), e a exportação de metadados que representam este tráfego, também conhecido como fluxo de rede, permite que a sumarização e contabilização de todo tráfego envolvido na comunicação entre os elementos da rede seja realizada a partir de um único ponto de exportação, dependendo da topologia e organização dos elementos que compõem a rede, como *switches* e roteadores.

A fim de padronizar as informações que compõem o fluxo de dados na Internet, o *Internet Engineering Task Force* (IETF, sigla em inglês,), órgão regulador da Internet, em 2013, redefiniu por meio do RFC 7011 o protocolo IPFIX, que especifica de forma detalhada quais campos devem compor um fluxo de dados (CLAISE; TRAMMELL; AITKEN, 2013). A criação e especificação de tal

protocolo uniformizou o processo de exportação e coleta de informações de dispositivos de rede, como os próprios roteadores e *switches*, proporcionando maior eficiência no processo de armazenamento e análise de tais informações.

2.1.1 *Netflow*

O *Netflow*, definido pelo RFC 3954, é um protocolo privado de exportação de fluxo de dados criado pela Cisco para ser incorporado a seus equipamentos (CLAISE, 2004). Em sua versão 5, o *Netflow* passou a ser amplamente difundido e adotado por outras empresas do mesmo segmento que sua criadora, e se tornou base para a definição do IPFIX. Em sua versão 9, o protocolo passou a agregar campos provenientes do protocolo de rede IPv6, e coletar e registrar características oriundas do fluxo de rede não possíveis de serem obtidas em versão anterior, o *Netflow* V5. Além disso, por meio da definição e utilização de *templates*, recurso de gerenciamento de fluxo incluído no *Netflow* V9, foi acrescentada ao protocolo a capacidade de definir quais informações compõem os metadados que representam o fluxo da rede. Enquanto no *Netflow* V5 todos os atributos do fluxo são exportados, os *templates* presentes no *Netflow* V9 permitem que sejam exportados do fluxo de rede apenas os dados relevantes para a análise.

2.1.2 *Openflow*

O *Openflow* é um protocolo de código aberto de exportação e manipulação de dados referentes ao tráfego de rede. O *Openflow* tem sua aplicabilidade voltada principalmente para a gerencia de tráfego em Redes Definidas por Software – SDN, permitindo que, por meio da utilização deste protocolo, os administradores de rede tenham controle amplo sobre a direção, regras de encaminhando e priorização de tráfego na rede (MCKEOWN et al., 2018).

2.1.3 *Sflow*

O *Sflow*, assim como o *Netflow*, é um protocolo de monitoramento, coleta e exportação de dados sumarizados relacionados ao tráfego de rede. A tecnologia que rege o protocolo, desenvolvida pela InMon Corporation's, e especificada no RFC 3176, tem como principais objetivos o monitoramento do tráfego de redes de alta velocidade, como redes *Gigabit* e superiores, além da implementação de agentes de monitoramento (PANCHEN; PHAAL; MCKEE, 2001). Ainda, o protocolo adota uma política de coleta de fluxo por meio da qual são obtidas informações completas de camadas aquém das observadas pelo *Netflow*. Com o *Sflow* apenas alguns exemplares do fluxo são coletados, e dados estatísticos são produzidos pelo protocolo para representar o comportamento global da rede, diferentemente do *Netflow* que exporta todas as informações do fluxo, sejam eles definidos por meio de *templates* ou não, e realiza um rastreio dos mesmos.

2.2 Ataques às Redes de Computadores

A fim de obter informações sigilosas a respeito de dados de usuários, ou até mesmo expor falhas de segurança de ambientes e/ou aplicações, ataques às redes de computadores fazem uso de vulnerabilidades para evadir sistemas de proteção e, por muitas vezes, causar danos alterando o funcionamento normal da rede e dos dispositivos conectados a ela (HOQUE et al., 2014).

Ataques às redes de computadores podem ser classificados em duas categorias: passivos e ativos. Ataques passivos são aqueles em que o invasor captura os dados transmitidos e analisa o tráfego de rede para obter informações e identificar padrões de uso da mesma. Por outro lado, em ataques ativos o invasor executa comandos, utiliza-se de aplicações maliciosas, conhecidas como *malware*, para executar atividades maliciosas na rede, ou induz usuários a realizarem operações que alterem o funcionamento normal da rede (KANDAN; KATHRINE; MELVIN, 2019).

Nas Seções 4.2.1 até 4.2.8 são apresentados os ataques às redes de computadores examinados neste trabalho.

2.3 Sistema de Detecção de Intrusão

Sistema de Detecção de Intrusão (IDSs) são sistemas que, por meio de uma abordagem passiva, têm como objetivo realizar a detecção de eventos, geralmente maliciosos, ocorrendo no ambiente sendo monitorado. De modo geral, IDSs podem variar de acordo com o método de detecção empregado, e de acordo com o ambiente sobre o qual o processo de detecção é aplicado. Ao considerar o ambiente a ser monitorado, os IDSs comumente se diferenciam entre IDS baseado em *host*[1], também conhecidos por HIDS (do inglês, *Host-Based Intrusion Detection System*), cujo processo de detecção é direcionado a elementos da rede, como estações de trabalho, *notebooks*, *smartphones* ou servidores; e IDS baseado em rede, comumente chamado de NIDS (do inglês, *Network Based Intrusion Detection System*), no qual o processo de detecção é direcionado ao ambiente de rede (FERREIRA, 2016). Além disso, ao considerar o método de detecção aplicado, independente do ambiente, os IDSs se diferenciam entre métodos de detecção baseados em abuso, e métodos de detecção baseados em anomalia, os quais são descritos nas seções 2.3.1 Sistema de detecção de intrusão por abuso, e 2.3.2 Sistema de detecção de intrusão por anomalia, respectivamente.

2.3.1 Sistema de detecção de intrusão por abuso

Sistemas de detecção de intrusão por abuso têm como característica principal a identificação de atributos e valores que caracterizam determinado comportamento presente no ambiente monitorado. A partir de tais atributos e valores, são definidas assinaturas que se utilizam de regras e comparações que determinam a ocorrência, ou não, do evento por elas descrito (HINDY et al., 2018). No contexto de IDSs baseados em fluxo, as assinaturas utilizadas no método de detecção por abuso são aplicadas sobre os dados do fluxo trafegado na rede para realizar a identificação de um ataque acorrendo na mesma, por exemplo. Ao serem encontradas similaridades no fluxo que atendam aos critérios definidos pela assinatura, ou conjunto de assinaturas, alguns

[1] *Host*, ou hospedeiro, como também é chamado, remete a qualquer elemento computacional que esteja conectado à rede.

IDSs incluem a função de emissão de alerta uma vez que tal correspondência for encontrada. Isso é feito com o intuito de notificar o analista de segurança a ocorrência do evento de interesse.

2.3.2 Sistema de detecção de intrusão por anomalia

Sistemas de detecção de intrusão por anomalia se baseiam na identificação de eventos de interesse por meio da análise e observação de comportamentos no ambiente, que sejam divergentes do usual, característica importante devido a sua aplicabilidade na detecção de atividades maliciosas nunca antes vistas no ambiente (HINDY et al., 2018). No entanto, a alta taxa de ocorrências e, consequentemente, alertas, gerados para eventos que são considerados de interesse, mas cuja identificação se deu de forma incorreta, ou seja, falso positivos, reduz a eficiência do uso de IDSs baseados em anomalia no que tange a perspectiva de aplicação em cenários reais, em que analistas de segurança alocam tempo e recursos para lidar com tais eventos.

2.3.3 Métricas de avaliação da qualidade de detecção

A avaliação do modelo proposto consiste essencialmente na análise binária da identificação, ou não, de um evento, ou conjunto de eventos, que sejam de interesse, neste caso, eventos maliciosos, por meio do conjunto de assinaturas geradas para os mesmos. Neste contexto, para a avaliação do método proposto, foi utilizada a métrica acurácia, descrita na Equação 3, e que compõe um conjunto de métricas amplamente empregadas para análise binária de sistema de detecção de intrusão descritas nas Equações 1, 2, 4 e 5. Quando um caso positivo é detectado corretamente a partir de uma assinatura pré-definida para o mesmo, tem-se um verdadeiro positivo (VP), caso contrário, tem-se um falso negativo (FN). No entanto, da perspectiva dos casos negativos, quando os eventos são detectados corretamente, têm-se um verdadeiro negativo (VN), caso contrário, tem-se um falso positivo (FP). A matriz de confusão apresentada na Tabela 1 descreve a correlação dos casos descritos.

Tabela 1: Matriz de confusão.

		Valor Predito	
		Positivo (Anomalia)	Negativo (Normal)
Valor Real	Positivo (Anomalia)	VP	FN
	Negativo (Normal)	FP	VN

Fonte: produzida pelo próprio autor.

A partir dos valores oriundos da matriz de confusão apresentada na Tabela 1, métricas expressivas a respeito do modelo de avaliação são derivadas. Algumas delas são:

a) Sensibilidade, ou *recall* (S): representada na Equação 1, descreve a relação entre todos os casos positivos corretamente classificados sobre o total de casos corretamente classificados.

$$S = \frac{VP}{VP + FN} \qquad \text{(Equação 1.)}$$

b) Especificidade, ou taxa de falsos positivos (E): representada na Equação 2, descreve a relação entre todos os casos negativos corretamente classificados sobre o total de casos erroneamente classificados.

$$E = \frac{VN}{VN + FP} \qquad \text{(Equação 2.)}$$

c) Acurácia (A): representada na Equação 3, descreve a relação entre todos os casos corretamente classificados sobre o total de casos analisados. Esta métrica é diretamente relacionada à performance do modelo.

$$A = \frac{VP + VN}{VP + VN + FP + FN} \qquad \text{(Equação 3.)}$$

d) Precisão (P): representada na Equação 4, descreve a relação entre todos os casos positivos corretamente classificados sobre o total de casos classificados como positivos.

$$P = \frac{VP}{VP + FP} \qquad \text{(Equação 4.)}$$

e) Pontuação F1, ou F1-*Score* (F): representada na Equação 5, descreve a relação entre a precisão (P) e a sensibilidade (S) das classificações realizadas.

$$F = \frac{2 * P * S}{P + S} \qquad \text{(Equação 5.)}$$

f) Coeficiente de Silhueta, ou *Silhouette-Score* (SC): descreve o quão bem o modelo que define o agrupamento condiz com os dados que estão sendo analisados. Para tal, o cálculo da métrica considera a distância Euclidiana entre os elementos pertencentes ao mesmo agrupamento, e a distância entre os elementos pertencentes aos agrupamentos vizinhos. O cálculo do Coeficiente de Silhueta, SC, para cada elemento, i, do conjunto analisado, é representado na Equação 6. O valor do SC varia de -1 (um negativo) à +1 (um positivo), inclusive. Valores negativos indicam que a atribuição do elemento ao agrupamento, neste caso, a atribuição do fluxo a um determinado segmento de rede, aconteceu de forma errônea. O valor um negativo indica que todas as atribuições foram incorretas. Valores próximos a zero, negativos ou positivos, indicam que há sobreposição nos segmentos formados. O valor zero indica sobreposição total, ou seja, todos os elementos pertencem a todos os segmentos. Valores próximos a um positivo indicam que os agrupamentos são compostos por elementos que realmente os representam bem. O valor um positivo representa o agrupamento totalmente correto e disjunto dos elementos.

$$SC(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \qquad \text{(Equação 6.)}$$

Onde $a(i)$ é a distância média do elemento i em relação a todos os demais elementos do mesmo grupo; e $b(i)$ é a distância média entre o elemento i em relação a todos os objetos do grupo vizinho mais próximo a ele.

2.4 MITRE ATT&CK

O MITRE ATT&CK é um *framework* que concentra uma extensa base de conhecimento a respeito de Ameaças Persistentes Avançadas (APT, sigla em inglês). Por meio da enumeração e categorização das principais táticas, técnicas e procedimentos utilizados por estes tipos de ameaças, o *framework* garante uma visão geral das diferentes fases do ciclo de vida dos mesmos, seus principais alvos e, principalmente, comportamento (STROM et al., 2018). Além disso, tendo como principal motivação tornar o mundo um lugar mais seguro, o *framework* MITRE ATT&CK foi construído de modo que qualquer pessoa ou organização, pública ou privada, possa utilizá-lo como base para o desenvolvimento de outros modelos de identificação de ameaças e, também, novos produtos de cibersegurança (MITRE, 2019).

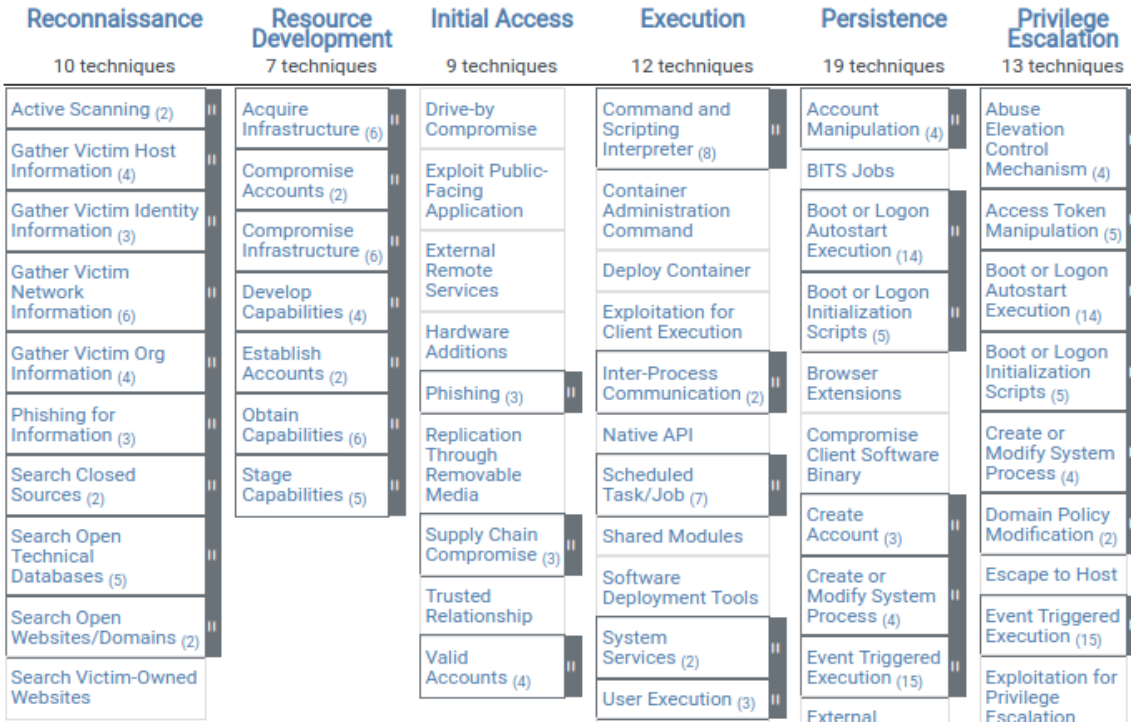
De modo geral, a aplicabilidade do MITRE ATT&CK é dada em função de sua completude e nível de detalhamento, além da forma estruturada como o mesmo organiza as informações relacionadas aos diferentes ataques utilizados em APTs, o que garante a fundamentação necessária não apenas para identificar comportamentos maliciosos em ambientes comprometidos, mas também a possibilidade de utilização do *framework* por equipes de *red team*[2], processos de *Cyber Threat Intelligence* (CTI, sigla em inglês), e gerenciamento e engenharia de ativos (MITRE, 2020). A título de exemplo, uma das formas de aplicação e utilização das informações disponibilizadas pelo *framework* consiste em utilizá-las para identificar e analisar eventos por meio de técnicas de CTI, as quais são técnicas que, por meio do correlacionamento de diversas fontes de informações sobre eventos e ameaças, permitem a identificação de ataques que possam estar acontecendo, ou indícios de que já tenham ocorrido, no ambiente observado. Inclusive, o MITRE, corporação

[2] Equipes de *red team* são equipes cuja função é replicar de modo mais fidedigno possível ataques reais executados por indivíduos mal-intencionados.

mantenedora do *framework*, fornece interfaces de comunicação via servidor TAXII[3] e padrão STIX[4] (BURNS, 2019), por meio das quais o MITRE ATT&CK pode ser consumido, e integrado a outras fontes de CTI.

A relação entre algumas das táticas e técnicas exploradas e descritas pelo MITRE ATT&CK é apresentada por meio da ATT&CK Matrix, ilustrada na Figura 1.

Figura 1: MITRE ATT&CK Matrix.



Fonte: Adaptado de (MITRE, 2021).

2.4.1 Táticas

O MITRE ATT&CK aborda um conjunto de 14 táticas que cobrem os principais objetivos de indivíduos mal-intencionados durante uma operação de ataque. Considerando o fato de que os objetivos dos atacantes não mudam, ou mudam muito pouco ao longo do tempo, a inclusão de novas táticas no *framework* passa por um processo extenso de levantamento e entendimento da mesma, antes que ela possa ser inserida no *framework*. Tal processo garante as táticas listadas no

[3] O TAXII™ (*Trusted Automated Exchange of Intelligence Information*) é um protocolo da camada de aplicação para a troca de informações sobre ameaças cibernéticas de maneira simples e escalável.
[4] O STIX™ (*Structured Threat Information Expression*) é um formato de idioma e serialização usado para trocar informações sobre ameaças cibernéticas (CTI).

framework um aspecto estático, mas também consistente e de profundo entendimento (STROM et al., 2017).

As táticas abordadas pelo MITRE ATT&CK são listadas na primeira linha da ATT&CK Matrix, ilustrada na Figura 1, sendo elas: Reconhecimento, Desenvolvimento de Recursos, Acesso inicial, Execução, Persistência, Escalonamento de Privilégios, Evasão de Defesa, Credencial de Acesso, Varredura, Movimento Lateral, Coleta, Comando e Controle, Exfiltração e Impacto. Cada tática apresentada relaciona o objetivo do atacante com a plataforma alvo e a superfície de ataque, estendendo-se aos diferentes sistemas operacionais, como *macOS*, *Linux* e *Windows*, e arquiteturas, como computadores pessoais e dispositivos móveis.

2.4.2 Técnicas

No MITRE ATT&CK as técnicas correspondem ao “como” um atacante pode alcançar um objetivo dentre as táticas cobertas pelo *framework*. Entretanto, diferentemente das táticas, que refletem os objetivos dos atacantes em um ataque e possuem pouca variabilidade, os métodos para a aplicação de determinada técnica são diversos e evoluem constantemente ao longo do tempo (STROM et al., 2017).

Quando vinculadas a diferentes táticas, equipes *red team* podem utilizar as técnicas enumeradas no MITRE ATT&CK para simular de modo mais fidedigno o comportamento de atividades maliciosas presente em cenários reais de ataque, algo que dificilmente pode ser feito ao abordar cada técnica de modo independente, ou mesmo a partir de outros *frameworks*.

2.5 Pilha Elastic

A Pilha Elastic consiste em um conjunto de ferramentas de código aberto, dentre as quais, aquelas utilizadas no desenvolvimento deste trabalho são o *Elasticsearch*, o *Filebeat* e o *Kibana* (ELASTIC, 2021). A estrutura formada pela Pilha Elastic compõe um poderoso motor de indexação e busca que se destaca pela capacidade de armazenar grandes volumes de dados e realizar consultas em tempo

real, características importantes em se tratando de análises de ameaças em redes de grande porte (LV et al., 2018).

2.5.1 *Elasticsearch*

O *Elasticsearch* é o componente da Pilha Elastic responsável pelo armazenamento e recuperação de dados. Considerado o coração da Pilha Elastic, o *Elasticsearch* desempenha o papel de indexar diferentes tipos de dados e realizar buscas sobre eles. Além disso, este componente agrega a possibilidade de utilizar uma estrutura distribuída para lidar com *Big Data*, na qual, tanto armazenamento, quanto capacidade de processamento são dispersos entre diversas instâncias da ferramenta. Ainda, explorando sua capacidade de busca, é possível criar consultas e filtros que identifiquem e extraiam padrões dos dados armazenados (ELASTICSEARCH, 2021).

2.5.2 *Filebeat*

Dentro da Pilha Elastic, o *Filebeat* desempenha o papel de coleta de dados, de uma ou mais fontes, e inserção destes no *Elasticsearch* por meio de *pipelines* que realizam o processamento e transformação dos dados. Também, por meio do *Filebeat*, dados estruturados e não estruturados podem ser transformados de modo que sejam obtidas informações consistentes e convergentes para um mesmo formato, o que corresponde à atividade de Extração, Transformação e Carregamento de Dados (ETL, sigla em inglês) presente em sistemas voltados para descoberta de conhecimento (FILEBEAT, 2021).

2.5.3 *Kibana*

A visualização clara e facilidade de interpretação de determinados indicadores de rede e ameaças têm se tornado um grande aliado no monitoramento de atividades

maliciosas e de comportamentos anômalos na rede (ZAO et al., 2019). Dentro da Pilha Elastic, o *Kibana* é a ferramenta por meio da qual são criadas visualizações sobre os dados coletados pelo *Filebeat* e armazenados no *Elasticsearch*.

Dentre as diferentes formas de visualização que podem ser utilizados no *Kibana*, se destacam aquelas voltadas para: a criação de histogramas, séries temporais e mapas de geolocalização. Ainda, métodos de visualização e correlação de eventos podem ser agrupados em painéis direcionados à representação integrada e facilitada de informações de determinado objeto de interesse, como é o caso, neste trabalho, dos eventos de rede maliciosos (KIBANA, 2021).

2.6 Aprendizado de Máquina

Permitir que máquinas aprendam mediante novas informações dadas como entrada, e as interpretem de maneira automatizada seguindo um modelo predeterminado, tem proporcionado estudos e avanços em diversos âmbitos, sejam eles públicos ou privados, na área financeira, hospitalar ou de estudo de consumo. O aprendizado de máquina consiste em um conjunto de técnicas matemáticas e estatísticas aplicadas em sistemas computacionais que realizam o processamento de dados a fim de extrair informações, detectar padrões e realizar inferências a partir dos dados analisados (CHIO; FREEMAN, 2018).

Algoritmos de aprendizado de máquina são classificados em dois grandes grupos, sendo eles: algoritmos de aprendizado supervisionado e algoritmos de aprendizado não supervisionado. Algoritmos de aprendizado supervisionado utilizam conjuntos de dados previamente classificados, pertencentes ao domínio do problema a ser tratado, para treinar o modelo, que posteriormente é avaliado ao se submeter um novo conjunto de testes, não conhecido pelo modelo, e pertencente ao mesmo domínio do problema, para averiguar a eficiência do mesmo a partir de suas predições (REESE et al., 2017). Neste trabalho, foram utilizados, por meio da biblioteca de aprendizado de máquina *Scikit-learn* (PEDREGOSA et al., 2011), os algoritmos de aprendizado de máquina supervisionado KNN (do inglês, *K-Nearest Neighbors*) e *Naive Bayes*; por meio do pacote *Python XGBoost* (XGBoost, 2020), o algoritmo *XGBoost*; e, por fim, a ferramenta de aprendizado de máquina

automatizado TPOT (OLSON; MOORE, 2016). Os algoritmos de aprendizado de máquina supervisionados foram utilizados para definir os modelos que representam IDSs.

Por sua vez, algoritmos de aprendizado não supervisionado se utilizam de modelos que têm como objetivo encontrar relações entre os dados fornecidos. No entanto, diferentemente do modelo supervisionado, tais dados não possuem qualquer tipo de classificação prévia, deixando a cargo do próprio algoritmo o agrupamento e classificação dos dados de acordo com parâmetros pré-estabelecidos (REESE et al., 2017). Neste trabalho, foram utilizados, por meio da biblioteca de aprendizado de máquina *Scikit-learn* (PEDREGOSA et al., 2011), os algoritmos de aprendizado de máquina não supervisionado *K-Means* e DBSCAN. Os algoritmos de aprendizado de máquina não supervisionados foram utilizados para realizar, de forma automatizada, a segmentação das redes analisadas, estas, por sua vez, representadas por meio dos conjuntos de dados avaliados.

CAPÍTULO 3 - Trabalhos Relacionados

Neste capitulo é apresentada a revisão bibliográfica de estudos relacionados ao trabalho desenvolvido.

3.1 Abordagens de Avaliação de Sistemas de Segurança

Estudos em torno da criação de IDSs baseados em fluxo por meio da aplicação de algoritmos de aprendizado de máquina e redes neurais foram explorados em trabalhos como os de (ALJAWARNEH; ALDWAIRI; YASSEIN, 2018), (SULTANA et al., 2019) e (GURUNG; GHOSE; SUBEDI, 2019). Em função do desenvolvimento desses sistemas, a análise e criação de métricas por meio das quais eles são avaliados se tornou um fator determinante para o aperfeiçoamento e desenvolvimento de novos métodos de análise.

Ao abordarem um novo método para analisar sistemas, MATOUŠEK, RYŠAVÝ e GRÉGR, 2018 destacaram em seu trabalho a importância de estratégias de monitoramento de segurança como parte relevante do processo de gerenciamento de rede. No modelo proposto pelos autores, direcionado ao monitoramento de dispositivos IoT (do inglês, *Internet of Things*), foram utilizados dois métodos estatísticos. Com o primeiro método, voltado para a análise estatística dos fluxos trocados entre os dispositivos, os autores avaliaram a visibilidade dos dispositivos na rede. Posteriormente, por meio do segundo método, incidentes de segurança como

varredura de recursos e ataques distribuídos de negação de serviço puderam ser identificados.

Já no trabalho de DONG e CHOPRA, 2018, os autores abordaram a segurança de Sistemas Ciberfísicos (CPS, sigla em inglês), como dispositivos médicos, veículos conectados e edifícios inteligentes, por meio da observação de seus estados. A fim de monitorar o comportamento desses sistemas e o estado de segurança dos mesmos, os autores propuseram uma medida quantitativa de observabilidade. A medida é desenvolvida por meio da análise comparativa de dois momentos diferentes do sistema: o primeiro, chamado estado original, e o segundo, chamado estado de saída.

Ao proporem o aprimoramento na confiança de sistemas de controle por meio da inclusão de pontos de verificação de segurança, LEE, SHIM e EUN, 2018, explicam em seu trabalho que, com base em um banco de observações parciais, aspectos de segurança podem ser observados de modo a comporem indicadores de segurança e aumentarem a capacidade de detecção de ataques realizadas por esses sistemas.

3.2 Segmentação de Rede

Explorando a estratégia de adotar observações parciais para analisar o todo, YANG et al., 2019 investigaram várias estruturas de rede para mostrar que, de fato, redes podem ser completamente monitoradas. Em um estudo voltado para a análise de ataques de espionagem, os autores implementaram um sistema de coleta no qual cada dispositivo avaliado envia informações de forma independente para um centralizador. Neste cenário, testes foram realizados considerando casos em que alguns dos dispositivos não eram monitorados, ou outros eram inseridos sem o conhecimento do sistema de segurança. Na abordagem proposta pelos autores, o monitoramento completo do ambiente foi obtido por meio do agrupamento de dispositivos em grupos de análise.

Em redes de computadores, uma das estratégias utilizadas para realizar o agrupamento de elementos é a segmentação de rede, que consiste na divisão da rede em diferentes partes, ou segmentos, por meio das quais controles e limitações podem

ser aplicados de acordo com o que compõe determinado segmento (DAN; BRETT, 2016). Embora seja considerado um mecanismo de segurança, no trabalho desenvolvido por WAGNER et al., 2016, os autores abordaram o fato de que o método utilizado para segmentar determinada rede, diante as diversas possibilidades de se fazê-lo, recai sobre o julgamento e expertise do analista responsável pelo processo de segmentação. Com o intuito de auxiliar na tomada de decisão sobre qual método de segmentação melhor se aplica ao contexto da rede, os autores propuseram um método que combina teste de mesa, modelagem hierárquica, análise estatística e técnicas de otimização para pesquisar dentre os modos segmentação aquele considerado ótimo, ou quase ótimo, para a rede.

Ainda, destinado a ambientes de rede que utilizam Tradução de Endereço de Rede (NAT, sigla em inglês), CRICHIGNO et al., 2019, propuseram em seu trabalho a caracterização de redes de pequeno e médio porte por meio da análise da entropia presente nos fluxos da rede. A análise realizada pelos autores identificou que, ao longo do dia, a entropia das redes pode variar significativamente. Diante disso, ao caracterizar os fluxos individualmente, os autores observaram que comportamentos associados a ataques produzem entropias que se desviam do padrão esperado, característica que, segundo eles, pode ser utilizada no processo de detecção de intrusão de anomalias, embora não tenha sido explorada pelos autores tal aplicação do método proposto.

Além disso, considerando o fato de que a identificação e manutenção manual de segmentos de rede são impraticáveis em redes que estão sujeitas a alterações constantes, em seu trabalho, SMERIGA e JIRSIK, 2019, exploram a segmentação automatizada da rede por meio da utilização de fluxos de rede, e de técnicas de aprendizado de máquina. A abordagem proposta pelos autores se baseia no agrupamento de elementos que compõem a rede, por meio da análise de seus comportamentos, neste caso, representados pelo fluxo. Por fim, dado o surgimento na rede de um dispositivo não conhecido, o modelo proposto pelos autores foi capaz de associá-lo a um segmento já existente com mais de 92% de acurácia.

3.3 Avaliação de Sistemas de Detecção de Intrusão

Voltando-se para a análise de IDSs, PARK e AHN, 2017 apresentaram em seu trabalho um estudo comparativo entre os IDSs SNORT e Suricata, dois dos IDSs mais utilizados em estratégias de segurança de rede, segundo os autores. A análise comparativa realizada no trabalho se deu por meio da avaliação da taxa de processamento de ambos os sistemas quando aplicados em ambientes *single-threat* ou *multi-threat*. O IDS considerado melhor, de acordo com o estudo realizado, foi aquele cujo nível de utilização da CPU (Unidade Central de Processamento, sigla em inglês) foi o menor durante a realização das análises e detecções de eventos no ambiente avaliado.

KWON et al., 2018, avaliaram a utilização de Redes Neurais Convolucionais (CNN, sigla em inglês) para detectar comportamentos anômalos em redes de computadores. No trabalho desenvolvido pelos autores foram avaliados três modelos simples de CNN, ora diferenciados pelo nível de profundidade nas camadas que os compõem. Após treinar cada modelo e aplicá-los nos conjuntos de testes, os autores avaliaram sua eficiência por meio da métrica F, apresentada na Equação 5. Segundo os autores, apesar da utilização de CNNs para construção de IDSs baseados em anomalia produziram bons resultados - pontuação F1 próxima a 80% sobre a base de fluxos NSL-KDD, os autores concluíram que tais resultados não estão relacionados à profundidade das camadas que compõem o modelo, visto que o melhor resultado foi obtido por meio da arquitetura mais rasa implementada pelos autores.

Também, em um estudo abrangente sobre IDSs baseados em aprendizado de máquina, CHAPANERI e SHAH, 2019, realizaram diversas comparações entre os diferentes modelos avaliados considerando fatores como: cenários de ataques de rede representados nos conjuntos de dados avaliados, técnicas de transformação e seleção de atributos aplicadas nos mesmos, e a taxa de acerto obtida pelos diferentes modelos de detecção. Além de apresentarem uma visão geral das técnicas de aprendizado de máquina aplicadas nos trabalhos que fizeram parte do estudo, apesar da obtenção de acurácias acima de 90% obtida pelos modelos, os autores chamaram a atenção para desafios ainda presentes neste campo de estudo, como o grande volume de dados a ser tratado, o alto custo associado à ocorrência de falsos positivo, e a ausência de

conjuntos de dados balanceados de amostras de comportamentos anômalos e reais em redes.

Ainda, com o propósito de avaliar o desempenho de IDSs baseados em rede, PÉREZ et al., 2019, em seu trabalho, compararam quatro algoritmos de detecção baseados em anomalia. Para cada algoritmo, os autores observaram a eficácia dos sistemas por meio de métricas bem conhecidas, como acurácia, e área sobre a curva ROC. Utilizando a estratégia de Análise dos Componentes Principais (PCA, sigla em inglês) para validar os diferentes modelos de IDS avaliados, o melhor resultado obtido pelos autores foi uma acurácia de 87%. As bases de fluxo de rede analisadas pelos autores foram: UNSW-NB15 (MOUSTAFA; SLAY, 2015), NSL-KDD (TAVALLAEE; BAGHERI; GHORBANI, 2009), CIC-IDS-2017 (SHARAFALDIN; LASHKARI; GHORBANI, 2018) e Kyoto (SONG et al. 2011).

Além disso, a fim de sanar a ausência de métricas e metodologias que calculem a observabilidade e a eficiência de uma estratégia de monitoramento de segurança, HALVORSEN, WAITE e HAHN, 2019, propuseram em seu trabalho a Ferramenta de Avaliação de Observabilidade e Monitoramento de Ameaças (TOMATO, sigla em inglês). Por meio da coleta de dados de diferentes fontes, como fluxos e registros dos sistemas, a ferramenta desenvolvida pelos autores avalia o quão bem determinado sistema de detecção identifica técnicas de ataques em andamento na rede, considerando, para tal, a quantidade de falsos positivo e os componentes da rede sendo observados.

Por fim, KUMAR, DAS e SINHA, 2020 avaliaram em seu trabalho a capacidade de detecção de IDSs segundo sua acurácia, sensibilidade, e taxa de falso positivo. Os autores avaliaram a performance de diferentes modelos de aprendizado de máquina aplicados para a detecção das diferentes classes de ameaças definidas na base UNSW-NB15. Explorando a análise dessas ameaças em fluxos gerados por dispositivos IoT e, utilizando a estratégia de ganho de informação para seleção dos atributos utilizados para classificar os eventos, o modelo desenvolvido pelos autores, utilizando o algoritmo de floresta aleatória, obteve como melhor resultado a taxa de acurácia de 89,86%.

3.4 Considerações Finais

Diante dos estudos apresentados, nota-se a aplicação e o desenvolvimento de diferentes formas de avaliação de IDSs e outros sistemas de segurança, assim como a utilização de estratégias de segmentação de rede como meio para agregar segurança e monitorar ambientes alvos de ataques, como é o caso dos ambientes de rede. No entanto, enquanto alguns trabalhos exploram a avaliação de IDSs em contextos restritos, como os associados a dispositivos IoT, outros fazem uso de diferentes fontes de dados para realizar tais análises. Ainda, no levantamento realizado observa-se a carência de trabalhos voltados para a avaliação de IDSs baseados em fluxo que utilizem o método de detecção por abuso, ou assinatura, como também é conhecido, para o desenvolvimento e análise dos modelos propostos. Desde modo, por meio da análise dos trabalhos expostos, o trabalho desenvolvido apresenta sua contribuição ao explorar uma abordagem de avaliação de IDSs baseados em fluxo que compreende alguns dos aspectos não abordados nos trabalhos correlatos, sejam de forma independente ou correlacionada, ressaltando a justificativa para o desenvolvimento do trabalho apresentada na Seção 1.2.

CAPÍTULO 4 - Metodologia

Este capítulo tem como objetivo apresentar a metodologia proposta para a elaboração deste trabalho. Na Seção 4.1 é apresentada uma visão geral da metodologia utilizada para o desenvolvimento do trabalho. Adentrando nas partes que compõem a mesma, na Seção 4.2 é apresentado como o *framework* MITRE ATT&CK a integra, e o modo como o mesmo foi utilizado para a geração das assinaturas das técnicas de ataque analisadas no trabalho. Na Seção 4.3 são abordadas as fontes de dados utilizadas para a avaliação do trabalho. Nas Seções 4.4, 4.5 e 4.6 são apresentados o método utilizado na estratégia de segmentação da rede, o método utilizado para o treinamento e validação dos modelos explorados no trabalho e, por fim, o método utilizado para avaliar os modelos que representam os IDS analisados, respectivamente.

4.1 Visão Geral

Para o treinamento e avaliação dos modelos utilizados no trabalho, foram mapeadas técnicas de ataques descritas no *framework* MITRE ATT&CK cujos comportamentos são perceptíveis utilizando como fonte de dados apenas o fluxo de rede. De posse do padrão de comportamento das técnicas mapeadas, os mesmos foram utilizados de modo distinto para os dois conjuntos de dados explorados no trabalho.

Para o primeiro conjunto de dados, correspondente à base de fluxos UNSW-NB15, os comportamentos mapeados a partir do MITRE ATT&CK foram utilizados para, dentre as nove categorias de ataques classificadas na base UNSW-NB15, selecionar aquelas correspondentes aos padrões de comportamento mapeados. Para o segundo conjunto de dados, correspondente ao fluxo coletado das unidades de ensino da Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), os comportamentos mapeados a partir do *framework* foram utilizados para identificar, a partir de suas assinaturas correspondentes, a presença, ou não, dos eventos de interesse na rede. Posteriormente, para o segundo conjunto de dados, as assinaturas foram utilizadas para rotular os fluxos que o compõem, assim, dando origem à base de fluxos utilizada durante os experimentos desenvolvidos no trabalho, denominada ACME'21.

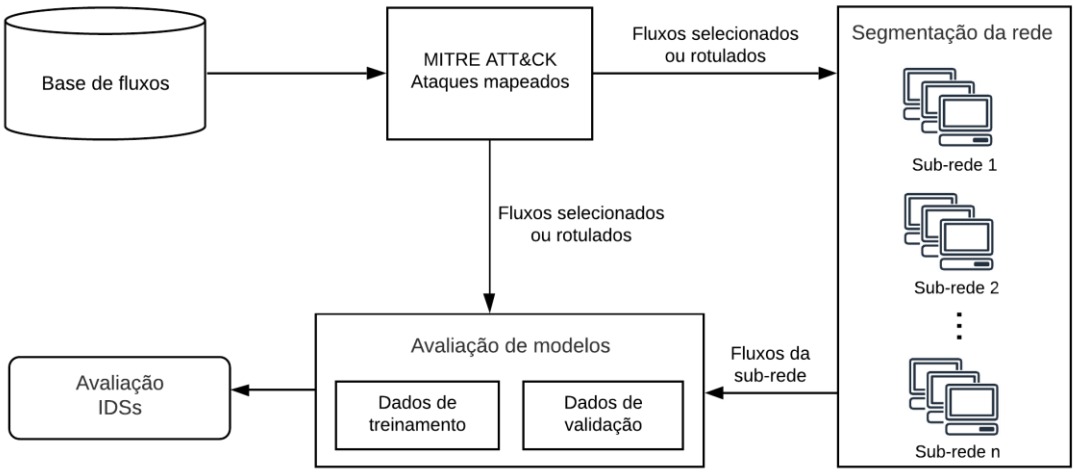
Além disso, diferentemente de IDSs que realizam análises sobre cada dispositivo - HIDS, IDSs baseados em fluxo possuem a capacidade de obter todos os dados que descrevem o comportamento da rede a partir de um único ponto de exportação. Com o intuito de explorar esta vantagem do ponto de vista de coleta, os fluxos que compõem o segundo conjunto de dados foram exportados de apenas um ponto por meio do qual todos os dados de fluxo das diferentes unidades da UNESP passam através. A coleta, armazenamento e análise destes fluxos foram realizados por meio da Pilha Elastic.

O treinamento dos modelos que descrevem os IDSs avaliados, representados pelos algoritmos de aprendizado de máquina supervisionado *Naive Bayes*, KNN, *XGBoost* e TPOT, foram realizados, para os diferentes conjuntos de dados avaliados, a partir de todos os fluxos que compõem cada conjunto, e a partir dos fluxos que compõem cada segmento de rede gerado por meio dos agrupamentos formados mediante a aplicação da estratégia de segmentação de rede, ora realizada com o auxílio dos algoritmos de aprendizado de máquina não supervisionado *K-Means* e DBSCAN.

A avaliação dos IDSs foi realizada por meio da análise da acurácia média obtida por cada modelo quando aplicado sobre cada segmento de rede observado. Ainda, a fim de demostrar a contribuição da abordagem proposta, a acurácia obtida por meio da aplicação dos modelos sobre todo o conjunto de dados, ou seja, não

considerando a segmentação realizada, também foi observada. O diagrama geral de funcionamento do sistema está ilustrado na Figura 2.

Figura 2: Diagrama geral de funcionamento do sistema.



Fonte: produzida pelo próprio autor.

4.2 MITRE ATT&CK

Para cada uma das técnicas descritas no *framework* MITRE ATT&CK, foram analisados seu comportamento, método de detecção, e as fontes de dados por meio das quais a detecção das mesmas são realizadas. Apesar do nível de detalhamento fornecido pelo *framework* para cada uma das técnicas enumeradas, o fato do mesmo ter seu desenvolvimento baseado na identificação e mapeamento de APTs faz com que, naturalmente, diversas fontes de dados, além do fluxo, sejam consideradas para a detecção das ameaças, tais como: registros de chamadas do sistema, monitoramento de processos, registros de eventos, entre outras.

Com isso, embora a agregação e correlação de diferentes fontes de dados forneçam maiores detalhes sobre o ambiente sendo analisado, a análise de dados além daqueles existentes no fluxo possui desvantagens, algumas das quais foram apontadas na Seção 1.2. Em vista disso, dentre todas as técnicas descritas no *framework* MITRE ATT&CK, foram objetos de estudo neste trabalho apenas aquelas cuja detecção pode ser realizada por meio de informações presentes no fluxo, única e exclusivamente.

Na Tabela 2 são apresentadas as técnicas do MITRE ATT&CK que podem ser identificadas por meio da análise do fluxo, a qual tática a técnica está associada, seu identificador junto ao *framework* (ID da técnica), e quais atributos do protocolo *Netflow* foram utilizados para determinar o comportamento da mesma na rede.

Tabela 2: Técnicas identificadas utilizando *Netflow*.

Tática	**Técnica**	**ID da Técnica**	**Atributo**
Persistência	Sinalização de tráfego: batida na porta	T1205.001	srcip, srcport, dstip, dstport, flags, time
Credencial de acesso	Autenticação forçada	T1187	srcip, dstip, flags, bytes, time, dstport {21,22,23,445}
Descoberta	Varredura de serviço de rede	T1046	dstip, flags, time, dstport {1-65535}
	Descoberta de compartilhamento em rede	T1135	dstip, flags, time dstport {21,445}
	Descoberta de sistemas	T1018	scrip, dstip, flags, proto, time
Comando e controle	Sinalização de tráfego: batida na porta	T1205.001	srcip, srcport, dstip, dstport, flags, time
Exfiltração	Limites de tamanho de transferência de dados	T1030	srcip, dstip, bytes, time
Impacto	Negação de serviço: exaustão do SO	T1499.001	dstip, bytes, time
	Negação de serviço: exaustão de serviço	T1499.002	dstip, dstport, time

Fonte: elaborado pelo próprio autor.

Para cada técnica elencada na Tabela 2, foram definidas assinaturas que representam seu comportamento esperado na rede por meio de condições e comparações que fazem uso de seus atributos correlatos. A especificação da assinatura gerada para cada uma das técnicas elencadas está descrita nas Seções 4.2.1 até 4.2.8.

4.2.1 Assinatura de sinalização de tráfego: batida na porta - T1205.001

A técnica sinalização de tráfego: batida na porta, baseia-se no estabelecimento prévio de uma determinada sequência de conexões antes que a conexão objetivo seja estabelecida. Comumente aplicadas em táticas de persistência, e comando e controle, a assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para um mesmo destino;
2. Ocorridos em intervalos de um minuto;
3. Envolvendo entre duas e sete portas de destino diferentes;
4. Com a *flag* SYN e/ou ACK habilitada;
5. A média de *bytes* enviados no intervalo observado seja menos que 150;
6. E com a quantidade máxima de *bytes* transmitidos em um fluxo do conjunto de fluxos enviados no intervalo observado maior que a média de *bytes* transmitidos nos demais fluxos enviados no mesmo intervalo. Deste modo, garante-se que, a princípio, o comportamento observado não seja confundido com um *scan*.

4.2.2 Assinatura de autenticação forçada - T1187

A técnica de autenticação forçada, também conhecida como força bruta ou ataque de dicionário, é baseada na tentativa de se obter acesso ao alvo por meio da adivinhação de suas credenciais. Pertencente a tática de credencial de acesso, esta técnica tem sua aplicação direcionada principalmente a serviços que fazem uso de

métodos de autenticação, como FTP, HTTP, HTTPS, SMB, SSH e *Telnet*. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para um mesmo destino;
2. Ocorridos em intervalos de três minutos;
3. Direcionados a portas de destino conhecidas por fazerem uso ou implementarem sistemas de autenticação, como FTP (21), SSH (22), *Telnet* (23), HTTP (80), HTTPS (443) e SMB (445);
4. Envolvendo vinte ou mais fluxos direcionados à mesma porta de destino, ou seja, vinte ou mais tentativas;
5. Utilizando o protocolo TCP;
6. Com a *flag* SYN ou ACK habilitada;
7. E a quantidade média de *bytes* menor que 150.

4.2.3 Assinatura de varredura de serviço de rede - T1046

A técnica de varredura de serviço de rede é utilizada por atacantes para obter a lista de serviços em execução no alvo. Como parte importante da tática de descoberta de conhecimento, está técnica se baseia na verificação da existência de portas do alvo que possuem serviços ativos, ou não. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para um mesmo destino;
2. Ocorridos em intervalos de cinco minutos;
3. Direcionados a trinta ou mais portas de destino diferentes;
4. Utilizando o protocolo TCP ou UDP;
5. Contendo apenas a *flag* SYN, ACK ou FIN habilitada, as *flags* FIN, PSH e URG habilitadas simultaneamente, ou nenhuma *flag* habilitada;
6. E a quantidade média de *bytes* menor que 150.

4.2.4 Assinatura de descoberta de compartilhamento em rede - T1135

Apesar de compor o conjunto de táticas de descoberta, a técnica de descoberta de compartilhamento em rede também está diretamente relacionada à tática de movimento lateral. Esta técnica se baseia na descoberta de serviços de compartilhamento presentes na rede e, uma vez executada com sucesso, pode levar à exfiltração, e perda da integridade e confidencialidade dos dados. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para um mesmo destino;
2. Ocorridos em intervalos de um minuto;
3. Direcionados apenas a portas de destino conhecidas por fornecerem serviços de compartilhamento, como FTP (21) e SMB (445);
4. Utilizando o protocolo TCP;
5. Contendo apenas a *flag* SYN ou ACK habilitada;
6. E a quantidade média de *bytes* menor que 150.

4.2.5 Assinatura de descoberta de sistemas - T1018

A técnica de descoberta de sistemas consiste na busca por sistemas ativos dentro da superfície de ataque estabelecida pelo atacante. Com o objetivo de conhecer melhor o ambiente alvo, esta técnica se relaciona diretamente ao mapeamento de endereços IPs. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para diferentes destinos;
2. Ocorridos em intervalos de cinco minutos;
3. Direcionados a vinte ou mais destinos diferentes;
4. Utilizando o protocolo ICMP;
5. Com a *flag* SYN e/ou ACK habilitada, ou nenhuma *flag* habilitada;
6. E a quantidade média de *bytes* menor que 100.

4.2.6 Assinatura de limites de tamanho da transferência de dados - T1030

A técnica de limites de tamanho da transferência de dados, associada à tática de exfiltração, delimita o tamanho dos pacotes utilizados no processo de transferência de grandes volumes de dados ou comandos de controle, de modo que evadam sistemas de detecção. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos partindo de uma mesma origem, para um mesmo destino;
2. Ocorridos em intervalos de cinco minutos;
3. Envolvendo dez ou mais fluxos direcionados à mesma porta de destino;
7. Com a *flag* SYN ou ACK habilitada.
4. E com a quantidade de *bytes* transmitido em cada fluxo do intervalo analisado fixa, como 23, 1500 ou 2048 *bytes*.

4.2.7 Assinatura de negação de serviço: exaustão do SO - T1499.001

A técnica de negação de serviço: exaustão do SO, associada à tática de impacto, resume-se em tornar o sistema operacional indisponível para uso por meio do esgotamento de seus recursos. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos para um mesmo destino;
2. Ocorridos em intervalos de cinco minutos;
3. Cuja média de *bytes* transmitidas no último intervalo analisado seja ao menos duas vezes maior que a média de *bytes* transmitidos nos últimos seis intervalos analisados (trinta minutos);
4. E a média de *bytes* transmitidas no próximo intervalo analisado seja ao menos tão grande quanto o último ou zero, caracterizando a continuação do ataque, ou o sucesso em tornar o sistema operacional indisponível.

4.2.8 Assinatura de negação de serviço: exaustão de serviço - T1499.002

A técnica de negação de serviço: exaustão de serviço, associada à tática de impacto, fundamenta-se em tornar o serviço indisponível para uso por meio do esgotamento de sua capacidade de atender requisições. A assinatura que descreve o comportamento desta técnica consiste em:

1. Identificar fluxos para um mesmo destino e uma mesma porta;
2. Ocorridos em intervalos de cinco minutos;
3. Cuja média de *bytes* transmitidas no último intervalo analisado seja ao menos duas vezes maior que a média de *bytes* transmitidos nos últimos seis intervalos analisados (trinta minutos);
4. Ou a média de fluxos transmitidas no último intervalo analisado seja ao menos duas vezes maior que a média de fluxos transmitidos nos últimos seis intervalos analisados (trinta minutos);
5. E a média de *bytes* ou fluxos transmitidos no próximo intervalo analisado seja ao menos tão grande quanto o último ou zero, caracterizando a continuação do ataque ou o sucesso em tornar o serviço indisponível.

4.3 Base de Fluxos

Os dados de fluxo utilizados no trabalho são provenientes de duas fontes diferentes, sendo elas: o fluxo coletado e classificado a partir da rede UNESP e a base de fluxos UNSW-NB15. A utilização da base de fluxos UNSW-NB15 teve como objetivo apresentar a eficácia da abordagem proposta para a avaliação de IDS diante um conjunto de fluxos que contém a representação de ameaças de rede bem conhecidas, e amplamente utilizada na literatura. Já o conjunto de fluxos coletado e classificado a partir da rede UNESP teve como objetivo explorar a eficácia da avaliação proposta em uma rede de grande porte, e cujo tráfego agrega atributos do protocolo *Netflow* V9, características não presentes na base UNSW-NB15. Detalhes sobre cada uma das fontes de dados utilizadas neste trabalho são descritos nas Seções 4.3.1 e 4.3.2.

4.3.1 Base de fluxo UNSW-NB15

A base de fluxos UNSW-NB15 contém 43 atributos que compõem o vetor de características de cada fluxo, dos quais, apenas alguns são intrínsecos do fluxo, e todos os demais foram obtidos a partir do enriquecimento dos dados coletados. A lista dos atributos presentes na base, ora utilizados no desenvolvimento do trabalho, assim como uma breve descrição dos mesmos, pode ser observada na Tabela 3.

Tabela 3: Atributos da base UNSW-NB15.

Nome	**Descrição**
proto	Tipo de protocolo.
state	*Flags* habilitadas na conexão.
dur	Duração total do fluxo.
sbytes	*Bytes* transmitidos da origem para o destino.
dbytes	*Bytes* transmitidos do destino para a origem.
sttl	Duração do fluxo da origem para o destino.
dttl	Duração do fluxo do destino para a origem.
sloss	Pacotes retransmitidos ou descartados partidos da origem.
dloss	Pacotes retransmitidos ou descartados partidos do destino.
service	Nome do serviço associado ao fluxo.
sload	*Bits* por segundo partidos da origem.
dload	*Bits* por segundo partidos do destino.
spkts	Pacotes enviados da origem para o destino.
dpkts	Pacotes enviados do destino para a origem.
swin	Valor do anúncio da janela TCP de origem.
dwin	Valor do anúncio da janela TCP de destino.
stcpb	Número de sequência base do TCP de origem.
dtcpb	Número de sequência base do TCP de destino.
smean	Média de *bytes* por pacote transmitido pela origem.
dmean	Média de *bytes* por pacote transmitido pelo destino.
trans_depth	Profundidade do *pipeline* de requisições/respostas do fluxo.
res_bdy_len	Tamanho do conteúdo (*payload*) transmitido no fluxo.
sjit	Variação do atraso dos pacotes transmitidos pela origem.

djit	Variação do atraso dos pacotes transmitidos pelo destino.
sintpkt	Hora de chegada dos pacotes de origem no destino.
dintpkt	Hora de chegada entre pacotes de destino na origem.
tcprtt	Tempo de ida e volta da configuração da conexão TCP; soma de “synack” e “ackdat”.
synack	Tempo de configuração da conexão TCP; tempo entre os pacotes SYN e SYN_ACK.
ackdat	Tempo de configuração da conexão TCP; tempo entre os pacotes SYN_ACK e ACK.
is_sm_ips_ports	Indica se a origem e o destino do fluxo, assim como as portas envolvidas na conexão, são os mesmos.
ct_state_ttl	Número associado a cada “state” de acordo com a faixa específica de valores de “sttl” e “dttl”.
ct_flw_http_mthd	Número de pacotes do fluxo que utilizam o método GET ou POST presente no serviço “HTTP”; 0 (zero) é o valor padrão.
is_ftp_login	Indica se foi estabelecida uma sessão autenticada no serviço “FTP”.
ct_ftp_cmd	Número de pacotes de sessão “FTP” que possuem comandos.
ct_srv_src	Número de pacotes que contêm o mesmo “service” e “srcip”.
ct_srv_dst	Número de pacotes que contêm o mesmo “service” e “dstip”.
ct_dst_ltm	Número de pacotes partindo de “dstip” de acordo com “ltime”.
ct_src_ltm	Número de pacotes partindo de “srcip” de acordo com “ltime”.
ct_src_dport_ltm	Número de pacotes de mesmo “srcip” e “dsport” de acordo com “ltime”.
ct_dst_sport_ltm	Número de pacotes de mesmo “dstip” e “sport” de acordo com “ltime”.
ct_dst_src_ltm	Número de pacotes de mesmo “srcip” e “dstip” de acordo com “ltime”.
attack_cat	Categoria do ataque.
label	Classificação do fluxo: 0 (zero) para fluxo legítimo de usuários da rede; 1 (um) para fluxo malicioso associado a alguma das categorias observadas.

Fonte: adaptado de (MOUSTAFA e SLAY, 2016).

Além disso, a base UNSW-NB15 contém nove grupos de ataques classificados, além do grupo de fluxo descrito como “normal”, sendo esta última a classificação que designa os fluxos que descrevem comportamentos legítimos de usuários na rede. Apesar da quantidade de ataques classificados existentes na base, para este trabalho foram considerados apenas os fluxos cujo comportamento descrito se assemelha aos observados pelo *framework* MITRE ATT&CK, sendo eles: Análise, Porta dos fundos, DoS, Exploração, *Fuzzers* e Reconhecimento. A correlação entre as categorias presentes na base UNSW-NB15 e as técnicas do MITRE ATT&CK foi realizada de forma associativa por meio da análise técnica e descritiva das mesmas. A abordagem associativa adotada se deu em função da ausência de alguns atributos no vetor de características da base UNSW-NB15. Alguns dos atributos não presentes no conjunto de fluxos classificados que compõem a base são: IP de origem, IP de destino, Porta de origem e Porta de destino. A falta de tais atributos impediu que a estratégia de correlação por meio de assinaturas, aplicada na Seção 4.3.2, fosse replicada para a base UNSW-NB15.

Na Tabela 4 são apresentadas todas as categorias de ataque classificadas na base UNSW-NB15 e cujo comportamento condiz com alguma das técnicas enumeradas a partir do *framework* MITRE ATT&CK, o ID da técnica correspondente, e a quantidade de fluxos classificados existentes em cada categoria. Do total de 2.576.673 fluxos classificados que compõem a base UNSW-NB15, para o presente trabalho foram utilizados apenas 197.120 fluxos, sendo este último o total de fluxos correlacionados unicamente às categorias e técnicas de interesse apontadas na Tabela 4.

Tabela 4: Ataques classificados na base UNSW-NB15.

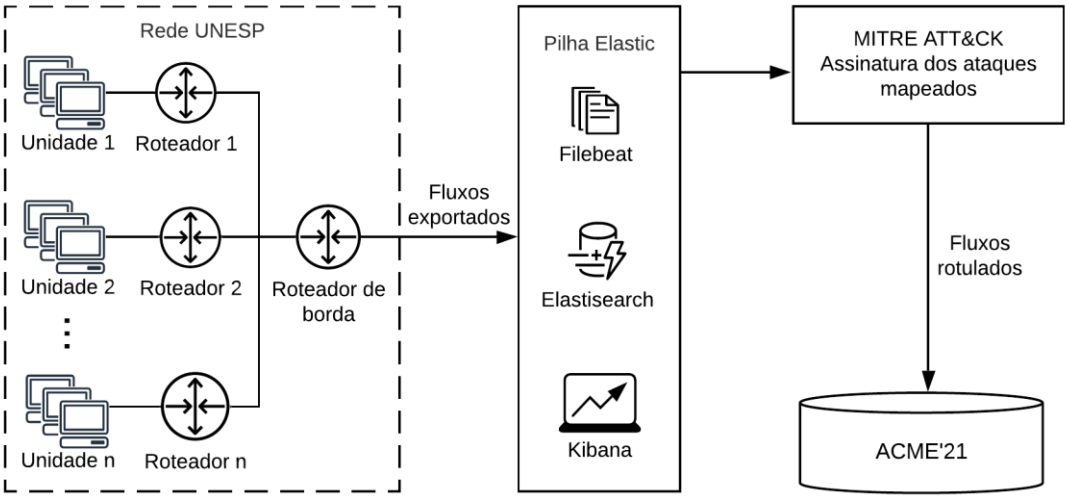
Categoria	**ID da técnica**	**Total de fluxos**
Análise	T1046	2.677
Porta dos fundos	T1205.001	2.332
DoS	T1499.001	16.353
Exploração	T1030	44.525
Fuzzers	T1499.002	24.246
Reconhecimento	T1018	13.987
Fluxo normal	---	93.000

Fonte: elaborada pelo próprio autor.

4.3.2 Coleta e classificação de fluxos da rede UNESP

A coleta do fluxo de rede das vinte e quatro unidades de ensino da rede UNESP foi realizada a partir de um único ponto de exportação por meio do qual todo tráfego gerado pelas unidades de ensino passa através. Este ponto de exportação é ilustrado na Figura 3 como o “Roteador de borda”. Também na Figura 3 está representada a rede UNESP e suas unidades de ensino a partir das quais os fluxos são coletados, a estrutura da Pilha Elastic utilizada para realizar a coleta, indexação, armazenamento e processamento dos fluxos exportados, assim como a aplicação das assinaturas geradas a partir das técnicas de ataque descritas nas Seções de número 4.2.1 até 4.2.8, sobre os fluxos coletados e armazenados. Esta última corresponde à etapa de processamento realizada para rotular os fluxos e criar a base utilizada nos experimentos realizados. O nome dado à base de fluxos coletados da rede UNESP, e classificados de acordo com as assinaturas derivadas do *framework* MITRE ATT&CK, é ACME’21.

Figura 3: Diagrama de coleta dos fluxos da rede UNESP.



Fonte: produzida pelo próprio autor.

Ao considerar o grande volume de dados de fluxo produzidos pela rede observada, com uma média de 50GB de dados gerados diariamente, foi definido um período de coleta a partir do qual os fluxos foram coletados e processados para a geração da base ACME’21. Duas horas do fluxo gerado pela rede observada foram coletados, englobando todos os fluxos gerados pela rede entre 00h00 e 01h59 do dia

13 de janeiro de 2021. O dia e o horário de coleta foram definidos de forma arbitrária. Já o tempo de coleta utilizado para a construção da base foi limitado pela quantidade de fluxos que o *hardware* utilizado no desenvolvimento do trabalho foi capaz de processar. O *hardware* utilizado consistiu em: um processador Intel Core i5-7300HQ de 2.50GHz com 4 núcleos, e 8GB de memória RAM.

Na Tabela 5 são exibidos os atributos oriundos do *Netflow* V9, bem como do processamento aplicado sobre os mesmos, que compõem o vetor de características dos fluxos que constituem a base ACME'21, assim como uma breve descrição dos mesmos.

Tabela 5: Atributos do fluxo coletado.

Nome	**Descrição**
time	Horário de exportação do fluxo.
srcip	Endereço IP de origem.
sport	Número da porta de origem.
dstip	Endereço IP de destino.
dsport	Número da porta de destino.
version	Versão do protocolo IP utilizado.
proto	Tipo de protocolo.
flags	*Flags* habilitadas na conexão.
dur	Duração total do fluxo.
bytes	*Bytes* transmitidos no fluxo.
pkts	Pacotes transmitidos no fluxo.
fwrd	Status de encaminhamento do fluxo.
svrf	Identificador da tabela de roteamento e encaminhamento (VRF) da origem do fluxo.
dvrf	Identificador da tabela de roteamento e encaminhamento (VRF) do destino do fluxo.
direction	Direção do fluxo.
bytes_per_sec	Média de b*ytes* transmitidos no fluxo por segundo.
pkts_per_sec	Média de pacotes transmitidos no fluxo por segundo.
category	ID da técnica do MITRE ATT&CK correspondente.
label	Classificação do fluxo: 0 para fluxo legítimo de usuários da

	rede; 1 para fluxo malicioso associado a alguma das técnicas observadas.

Fonte: elaborada pelo próprio autor.

Na Tabela 6 são apresentados os IDs das técnicas de ataque classificadas na base ACME'21, e suas respectivas táticas associadas, oriundas do *framework* MITRE ATT&CK, bem como a quantidade de fluxos classificados e associados a cada técnica.

Tabela 6: Técnicas de ataque classificadas na base ACME'21.

ID da técnica	**Tática**	**Total de fluxos**
T1018	Descoberta	28.141
T1030	Exfiltração	193
T1035	Descoberta	0
T1046	Descoberta	24.975
T1187	Credencial de acesso	286
T1205.001	Persistência, Comando e controle	0
T1499.001	Impacto	128.375
T1499.002	Impacto	54.275
Fluxo normal	---	4.236.550

Fonte: elaborada pelo próprio autor.

Um outro item que pode ser observado com relação aos fluxos que compõem a base é que, dentre aqueles relacionados a alguma técnica de ataque, os mesmos podem ser analisados de acordo com sua tática relacionada. Na Tabela 7 é apresentada a quantidade de fluxos presentes na base relacionados as táticas do MITRE ATT&CK que foram abordadas no trabalho.

A quantidade total de fluxos associados as táticas, 183.379, difere da quantidade total de fluxos associados as técnicas de ataque observadas, 236.245, devido ao fato de que a nível de rede, observando as características dos fluxos, um mesmo fluxo pode se enquadrar no comportamento esperado de mais de uma técnica de ataque. Este comportamento foi observado na base construída. No entanto, ao observar a classificação dos fluxos com relação a suas táticas, essa correlação

múltipla não acontece, ou seja, cada fluxo é associado unicamente a uma tática correspondente; uma exceção a esta correlação única entre fluxo e tática pode acontecer quando, ao analisar o *framework* MITRE ATT&CK, uma mesma técnica está relacionada a diferentes táticas, como é o caso da técnica Assinatura de sinalização de tráfego: batida na porta - T1205.001, que está relacionada às táticas de Persistência e Comando e controle. No entanto, na base construída não foram identificados fluxos relacionados a esta técnica de ataque.

Tabela 7: Táticas de ataque classificadas na base ACME'21.

Tática	**Total de fluxos**
Persistência	0
Comando e controle	0
Credencial de acesso	141
Descoberta	49.935
Exfiltração	138
Impacto	133.165

Fonte: elaborada pelo próprio autor.

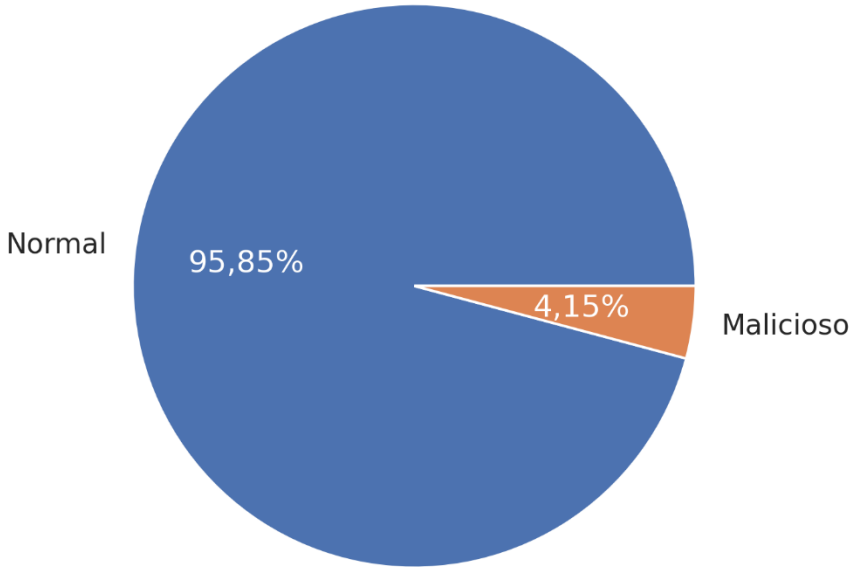
Ainda, vale destacar que, diferentemente da estratégia utilizada para a construção da base de fluxos UNSW-NB15, a qual, de forma híbrida, considerou o tráfego gerado pela rede observada, mas também executou ataques à rede de forma programada para gerar fluxos maliciosos de acordo com as classificações realizadas pelos autores; para a base de fluxos ACME'21, desenvolvida durante este trabalho, considerou-se para seu desenvolvimento apenas os fluxos gerados pela rede, ou seja, não foram executados ataques programados sobre a mesma durante o período de coleta estabelecido para sua composição.

Devido à estratégia utilizada para a construção da base ACME'21, nota-se a ausência de fluxos relacionados às técnicas 1035 e T1205.001, e a baixa ocorrência de fluxos relacionados as técnicas T1030 e T1187. Assim como se observa a ausência de fluxos relacionados às táticas de Persistência e Comando e controle. Tal comportamento demostra que, embora sejam técnicas e táticas de interesse a serem observadas, para a rede analisada, durante o período de coleta estabelecido, as respectivas técnicas não foram amplamente exploradas por atacantes, considerando o

padrão de assinatura definido para as mesmas. Um total de 4.472.795 fluxos compõem a base ACME'21.

Além disso, conforme exibido na Figura 4, também em decorrência da estratégia utilizada para a construção da base, nota-se que a mesma é composta majoritariamente por fluxos associados a atividades de uso normal da rede, 95,85%, enquanto isso, apenas 4,15% dos fluxos que compõem a base estão relacionados a alguma das técnicas de ataque sendo observadas.

Figura 4: Proporção de fluxo normal e malicioso na base ACME'21.



Fonte: produzida pelo próprio autor.

4.4 Segmentação da Rede

Em diversos cenários a segmentação de rede é aplicada como um meio para estabelecer regras e controles para os elementos que compõem cada segmento, deste modo, proporcionando maior domínio sobre a observação da ocorrência de eventos. Neste trabalho a avaliação dos IDSs é realizada por meio da análise de sua capacidade de detecção, ou seja, sua capacidade de identificar eventos de natureza maliciosa, quando são aplicados sobre os diferentes segmentos que compõem a rede.

Embora em redes de grande porte, ou mesmo redes menores, mas estruturadas, seja comum a utilização de segmentação a nível de rede, ao observar os fluxos trafegados, é necessário que de antemão sejam conhecidos os endereços IP de cada

elemento que compõe o segmento para que análises diferenciadas passam ser aplicadas sobre cada segmento. Com o intuito de evitar a necessidade deste conhecimento prévio, a estratégia de segmentação de rede utilizada neste trabalhou consistiu na definição dos segmentos por meio da análise comportamental dos elementos que compõem a rede. Para tal, sobre cada conjunto de fluxos analisados - bases de fluxos UNSW-NB15 e ACME'21 - foram aplicados os algoritmos de aprendizado de máquina não supervisionado, DBSCAN e *K-Means*, de modo que cada agrupamento de fluxo formado por meio da aplicação dos algoritmos sobre as bases que representam as redes analisadas configurou um segmento da rede. A escolha dos algoritmos se deu em função de sua aplicabilidade em estratégias automáticas de agrupamento de dados que não possuem rotulagens previamente conhecidas, como é o caso das bases de fluxos analisadas, com relação a sua segmentação.

Os segmentos formados a partir de cada base de fluxos são definidos por: $\{R, \{r_1, r_2, \ldots, r_n\} \epsilon R\}$, onde R é o conjunto de todos os fluxos que compõe a base de fluxos analisada, e r_n representa o conjunto de fluxos que compõem o segmento de rede n, originado por meio da estratégia de segmentação aplicada sobre o conjunto R, sendo $n \epsilon \{1, 2, \ldots, total\ de\ registros\ em\ R\}$.

Dentre os algoritmos utilizados para realizar a segmentação da rede, a avaliação daquele que melhor descreve os segmentos construídos foi realizada a partir do cálculo do Coeficiente de Silhueta (SC, sigla em inglês), definido pela Equação 6.

4.5 Treinamento e Validação dos Modelos que representam os IDSs

Os IDSs baseados em fluxo abordados neste trabalho foram representados por modelos que descrevem sua capacidade de identificação de eventos na rede. Os modelos explorados foram treinados e avaliados utilizando os seguintes algoritmos de aprendizado de máquina supervisionado: *Naive Bayes*, KNN, *XGBoost* e TPOT. A escolha dos algoritmos *Naive Bayes*, KNN e *XGBoost* ocorreu em função da utilização dos mesmos em trabalhos correlatos, alguns dos quais foram citados no Capítulo 3. Já a escolha do algoritmo TPOT se deu em função da abrangência de sua

aplicabilidade, uma vez que utiliza técnicas automatizadas de aprendizagem de máquina para solucionar problemas das mais diversas áreas de estudo.

Para os diferentes modelos que representam os IDSs, o modelo de treinamento de validação conhecido como *hold-out* foi empregado. Para treinamento dos modelos, foram utilizados 70% dos dados que compõem o conjunto de fluxos analisado. Essa proporção foi utilizada tanto para a análise de cada conjunto de dados como um todo, bem como para a análise realizada sobre cada um de seus segmentos formados. Já a avaliação dos modelos foi realizada utilizando 30% dos fluxos que compõem cada conjunto observado. De modo similar à etapa de treinamento, essa proporção foi utilizada tanto para a análise de cada conjunto de dados como um todo, bem como para a análise realizada sobre cada um de seus segmentos formados por meio da estratégia de segmentação.

Além disso, visto que o objetivo do IDS é identificar corretamente a presença de eventos de ameaças na rede, o critério utilizado para definir aquele que melhor desempenha tal papel foi a observação daquele que apresentou maior acurácia na detecção de tais eventos.

4.6 Avaliação dos Modelos que representam os IDSs

A acurácia obtida por cada modelo, para cada conjunto treinado e avaliado, representa o quão bem determinado IDS identifica os comportamentos observados dentro do conjunto. No entanto, as acurácias obtidas para os conjuntos que representam os segmentos da rede, quando analisadas isoladamente, não fornecem uma medida de avaliação que represente a qualidade do IDS para a rede como um todo. Para tal, foi definida a Acurácia Média dos Segmentos (AMS).

A AMS, definida na Equação 7, é uma métrica que, considerando a acurácia obtida para cada um dos segmentos da rede, calcula a acurácia média obtida por meio da aplicação de cada modelo sobre cada segmento do conjunto analisado. Desse modo, mensura a capacidade global do IDS de identificar ameaças na rede a partir da sua capacidade de observar de ameaças nos segmentos que a compõem.

$$AMS = \frac{Ar_1 + Ar_2 + \cdots + Ar_n}{n} = \frac{\sum_{i=1}^{n} Ar_i}{n} \qquad \text{(Equação 7.)}$$

sendo n o total de segmentos que compõem a rede, e Ar_i, a acurácia, A, do IDS, r, quando aplicado sobre o segmento de rede i. Para fins de comparação, também foi calculada a Acurácia Geral do Conjunto (AGC), a qual corresponde à acurácia obtida por meio da aplicação do IDS sobre todo o conjunto de dados, ou seja, desconsiderando os segmentos formados pela estratégia de segmentação proposta.

Ademais, a fim de evitar aspectos de enviesamento e agregar maior credibilidade sobre os resultados obtidos, a análise comparativa dos valores de AMS e AGC foi realizada a partir dos valores obtidos por meio da repetida aplicação do processo de Treinamento e Avaliação dos Modelos, descrito na Seção 4.5, sobre os conjuntos analisados. Foram realizadas dez repetições do processo descrito para se obter o valor das métricas, AMS e AGC, analisadas.

CAPÍTULO 5 - Experimentos e Resultados

Neste capítulo são apresentados os testes executados e os resultados obtidos por meio da abordagem proposta neste trabalho. Na Seção 5.1 são apresentados os experimentos executados e resultados obtidos durante a execução do processo de segmentação de rede aplicado sobre as bases de fluxo UNSW-NB15 e ACME'21. Na Seção 5.2 são descritos os experimentos de treinamento, validação e avaliação dos modelos que representam os IDS, assim como os resultados obtidos por meio da aplicação dos mesmos sobre as bases UNSW-NB15 e ACME'21. Por fim, na Seção 5.3 é apresentada a discussão dos resultados obtidos.

5.1 Segmentação da Rede

Os resultados da segmentação de rede realizada sobre os conjuntos de dados explorados no trabalho são apresentados na Tabela 8, na qual, para cada um dos algoritmos de agrupamento aplicado, sendo eles: DBSCAN e *K-Means*, é exibido o valor do melhor coeficiente de silhueta obtido para cada base de dados analisada, sendo elas: UNSW-NB15 e ACME'21.

Tabela 8: Resultado da segmentação da base UNSW-NB15.

Base de Fluxos	Algoritmo	Coeficiente de Silhueta
UNSW-NB15	DBSCAN	0,49
	K-Means	0,62
ACME'21	DBSCAN	0,39
	K-Means	0,92

Fonte: elaborada pelo próprio autor.

Para o algoritmo DBSCAN, a busca pela melhor configuração do algoritmo foi realizada por meio da alteração dos parâmetros *min_samples* e *eps*. O parâmetro *min_samples* define a quantidade mínima de fluxos necessários para compor o segmento, neste caso, foram atribuídos os valores $\{2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096\}$ ao parâmetro. Enquanto isso, para o parâmetro *eps*, que define a distância mínima entre dois fluxos para que um possa ser considerado, ou não, vizinho do outro e, consequentemente, passe a compor o mesmo segmento de rede, foram atribuídos os valores $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. A métrica de distância utilizada foi a Distância Euclidiana. Durante o processo de busca da melhor configuração do algoritmo, de acordo com os dados fornecidos como entrada para o mesmo, para cada valor de *min_samples* analisado, foram analisados cada um dos valores definidos para o parâmetro *eps*. Deste modo, foram exploradas 120 combinações diferentes de configuração do algoritmo. Os demais parâmetros utilizados pelo algoritmo DBSCAN foram mantidos em sua configuração padrão.

De maneira similar, para o algoritmo *K-Means*, a busca pela melhor configuração do algoritmo foi realizada por meio da alteração dos parâmetros *n_clusters* e *random_state*. O parâmetro *n_clusters* define a quantidade de segmentos formados pelo algoritmo, nesse caso, os valores assumidos pelo parâmetro foram $\{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100\}$. Para o parâmetro *random_state*, que define a aleatoriedade aplicada na geração dos centroides de inicialização dos segmentos, o que impacta diretamente a formação dos mesmos, foram atribuídos os valores $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Para cada valor de *n_clusters* analisado, também foram analisados cada valor definido para o parâmetro *random_state*. Deste modo, também foram exploradas 120 combinações diferentes

de configuração do algoritmo. Os demais parâmetros utilizados pelo algoritmo *K-Means* foram mantidos em sua configuração padrão.

Além disso, para a estratégia de segmentação de rede aplicada, foi realizado, para cada uma das bases de fluxo analisadas, um balanceamento entre os fluxos considerados maliciosos – fluxos relacionados a alguma das técnicas de ataque observadas - e os fluxos considerados normais – fluxos que remetem ao comportamento usual da rede. Em função disso, dentre o total de fluxos que compõem as bases analisadas, 197.200 e 4.472.795, para as bases UNSW-NB15 e ACME'21, respectivamente, foram mantidos, a partir desta etapa de segmentação, apenas 186.000 fluxos para a base UNSW-NB15, e 472.490 para a base ACME'21. A quantidade final de fluxos utilizada durante os experimentos remete ao maior conjunto balanceado de fluxos possível de ser obtido para cada base. Para a base UNSW-NB15, composta por 93.000 fluxos normais e 104.120 fluxos maliciosos, foram mantidos todos os fluxos normais para a composição do conjunto utilizado nos experimentos e, dentre os fluxos maliciosos, foram selecionados, de forma aleatória, outros 93.000 fluxos, totalizando os 186.000 fluxos considerados para este experimento. Para a base ACME'21, composta por 4.236.550 fluxos normais e 236.245 fluxos maliciosos, foram mantidos todos os fluxos maliciosos para a composição do conjunto utilizado nos experimentos e, dentre os fluxos normais, foram selecionados, de forma aleatória, outros 236.245 fluxos, totalizando os 472.490 fluxos considerados para este experimento e, consequentemente, para os experimentos descritos na Seção 5.2.

Os valores dos parâmetros utilizados em cada algoritmo, e que produziram os resultados apresentados na Tabela 8, assim como a quantidade de segmentos produzidos por cada respectiva configuração, quando aplica sobre a cada base de fluxos, são apresentados na Tabela 9.

Tabela 9: Parâmetros e valores utilizados na segmentação.

Base de Fluxos	Algoritmo	Parâmetro	Valor	Quantidade de Segmentos
UNSW-NB15	DBSCAN	*n_clusters*	16	10
		random_state	100	
	K-Means	*min_samples*	4	4
		eps	1	
ACME’21	DBSCAN	*n_clusters*	64	6
		random_state	70	
	K-Means	*min_samples*	2	2
		eps	1	

Fonte: elaborada pelo próprio autor.

5.2 Treinamento, Validação e Avaliação dos Modelos que representam os IDSs

Para o treinamento e validação dos modelos que representam os IDSs analisados, foi utilizada a estratégia *hold-out*, de forma que cada conjunto de dados, representados por meio das bases de fluxo UNSW-NB15 e ACME’21, foi dividido de modo que 70% do conjunto foi utilizado para o treinamento do modelo, e 30% do conjunto foi utilizado para a validação do mesmo.

Quanto à configuração dos modelos que representam os IDSs, os algoritmos *Naive Bayes* e *XGBoost* foram utilizados em suas respectivas configurações padrão, ou seja, não houve qualquer alteração nos parâmetros utilizados pelos algoritmos.

Ao considerar o algoritmo KNN, para o parâmetro *n_neighbors*, que define o número de vizinhos a serem observados pelo algoritmo a fim de realizar a classificação do fluxo observado, foi atribuído o valor 10. Também para o algoritmo KNN, a métrica de distância utilizada para calcular a proximidade entre vizinhos, definida pelo parâmetro *metric*, foi a métrica de Distância Euclidiana. Todos os demais parâmetros utilizados pelo algoritmo KNN foram mantidos em sua configuração padrão.

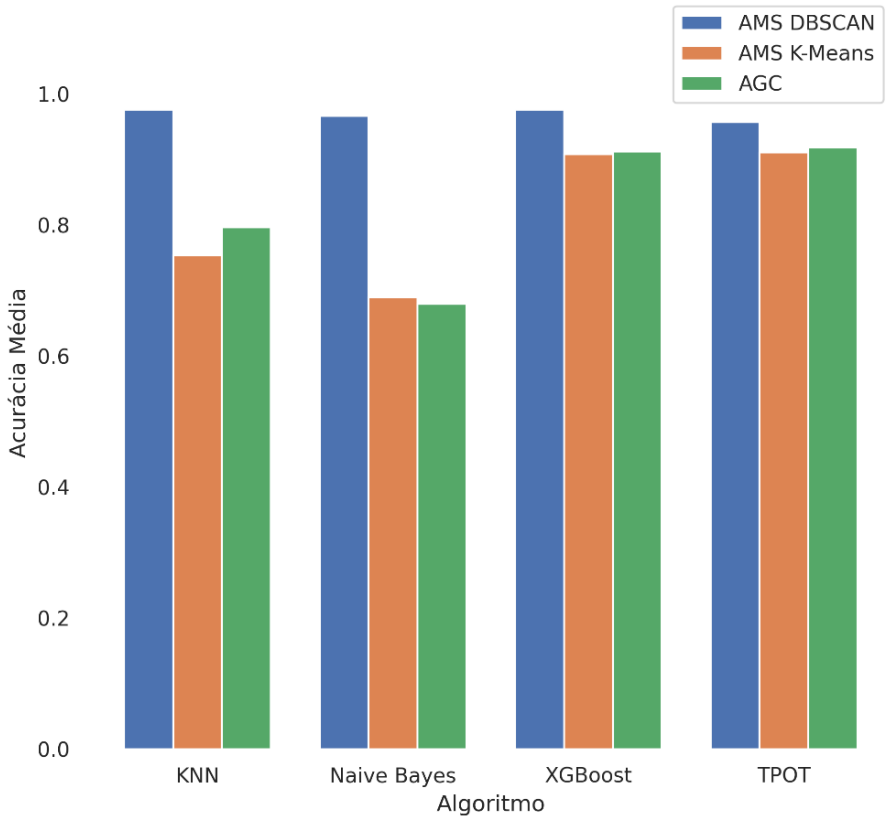
Para o algoritmo TPOT, algoritmo de aprendizagem de máquina automatizado explorado no trabalho, foi atribuído o valor 2 para o parâmetro *generations*, o valor 100 para o parâmetro *population_size*, e o valor 10 para o parâmetro *cv*. O parâmetro *generations* define a quantidade de vezes que o processo de busca pela configuração otimizada do *pipeline*[5] do algoritmo é realizado. Enquanto isso, o parâmetro *population_size* define quantos indivíduos de uma determinada geração, neste caso representados pelos fluxos de rede, são mantidos para compor a próxima geração. Já o parâmetro *cv* define quantas vezes a estratégia de validação *cross-validation* é aplicada sobre os *pipelines* formados pelo algoritmo para avaliar as configurações encontradas. Todos os demais parâmetros utilizados pelo algoritmo TPOT foram mantidos em sua configuração padrão.

Os valores definidos para os parâmetros mencionados para os diferentes algoritmos foram atribuídos de acordo com utilização dos mesmos em trabalhos correlatos, ou de acordo a utilização dos mesmos em sua respectiva documentação de uso, ou de forma empírica. Cada um dos algoritmos mencionados, com sua respectiva parametrização definida, foi executado 10 vezes sobre cada conjunto de dados, utilizando a estratégia *hold-out*, e seguindo a proporção 70/30, descrita no início desta seção. Os valores de AGC e AMS utilizados para avaliar os algoritmos de acordo a abordagem proposta foram definidos por meio do cálculo da média dos valores de AGC e AMS obtidos durante cada uma de suas 10 execuções.

Os resultados obtidos por meio do treinamento, validação e avaliação dos algoritmos que representam os IDSs, quando aplicados sobre a base de fluxos UNSW-NB15, são apresentados na Figura 5. Nela, para cada um dos modelos treinados e avaliados, sendo eles *Naive Bayes*, KNN, *XGBoost* e TPOT, é exibida a acurácia média obtida por meio da repetida aplicação do algoritmo sobre os segmentos formados (AMS) a partir de cada algoritmo aplicado na estratégia de segmentação utilizada, sendo eles DBSCAN e *K-Means*, e a acurácia média obtida por meio da repetida aplicação do algoritmo sobre todo o conjunto de dados (AGC).

[5] Fluxo de atividades que devem ser executadas, e respectivas configurações que devem ser utilizadas. Em aprendizagem de máquina automatizado envolve: limpeza de dados, processamento e seleção de características, seleção do modelo, otimização de parâmetros, e validação do modelo, entre outros.

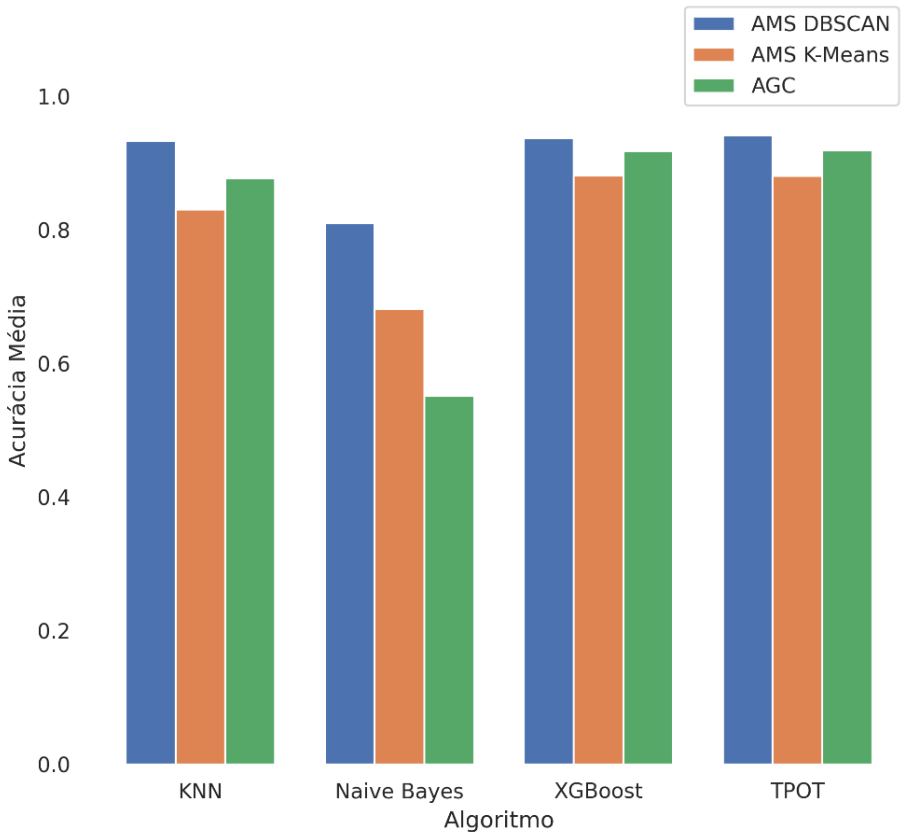
Figura 5: Resultado da avaliação dos IDSs sobre a base UNSW-NB15.



Fonte: elaborada pelo próprio autor.

De maneira similar, os resultados obtidos por meio do treinamento, validação e avaliação dos algoritmos que representam os IDSs, quando aplicados sobre a base de fluxos ACME'21, são apresentados na Figura 6. Nela, para cada um dos modelos treinados e avaliados, é exibida a acurácia média obtida por meio da repetida aplicação do algoritmo sobre os segmentos formados (AMS) a partir de cada algoritmo aplicado na estratégia de segmentação utilizada, e a acurácia média obtida por meio da repetida aplicação do algoritmo sobre todo o conjunto de dados (AGC).

Figura 6: Resultado da avaliação dos IDSs sobre a base ACME'21.



Fonte: elaborada pelo próprio autor.

5.3 Discussão dos Resultados

Para a segmentação da rede realizada sobre a base de fluxos UNSW-NB15, o melhor Coeficiente de Silhueta, 0,62, foi obtido por meio do algoritmo *K-Means*, utilizando *n_clusters* igual a 5, e *random_state* igual a 1, com uma média de 20.823 fluxos compondo cada segmento formado. Para a base de fluxos ACME'21, o melhor Coeficiente de Silhueta, 0,92, foi obtido por meio também do algoritmo *K-Means*, utilizando *n_clusters* igual a 2, e *random_state* igual a 1, com uma média de 93.000 fluxos compondo cada segmento formado. Os resultados do processo de segmentação de rede aplicado sobre as bases indicam que, mesmo para o mais disjunto conjunto de segmentos formado – Coeficiente de Silhuete mais próximo de +1, em ambas as bases analisadas há certa sobreposição sobre os fluxos que compõem seus respectivos segmentos, o que é esperado, visto que, a nível de rede¸ o

mesmo comportamento observado via fluxo pode se manifestar em diferentes segmentos da mesma rede.

Analisando os resultados obtidos por meio da aplicação da abordagem proposta para a avaliação de IDSs, expostos nas Figuras 5 e 6, para as bases UNSW-NB15 e ACME'21, respectivamente, observou-se que, em diversos cenários, a acurácia obtida por determinado modelo de detecção, quando avaliado utilizando a estratégia de segmentação de rede proposta, produziu ganhos na taxa de acurácia de detecção dos modelos, se comparado aos resultados obtidos pela avaliação realizada utilizando o método convencional, o qual não utiliza da estratégia de segmentação. Ao comparar o resultado obtido por meio do algoritmo *Naive Bayes* quando executado sobre os segmentos formados pelo algoritmo DBSCAN, com o resultado obtido pelo mesmo algoritmo quando aplicado sobre todo o conjunto de fluxos da base UNSW-NB15, observou-se um ganho de 09,57% na acurácia obtida ao utilizar a abordagem proposta. De maneira similar, a aplicação do algoritmo de detecção *Naive Bayes* sobre os segmentos formados pelo algoritmo DBSCAN produziu um ganho de 46,84% na taxa de acurácia da detecção realizada pelo modelo, se comparado ao resultado obtido pelo mesmo algoritmo quando aplicado sobre todo o conjunto de fluxos da base ACME'21, ou seja, utilizando a estratégia de avaliação convencional.

Ainda, para a segmentação realizada utilizando o algoritmo DBSCAN, e considerando todos os modelos avaliados, KNN, *Naive Bayes*, XGBoost e TPOT, observou-se para as bases UNSW-NB15 e ACME'21, um ganho médio de 03,15% e 10,02%, respectivamente, na acurácia obtida por meio das análises realizadas a partir da abordagem proposta, se comparado aos resultados obtidos por meio da avaliação dos modelos quando aplicados sobre todo o conjunto. Para a segmentação realizada utilizando o algoritmo *K-Means*, algoritmo por meio do qual foram obtidos os maiores Coeficientes de Silhueta para os segmentos formados, ao considerar todos os modelos avaliados, observou-se para as bases UNSW-NB15 e ACME'21 um ganho médio de 00,25% e 00,68%, respectivamente, na acurácia obtida por meio das análises realizadas a partir da abordagem proposta, se comparado à avaliação dos modelos quando aplicados sobre todo o conjunto.

A maior acurácia obtida durante os experimentos realizados sobre a base de fluxos UNSW-NB15 foi de 97,67%, correspondente à AMS obtida por meio do algoritmo KNN, quando executado sobre os segmentos de rede formados pelo

algoritmo DBSCAN. Enquanto isto, a maior acurácia obtida durante os experimentos realizados sobre a base de fluxos ACME'21 foi de 94,20%, correspondente à AMS obtida pelo algoritmo TPOT, quando executado sobre os segmentos de rede formados pelo algoritmo DBSCAN.

Diante dos resultados obtidos durante o processo de avaliação dos modelos que representam os IDSs, observou-se que os maiores ganhos na taxa de acurácia alcançados pela abordagem proposta ocorreram em função da aplicação dos algoritmos analisados sobre os segmentos formados pelo o algoritmo DBSCAN, para ambos os conjuntos de fluxo analisados. Com isso, embora os resultados obtidos tenham demonstrado a contribuição da abordagem em grande parte dos cenários analisados, o ganho de acurácia proporcionado pela mesma se demonstrou maior em segmentos de rede com maior sobreposição de elementos, haja visto que o Coeficiente de Silhueta obtido pelo algoritmo DBSCAN foi menor que o valor obtido pelo algoritmo *K-Means* para ambas as bases de dados, conforme exposto na Tabela 8. Os resultados obtidos também indicam que a criação de segmentos totalmente disjuntos, ou seja, que possuam um Coeficiente de Silhueta mais próximo do valor +1, faz com que cada segmento formado seja interpretado pelos algoritmos como redes diferentes, justificando os menores ganhos da abordagem proposta em tais cenários, uma vez que se assemelham às análises realizadas sobre todo o conjunto.

CAPÍTULO 6 - Conclusões

Este capítulo traz as conclusões obtidas para o trabalho proposto, os desafios encontrados, e propostas de atividades a serem desenvolvidas em trabalhos futuros.

6.1 Conclusões Gerais

Diante da necessidade de se desenvolver e aprimorar sistemas de segurança, como é o caso de IDSs, o desenvolvimento de métodos de avaliação que realcem sua eficiência se tornaram imprescindíveis. Ao explorar Sistemas de Detecção de Intrusão que utilizam o método de detecção baseado em assinaturas e em fluxo de rede, a abordagem de avaliação de IDSs proposta neste trabalho considerou em sua análise a estratégia de segmentação de rede como meio para realçar a eficiência de IDSs. A estratégia de segmentação abordada possibilitou que toda a rede, representada pelos conjuntos de dados analisados, fosse monitorada por meio da análise dos segmentos que a compõem, com ganhos na taxa de acurácia de detecção de eventos para os diferentes modelos de IDSs avaliados, quando comparados com a análise de toda a rede, ou seja, desconsiderando a segmentação de rede aplicada. Além disso, por meio da análise das técnicas apresentadas no *framework* MITRE ATT&CK, ora identificáveis utilizando apenas fluxo de rede como fonte de dados, o estudo garantiu que os diferentes modelos de IDS abordados fossem analisados diante de técnicas atuais, ressaltando sua aplicabilidade em cenários reais.

Os resultados obtidos mostraram que os modelos analisados seguindo a abordagem proposta tiveram ótimos resultados, alcançando mais de noventa por cento de acurácia na maioria dos casos analisados, considerando ambas as bases UNSW-NB15 e ACME'21. Os resultados também mostraram que as segmentações de rede com maiores Coeficientes de Silhueta, ou seja, as segmentações mais heterogêneas, alcançadas por meio do algoritmo *K-Means*, obtiveram resultados semelhantes à taxa de acurácia obtida por meio da análise de todo o conjunto de dados. Sendo assim, os maiores ganhos de acurácia obtidos ao utilizar a estratégia de segmentação proposta para avaliar os IDSs foram observados sobre segmentos gerados de forma mais homogênea por meio do algoritmo DBSCAN, o qual apresentou um ganho médio de 03,15% e 10,02%, na taxa de acurácia da detecção dos modelos, quando comparado aos resultados obtidos por meio dos IDSs quando aplicados sobre as bases UNSW-NB15 e ACME'21, respectivamente, como um todo, ou seja, desconsiderando a segmentação realizada sobre as mesmas.

Além disso, as maiores taxas de acurácia obtidas foram por meio da aplicação da abordagem proposta. Para a base de fluxos UNSW-NB15, foi obtida a acurácia máxima de 97,67%; já para a base de fluxos ACME'21, foi obtida a acurácia máxima de 94,20%. Para ambas as bases de fluxo analisadas os resultados obtidos pela abordagem proposta se mostraram de grande valia, uma vez que trabalhos de avaliação de IDS, como HALVORSEN; WAITE e HAHN, 2019, LEE, SHIM e EUN, 2018, CHAPANERI e SHAH, 2019, obtiveram taxas de acurácia em torno de 89,86%, 90% e 87%, respectivamente.

Portanto, considerando-se cenários reais, além de fornecer controles de segurança refinados, a aplicação da estratégia de segmentação de rede proposta neste trabalho para avaliar IDSs se mostrou válida diante o ganho nas taxas de acurácia de detecção obtidas para os diferentes modelos avaliados. Além disso, os resultados obtidos demonstraram-se melhores onde houve maior sobreposição sobre os segmentos, característica do fluxo de rede que se destaca em redes de médio e grande porte, ou mesmo em redes menores que possuem comportamentos e uso mais homogêneo, mostrando a contribuição da abordagem proposta também para tais cenários. Sendo assim, conclui-se que este projeto obteve sucesso em relação àquilo que foi proposto, seus objetivos foram integralmente cumpridos e se contribuiu em caráter inovador com a comunidade de pesquisa em segurança.

6.2 Trabalhos Futuros

Para trabalhos futuros, propõe-se uma análise estendida da abordagem proposta sobre conjuntos de dados maiores e também utilizando algoritmos de redes neurais (FAKER; DOGDU, 2019), a fim de explorar a escalabilidade da técnica de avaliação proposta.

Também propõe-se a extensão do conjunto de ataques observados e o refinamento das assinaturas descritas neste trabalho por meio da análise de outros *frameworks* e ferramentas de detecção, como é o caso dos *frameworks* MITRE Shield (MITRE, 2021), MITRE D3FEND (MITRE, 2021), e da ferramenta DeTTECT (DeTT&CT, 2021).

Por fim, considerando-se a influência da estratégia de segmentação sobre os resultados obtidos, sugere-se o desenvolvimento de trabalhos que objetivem identificar o nível de sobreposição ideal dos elementos que compõem os segmentos.

6.3 Dificuldades Encontradas

A principal dificuldade encontrada durante o desenvolvimento do trabalho foi a percepção da ausência de fontes de dados que contenham fluxos classificados de acordo com ataques recentes. Com isso, o desenvolvimento da base ACME'21 a partir dos fluxos coletados da rede UNESP foi fundamental para ampliar o espectro do trabalho proposto e aferir maior validade sobre os resultados obtidos.

Outra dificuldade encontrada foi com relação à parametrização das assinaturas que descrevem as técnicas de ataque observadas. Embora o *framework* MITRE ATT&CK possua um alto nível de detalhamento a respeito das técnicas de ataque, e proporcione informações a respeito das formas de detecção das mesmas, o *framework*, bem como o conjunto de trabalhos observados, não disponibilizam para todas as técnicas de ataque os valores e condições que devem ser observados para realizar a detecção das mesmas em uma abordagem baseada em assinatura, como a utilizada neste trabalho. Em vista disso, a definição empírica de alguns valores foi necessária para a construção das assinaturas utilizadas.

REFERÊNCIAS

ALI, M. H.; AL M., B. A. D.; ISMAIL, A.; e ZOLKIPLI, M. F. **A new intrusion detection system based on fast learning network and particle swarm optimization**. IEEE Access, 6, 20255-20261. 2018.

ALJAWARNEH, S.; ALDWAIRI, M.; YASSEIN, M. B. **Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model**. Journal of Computational Science, v. 25, p. 152-160, 2018.

BAO, H.; HE, H.; LIU, Z; LIU, Z. **Research on information security situation awareness system based on big data and artificial intelligence technology**. International Conference on Robots and Intelligent System (ICRIS). IEEE. p. 318-322. 2019.

BRASIL, **Lei No 12.965**. Brasília: Congresso Nacional, 2014. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>. Acesso em: 10 ago. 2020.

BRASIL, **Lei No 13.709**. Brasília: Congresso Nacional, 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 10 ago. 2020.

BURNS, J. **ATT&CK™ content available in STIX™ 2.0 via public TAXII™ 2.0 server**. Disponível em:<https://medium.com/mitre-attack/att-ck-content-available-in-stix-2-0-via-public-taxii-2-0-server-317e5c41e214>. Acesso em: 10 de ago. de 2020.

CERT.BR. **Incidentes Reportados ao CERT.br -- Janeiro a Dezembro de 2019**. Disponível em: < https://www.cert.br/stats/incidentes/2019-jan-dec/tipos-ataque.html>. Acesso em: 10 ago. 2020.

CHAPANERI, R.; SHAH, S. **A Comprehensive Survey of Machine Learning-Based Network Intrusion Detection**. In: Smart Intelligent Computing and Applications, p. 345-356. Springer. 2019.

CHIO, C.; FREEMAN, D. **Machine Learning and Security**. O'Reilly Media, 2018.

CLAISE, B. **Cisco systems netflow services export version 9**. 2004. Disponível em: <https://tools.ietf.org/html/rfc3954>. Acesso em: 10 ago. 2020.

CLAISE, B.; TRAMMELL, B.; AITKEN, P. **Specification of the IP flow information export (IPFIX) protocol for the exchange of flow information**. 2013. Disponível em: <https://tools.ietf.org/html/rfc7011>. Acesso em: 10 ago. 2020.

CRICHIGNO, J.; KFOURY, E.; BOU-HARB, E.; GHANI, N.; PRIETO, Y.; VEGA, C.; PEZOA J.; HUANG, C; e TORRES, D. **A flow-based entropy characterization of a NATed network and its application on intrusion detection**. In ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE. pp. 1-7. 2019.

DAN, M.; BRETT, Y. **ICS Network Segmentation**. SANS Institute. 2016. Disponível em: <https://www.sans.org/webcasts/ics-network-segmentation-101347>. Acesso em: 10 ago. 2020.

DeTT&CT. **Detect Tactics, Techniques & Combat Threats**. 2021. Disponível em: < https://github.com/rabobank-cdc/DeTTECT>. Acesso em: 01 ago. 2021.

DONG, Y.; CHOPRA, N. **Observability-based secure state encryption design for cyberphysical systems**. Indian Control Conference (ICC). IEEE. p. 24-29. 2018.

ERNST, J.; HAMED, T.; KREMER, S. **A survey and comparison of performance evaluation in intrusion detection systems**. In: Computer and network security essentials. Springer, Cham. p. 555-568. 2018.

ELASTICSEARH. **Elasticsearch introduction**. Elastic Co. Disponível em: <https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html#elasticsearch-intro>. Acesso em: 01 ago. 2021.

FAKER, O.; DOGDU, E. **Intrusion detection using big data and deep learning techniques**. In: Proceedings of the 2019 ACM Southeast Conference. p. 86-93, 2019.

FERREIRA, V. O. **Classificação de anomalias e redução de falsos positivos em sistemas de detecção de intrusão baseados em rede utilizando métodos de agrupamento**, 2016. Dissertação (Mestrado em Ciência da Computação) - Instituto de Biociências Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2016.

FILEBEAT. **Filebeat Reference**. Elastic Co. Disponível em: < https://www.elastic.co/pt/beats/filebeat>. Acesso em: 01 ago. 2021.

GARTNER, R. **What Metadata Is and Why It Matters**. In: Metadata. Springer, Cham, p. 1-13. 2016.

GONÇALVES, L. B. L. **Abordagem para geração automática de assinatura de ataques baseada em fluxos de redes de computadores**. 2019.

GURUNG, S.; GHOSE, M. K.; SUBEDI, A. **Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset**. International Journal of Computer Network and Information Security (IJCNIS), v. 11, n. 3, p. 8-14, 2019.

HALVORSEN, J.; WAITE, J.; HAHN, A. **Evaluating the Observability of Network Security Monitoring Strategies With TOMATO**. IEEE Access, v. 7, p. 108304-108315, 2019.

HINDY, H.; BROSSET, D.; BAYNE, E.; SEEAM, A.; TACHTATZIS, C.; ATKINSON, R.; e BELLEKENS, X. **A taxonomy and survey of intrusion detection system design techniques, network threats and datasets**. arXiv preprint arXiv:1806.03517. 2018.

HOQUE, N.; BHUYAN, M. H.; BAISHYA, R. C.; BHATTACHARYYA, D. K.; KALITA, J. K. **Network attacks: Taxonomy, tools and systems**. Journal of Network and Computer Applications, v. 40, p. 307-324, 2014.

HOSSAIN, M. D., OCHIAI, H., DOUDOU, F., e KADOBAYASHI, Y. **SSH and FTP brute-force Attacks Detection in Computer Networks: LSTM and Machine Learning Approaches**. In: 2020 5th International Conference on Computer and Communication Systems (ICCCS), p. 491-497. 2020.

IX.br. **Ponto de Troca de Tráfego da Internet Brasileira**. Disponível em: <https://ix.br/>. Acesso em: 01 ago. 2021.

JACKSON, J. **We Are Wisconsin Speech**. Wisconsin, 04 abr. 2011. Disponível em: <https://www.youtube.com/watch?v=n54ajG2GcnA>. Acesso em: 01 ago. 2021.

JOY, P.; MHAMDI, A.; MITSOS, A. **Optimization-based observability analysis**. Computers and Chemical Engineering. 2020.

KAKIHATA, E. M.; SAPIA, H. M.; OIAKAWA, R. T.; PEREIRA, D. R.; PAPA, J. P.; DE ALBUQUERQUE; V. H. C., e DA SILVA, F. A. **Intrusion detection system based on flows using machine learning algorithms**. IEEE Latin America Transactions, 15(10). 2017.

KANDAN, A. M.; KATHRINE, G. J.; MELVIN, A. R. **Network Attacks and Prevention techniques - A Study**. In: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), p. 1-6, 2019.

KIBANA. **Kibana**. Elastic Co. Disponível em: <https://www.elastic.co/pt/products/kibana>. Acesso em: 01 ago. 2021.

KHRAISAT, A., GONDAL, I., VAMPLEW, P., e KAMRUZZAMAN, J. **Survey of intrusion detection systems: techniques, datasets and challenges**. Cybersecurity, v. 2, n. 1, p. 20, 2019.

KUMAR, V.; DAS, A. K.; SINHA, D. **Statistical analysis of the UNSW-NB15 dataset for intrusion detection**. Computational Intelligence in Pattern Recognition. Springer, p. 279-294. 2020.

KUROSE, J. F.; ROSS, K. W. **Computer networking: a top-down approach**, 6th edition,. Addison-Wesley, 2016.

KWON, D.; NATARAJAN, K.; SUH, S. C.; KIM, H., e KIM, J. **An empirical study on network anomaly detection using convolutional neural networks**. IEEE 38th International Conference on Distributed Computing Systems (ICDCS). 2018.

LEE, C.; SHIM, H.; EUN, Y. **On redundant observability: from security index to attack detection and resilient state estimation**. IEEE Transactions on Automatic Control, v. 64, n. 2, p. 775-782, 2018.

LI, C.; WU, Y.; YUAN, X.; SUN, Z.; WANG, W.; LI, X., e GONG, L. **Detection and defense of DDoS attack–based on deep learning in OpenFlow-based SDN**. International Journal of Communication Systems, v. 31, n. 5, p. e3497, 2018.

LV, B.; YU, X.; XU, G.; YIN, Q., e SHI, Z. **Network Traffic Monitoring System Based on Big Data Technology**. In Proceedings of the 2018 International Conference on Big Data and Computing (pp. 27-32). ACM. 2018.

MATOUŠEK, P.; RYŠAVÝ, O.; GRÉGR, M. **Security Monitoring of IoT Communication Using Flows**. In: Proceedings of the 6th Conference on the Engineering of Computer Based Systems. ACM. p. 18. 2019.

MCKEOWN, N.; ANDERSON, T.; BALAKRISHNAN, H.; PARULKAR, G.; PETERSON, L.; REXFORD, J.; e TURNER, J. **OpenFlow: enabling innovation in campus networks**. ACM SIGCOMM Computer Communication Review, 38(2), 69-74. 2018.

MILLER, N. J.; ALIASGARI, M.. **Benchmarks for evaluating anomaly-based intrusion detection solutions**. California State University, Long Beach, 2018.

MITRE. **MITRE ATT&CK**. 2021. Disponível em: <https://attack.mitre.org/>. Acesso em: 01 ago. 2021.

MITRE. **MITRE D3FEND**. 2021. Disponível em: < https://d3fend.mitre.org/>. Acesso em: 01 ago. 2021.

MITRE. **MITRE Shield**. 2021. Disponível em: <https://shield.mitre.org/>. Acesso em: 01 ago. 2021.

MOUSTAFA, N.; SLAY, J. **The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set**. Information Security Journal: A Global Perspective, v. 25, p. 18-31, 2016.

MOUSTAFA, N., SLAY, J. **UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)**. Military Communications and Information Systems Conference (MilCIS), pp. 1–6. 2015.

OLSON, R. S.; MOORE, J. H. **TPOT: A tree-based pipeline optimization tool for automating machine learning**. In: Workshop on automatic machine learning. PMLR, p. 66-74, 2016. Disponível em: < http://epistasislab.github.io/tpot/>. Acesso em: 01 ago. 2021.

PANCHEN, S.; PHAAL, P.; MCKEE, N. **sFlow: A Method for Monitoring Traffic in Switched and Routed Networks**. RFC 3176. 2001. Disponível em: <https://tools.ietf.org/html/rfc3176>. Acesso em: 10 ago. 2020.

PARK, W.; AHN, S. **Performance comparison and detection analysis in snort and suricata environment**. Wireless Personal Communications, v. 94, n. 2, p. 241-252, 2017.

PATEL, S. K.; SONKER, A. **Rule-based network intrusion detection system for port scanning with efficient port scan detection rules using snort**. International Journal of Future Generation Communication and Networking, v. 9, n. 6, p. 339-350, 2016.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, v12, pp. 2825-2830, 2011. Disponível em: <https://jmlr.csail.mit.edu/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Acesso em: 01 ago. 2021.

PÉREZ, D.; ALONSO, S.; MORÁN, A.; PRADA, M. A.; FUERTES, J. J.; e DOMÍNGUEZ, M. **Comparison of Network Intrusion Detection Performance Using Feature Representation**. International Conference on Engineering Applications of Neural Networks, pp. 463-475. 2019.

REESE, R. M; REESE, J. L.; KALUZA, B.; KAMATH, U., e CHOPPELLA, K. **Machine Learning: End-to-End guide for Java developers: Data Analysis, Machine Learning, and Neural Networks simplified**. Packt Publishing, 1413 p, 2017.

SCHEERES, D. J.; ALFRIEND, K. T.; FRUEH, C. **Modeling Observability and Change Detection in Space Situational Awareness**. University of Colorado Boulder Boulder United States. 2018.

SHARAFALDIN, I.; LASHKARI, A.H.; GHORBANI, A.A. **Toward generating a new intrusion detection dataset and intrusion traffic characterization**. ICISSP, pp. 108–116. 2018.

SMERIGA, J.; JIRSIK, T. **Behavior-Aware Network Segmentation using IP Flows**. Proceedings of the 14th International Conference on Availability, Reliability and Security. p. 1-9. 2019.

SNORT. **SNORT - Network Intrusion Detection & Prevention System**. Disponível em: <https://www.snort.org/>. Acesso em: 10 ago. 2020.

SONG, J.; TAKAKURA, H.; OKABE, Y.; ETO, M.; INOUE, D., e NAKAO, K. **Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation**. Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, pp. 29–36. 2011.

STIX. **A structured language for cyber threat intelligence**. Disponível em: < https://oasis-open.github.io/cti-documentation/stix/intro>. Acesso em: 10 ago. 2019.

STROM, B. E.; APPLEBAUM, A.; MILLER, D. P.; NICKELS, K. C.; PENNINGTON, A. G., e THOMAS, C. B. **MITRE ATT&CK: Design and Philosophy**. MITRE Product MP, 18-0944. 2018.

STROM, B. E., BATTAGLIA, J. A., KEMMERER, M. S., KUPERSANIN, W., MILLER, D. P., WAMPLER, C., e WOLF, R. D. **Finding Cyber Threats with ATT&CK™-Based Analytics**. MITRE Technical Report MTR170202. The MITRE Corporation. 2017.

SULTANA, N.; CHILAMKURTI, N.; PENG, W., e ALHADAD, R. **Survey on SDN based network intrusion detection system using machine learning approaches**. Peer-to-Peer Networking and Applications, v12, pp. 493-501. 2019.

SURICATA. **Suricata | Open Source IDS / IPS / NSM engine**. Disponível em: < https://suricata-ids.org/>. Acesso em: 10 ago. 2020.

TAVALLAEE, M.; BAGHERI, E.; LU, W., e GHORBANI, A.A.: **A detailed analysis of the KDD CUP 99 data set**. Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications. 2009.

TAXII. **A transport mechanism for sharing cyber threat intelligence**. Disponível em: <https://oasis-open.github.io/cti-documentation/taxii/intro>. Acesso em: 10 ago. de 2020.

TERZI, D. S.; TERZI, R.; SAGIROGLU, S. **Big data analytics for network anomaly detection from netflow data**. International Conference on Computer Science and Engineering (UBMK). IEEE, p. 592-597. 2017.

VINAYAKUMAR, R.; ALAZAB, M.; SOMAN, K. P.; POORNACHANDRAN, P.; AL-NEMRAT, A.; e VENKATRAMAN, S. **Deep Learning Approach for Intelligent Intrusion Detection System**. IEEE Access. 2019.

WAGNER, N.; ŞAHIN, C. Ş.; WINTERROSE, M.; RIORDAN, J.; PENA, J.; HANSON, D.; e STREILEIN, W. W. **Towards automated cyber decision support: A case study on network segmentation for security**. IEEE Symposium Series on Computational Intelligence (SSCI), p. 1-10. 2016.

XGBOOST. **XGBoost Documentation**. Disponível em: https://xgboost.readthedocs.io/en/latest/index.html#. Acesso em: 10 ago. 2020.

YANG, W.; ZHENG, Z.; CHEN, G.; TANG, Y., e WANG, X. **Security analysis of a distributed networked system under eavesdropping attacks**. IEEE Transactions on Circuits and Systems II: Express Briefs. 2019.

ZHAO, H.; TANG, W.; ZOU, X.; WANG, Y., e ZU, Y. **Analysis of Visualization Systems for Cyber Security**. In Recent Developments in Intelligent Computing, Communication and Devices, pp. 1051-1061. 2019.