Aplicações de Machine Learning na Deteção de APT's

Alexandre Manuel Cruz Bastos
ISTEC
alexandre.bastor@my.istec.pt

Resumo

O ciberespaço tem sido constantemente ameaçado ao longo dos anos com o surgimento de novas ameaças, ataques e técnicas sofisticadas usadas pelos cibercriminais. Com o desenvolvimento das novas tecnologias e da Inteligência Artificial, novas formas de ciberataques têm surgido e novos métodos têm-se desenvolvido com o objetivo de proteger entidades contra este tipo de ameaças. Em particular, os Advanced Persistent Threats (APT) supõem uma ameaça severa no âmbito da cibersegurança devido aos métodos e ferramentas sofisticadas envolvidas nos mesmos e à constante adaptação que aplicam para conseguirem alcançar os seus objetivos e causar danos significativos nas organizações. Assim, vários métodos de Machine Learning são estudados e desenvolvidos com o intuito de proteger o meio digital da melhor forma possível. Neste artigo apresentam-se algumas soluções com o objetivo de demonstrar novas técnicas e métodos de ML para a deteção de APTs, e o contributo que as mesmas supõem nos dias de hoje no âmbito da cibersegurança.

Abstract

Cyberspace has been constantly threatened over the years with the emergence of new threats, attacks and sophisticated techniques used by cybercriminals. With the development of new technologies and Artificial Intelligence, new forms of cyberattacks have emerged and new methods have been developed in order to protect entities against this type of threats. In particular, Advanced Persistent Threats (APT) pose a severe threat in the field of cybersecurity due to the sophisticated methods and tools involved in them and the constant adaptation they apply to achieve their goals and cause significant damage to organizations. Thus, several Machine Learning methods are studied and developed in order to protect the digital environment in the best possible way. This article presents some solutions with the objective of demonstrating new ML techniques and methods for the detection of APTs, and the contribution that they represent nowadays in the field of cybersecurity.

Keywords: *CiberSegurança, Machine Learning, APT, algoritmos ML, métodos ML.*

Introdução

A cibersegurança é responsável por estabelecer políticas de segurança no ciberespaço, o que determina a forma com que se devem tratar os dados dentro de uma infraestrutura tecnológica e os limites que devem ser determinados. No entanto, existem vulnerabilidades no dia a dia como equipamentos não atualizados, uso de políticas antigas, etc., que permitem a intrusão de cibercriminais nas organizações e o comprometimento de dados valiosos.

O continuo desenvolvimento dos Sistemas de Informação e das novas tecnologias permitiu o acesso generalizado de usuários à Internet e a implementação de empresas no meio cibernético, o que despertou o interesse ao longo do tempo por parte deste tipo de criminais e, consequentemente, das suas práticas maliciosas. Estima-se que nos últimos anos tenham aumentado significativamente o número de ciberataques APT reportados pelas autoridades de cibersegurança [1] [2]. Desenvolveram-se ferramentas sofisticadas, como as vulnerabilidades "*Zero-Day*" e os ataques de Negação de Serviços (DoS), deixando para trás métodos de proteção que não conseguissem acompanhar a complexidade destes problemas.

Atualmente, os Advanced Persistent Threats, conhecidos como APTs, representam uma série de ameaças para o meio cibernético, deixando empresas e organizações a nível mundial em constante

preocupação com as suas infraestruturas digitais [3]. Tendo o objetivo de passar desapercebidos o maior tempo possível e de serem executados de forma sigilosa, estes ciberataques são difíceis de detetar/prever devido à utilização de ferramentas e métodos avançados. Estes ataques caracterizam-se pela sua estrutura, pela sua duração (podendo durar meses ou até anos) e pelos danos severos que os mesmos podem chegar a provocar numa organização, comprometendo os dados valiosos da mesma.

Tanto a comunidade académica como investigadores de todo o mundo propuseram diferentes abordagens ao estudo, tratamento, proteção e deteção deste tipo de ciberataques. Como conclusão, determinou-se que o Ciclo de Vida dos APTs pudesse ser um indicador para perceber como funcionam e como se desenvolvem [4] [5]. Graças a implementação do Machine Learning no âmbito da cibersegurança, foi possível o desenvolvimento de métodos que contribuíram de forma positiva para a prevenção, deteção e mitigação das ameaças que os APT supunham.

Neste artigo, abordam-se alguns métodos, algoritmos e ferramentas de Machine Learning como contributo para a proteção dos Sistemas de Informação no meio cibernético. Na secção 2, por um lado, descreve-se as características e a estrutura principal de um APT e, por outro, a definição dos algoritmos mais importantes de ML e as suas aplicações, o contributo do ML na deteção dos APTs e as suas aplicações. Na secção 3, procede-se a demonstração de modelos ML com o objetivo de demonstrar as suas estruturas e os algoritmos implementados, como as suas vantagens e desvantagens. Na secção 4, representa-se uma avaliação geral dos modelos ML e, por último, na secção 5, a conclusão do trabalho.

Metodologia

APT'S

O termo **Advanced Persisten Threat** (APT), traduzido como "Ameaça Avançada Persistente", é um tipo de ciberataque caracterizado pela sua finalidade e pelas etapas que o compõem. Por norma, o objetivo final deste tipo de ataques é aceder e comprometer informação confidencial e valiosa pertencente a uma empresa, indústria ou até mesmo uma organização governamental [6] [7]. Com o passar do tempo e também com a aparição da Stuxnet [8], este tipo de ataques tem-se tornado cada vez mais destrutivo e popular entre as organizações com fins maliciosos, já que representa características e métodos sofisticados que permitem o seu sigilo perante as defesas tradicionais no meio digital. Após a deteção destes ciberataques (se for o caso) por parte de organizações de cibersegurança, os mesmos são estudados e novamente desenvolvidos pelos cibercriminais, mas com modificações e alterações com a intenção de não repetir padrões de ataques passados que possam estar identificados, passando, assim, desapercebidos a possíveis medidas tomadas pelas entidades de segurança.

CARACTERÍSTICAS DOS APT'S E DIFERENÇA ENTRE CIBER-ATAQUES TRADICIONAIS

Em 2006, analistas da Força Aérea dos Estados Unidos (USAF) usaram o termo APT de forma a facilitar a discussão em relação a ataques de intrusão no ciberespaço [7]. Desta forma, poder-se-ia discutir as características destes ataques sem a necessidade de revelar entidades classificadas. Os componentes definidos pela USAF em relação a este tipo de ataques são descritos da seguinte forma:

- **Advanced**: os atacantes estão familiarizados com o tipo de métodos e ferramentas usadas para realizar este tipo de ataques, o que lhes permite ter um conhecimento abrangente sobre as mesmas. Alguns destes métodos avançados como engenharia social, vulnerabilidades "Zero-Day", comunicação disfarçada e Machine Learning são usados para alcançar os seus objetivos finais. A composição dos mesmos engloba mais de um tipo de ataque usados em conjunto, o que determina também a sua capacidade "avançada".

- **Persistent**: o atacante procura um controlo a longo prazo sobre a vulnerabilidade encontrada ou até mesmo o próprio sistema, com o objetivo de passar desapercebido até que o objetivo final (ataque) seja alcançado, o que poderá levar anos até ser concretizado.

- **Threat**: o ataque é coordenado, suportado e motivado pela sua finalidade. O tipo de ameaça é considerável já que estes ataques direcionam-se em grande parte a estruturas e empresas de grande valor que possam conter informações confidenciais e valiosas. Uma vez que o sistema se encontra comprometido, os danos podem ser bastante incomodativos a nível económico.

Os autores deste tipo de ciberataques têm propósitos e finalidades que se diferenciam dos ataques vulgares no ciberespaço pela sua natureza e composição. Por exemplo, a espionagem em diferentes setores como a indústria, economia, setor financeiro e entidades governamentais, são praticadas em grande medida nos dias de hoje pela intrusão nos Sistema de Informação deste tipo de atacantes com um tipo de ferramentas e métodos usados nos APT. Segundo [6] são apresentadas algumas diferenças entre um ataque convencional e um APT, tendo em conta os seguintes elementos: atacante, alvo, finalidade e método.

Table 1. Differences between an advanced persistent threat (APT) attack and common malware attacks [2].

Feature	APT Attacks	Common Malware Attacks
Definition	APT is a sophisticated, targeted and highly organised attacks. (e.g., Stuxnet)	Malware is malicious software used to attack and disable any system. (e.g., ransomware)
Attacker	Government actors and organised criminal groups	A cracker (a hacker in illegal activities)
Target	Diplomatic organisations, information technology industry and others sectors	Any personal or business computer
Purpose	Filter confidential data or cause damage to a specific target	Personal recognition
Attack life cycle	Maintain persistence as possible using different ways	It ends when it is detected by the security actions (e.g., anti-virus software)

PROCESSOS DE ATAQUE E MÉTODOS

O **ciclo de vida/estrutura** deste tipo de ataques é fundamental para entender a funcionalidade e os métodos maliciosos utilizados nos mesmos; existem inúmeras formas de um ataque APT passar desapercebido pelas etapas que o compõem de forma a não levantar suspeitas contra os sistemas de reconhecimento de anomalias na rede.

Recentemente, investigadores no ramo têm proposto a análise aos "ciclos de vida" que se têm determinado em certas campanhas de ataques APT, divididos em etapas, sendo estas compostas por técnicas, métodos e ferramentas apropriadas para proceder à intrusão no sistema de forma sigilosa. A quantidade de etapas define a anatomia do ataque APT e esta pode diferir, consoante o proposto no estudo [3] [9] entre **3 e 11 etapas**. Isto é devido às diferentes finalidades e objetivos que estas organizações possam ter.

No entanto, pode-se estabelecer uma estrutura primária composta por **7 etapas**, pela qual se baseiam estes ataques, útil para os atacantes e também para as entidades de cibersegurança no que toca à análise e prevenção dos mesmos: o "***Cyber Kill Chain***".

O conceito de "*kill chain*" não foi utilizado inicialmente no ciberespaço, mas sim no meio militar. É um modelo composto por **6 etapas** que descrevem todo o processo de um ataque militar: "*discover, locate, track, target, strike, achieve*" [1]. Quanto antes se procede à deteção do ataque nestas etapas, menor a probabilidade de o mesmo causar danos significativos na organização. Lockheed Martin propôs também uma sequência de etapas semelhante ao "kill chain", denominado **Cyber-Kill-Chain**. Dentro do meio cibernético, este ataque divide-se em **7 etapas** representadas na figura abaixo [2]. Esta teoria pode ser útil não só para a estruturação de um ciberataque, mas também para a sua deteção.

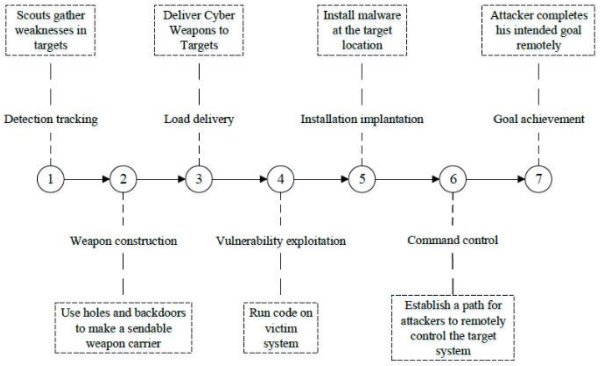

Figure 1. Cyber-kill-chain.

As 7 etapas representadas na figura são as seguintes:

- **Detection tracking**
- **Weapon construction**
- **Load delivery**
- **Vulnerability explotation**
- **Installation implantation**
- **Command control**
- **Goal achievement**

Os APTs usam uma grande variedade de métodos e técnicas para conseguir cumprir com o seu objetivo principal. O processo de ataque inicia-se com o estudo de possíveis vulnerabilidades no sistema da vítima, sendo utilizados o **spear-phishing** ou emails com ficheiros mailicosos juntamente com **engenharia social** de forma a convencer a vítima da sua veracidade dentro do meio em que se encontra e realizar o download do ficheiro infetado para começar a primeira etapa do ataque APT. A partir desse momento, o malware iniciaria a infeção de outros computadores pela rede de forma sigilosa.

Os ataques APT mais avançados são aqueles estão estruturados em volta de ferramentas e métodos nunca antes analisados, como o uso dos ***Zero-Day exploits*** ou uso de vetores infetados não identificados. Entre as vítimas deste tipo de ataques podem-se destacar uma série de órgãos governamentais de certos países com prejuízos milionários.

Algumas das **técnicas** combinadas e utilizadas num ataque APT (nas suas correspondentes fases) são [10]:

- **Engenharia Social**
- **Spear-phishing**
- **Watering hole**
- **Drive-by-Download**

EXEMPLOS DE ATAQUES APT

Os ciberataques realizados por organizações governamentais e nações unidas têm-se tornado cada vez mais frequentes nos dias de hoje. Estes autores são determinados e associados a estas ameaças após a análise forense intensiva por parte de expertos na área de cibersegurança, tendo conseguido investigar os respetivos endereços IP, emails com conteúdo malicioso ou até mesmo o próprio código destes programas. Consoante o tipo de organização, podem-se dividir em dois grupos: **autores governamentais e grupos criminais**.

Por outro lado, cabe destacar o exercício que os mesmos praticam, sendo ataques caracterizados pelas ações, ferramentas utilizadas e métodos aplicados para executar os APTs contra as vítimas e, assim, ter a capacidade de extrair e comprometer dados sensíveis e valiosos com certa finalidade. Alguns destes ataques identificados nos últimos anos estão determinados na tabela abaixo.

Table 1. APT attacks in recent years.

Organization	Target	Tools
Hades	Korea	Olympic Destroyer
APT28(Suspect)	Ukraine	VPNFilter
APT28	North America,Europe	Cannon, Zebrocy
BlueMushroom	China	PowerShell backdoor
OceanLotus	Southeast Asian, China	Denis, Cobalt Strike
BITTER	China, Pakistan	Unique backdoor procedures
APT38	Global, SWIFT	Multiple homemade malicious programs
DarkHotel	China	Plug-in Trojan Backdoor
APT33(Suspect)	Middle East, Europe	Shamoon V3

MACHINE LEARNING

O **Machine Learning** (ML) é um sub-ramo da ciência da Inteligência Artificial (AI) capaz de gerar através de poder computacional modelos de aprendizagem automática a partir de uma coleção de dados. O Machine Learning caracteriza-se pelo estudo de técnicas e algoritmos com o objetivo de automatizar tarefas e soluções que são consideradas complexas para a programação convencional. Graças ao uso de modelos matemáticos e funções estatísticas é possível determinar dependências entre os dados analisados, correlações ou até padrões que se repitam ao longo do tempo.

A finalidade do ML é ajudar o ser humano a tomar certas decisões e a resolver certos problemas, seja em que área científica for. Algumas das aplicações que o Machine Learning pode ter são: deteção de sistemas fraudulentos, reconhecimento facial, reconhecimento por voz, análise de padrões repetitivos ao longo do tempo, deteção de anomalias em sistemas, etc.

ALGORITMOS E TÉCNICAS EXISTENTES NO ML

Exitem diversos algoritmos e técnicas que serão apresentados neste estudo com o intuito de demonstrar a caracterização de cada um e as suas finalidades. Para isso, cabe destacar a diferença entre os dados rotulados e os dados não rotulados. No primeiro caso, os dados estão referenciados a través de classes ou "*tags*" e, no segundo, não existe qualquer tipo de identificação ou referencia sobre os dados. Quando a resposta correta a

uma pergunta relacionada aos dados é conhecida, são obtidos dados rotulados; no entanto, quando a resposta correta é desconhecida, obtêm-se dados não rotulados.

Os algoritmos de Machine Learning tem o poder e a capacidade de poder aprender de forma automática com os dados disponíveis e as decisões tomadas anteriormente sobre os mesmos. Assim, os **modelos de aprendizagem podem ser classificados em dois grandes grupos: modelos de aprendizagem supervisionada e não supervisionada** (Figura).

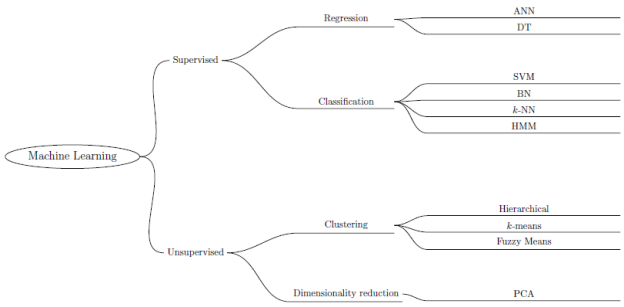

Figure 1. Machine learning algorithms.

O objetivo da **aprendizagem supervisionada** é o de desenvolver um modelo que crie previsões com base em evidências à priori na presença da incerteza. Tendo em conta um conjunto de dados iniciais (inputs) e as suas respetivas decisões/respostas (outputs), este tipo de algoritmo tem como objetivo treinar o modelo ML e gerar previsões analíticas em base do estudo estatístico e matemático como resposta a novos dados que lhes sejam introduzidos para análise. Uma vez treinado e adaptado a um grande conjunto de dados, estes modelos são capazes de determinar percentualmente um output final como resposta.

Técnicas de regressão e de classificação estatística são usadas no desenvolvimento destes modelos. Os **métodos de ML supervisionados baseados em regressão** são:

- **Artificial Neural Networks** (ANN): modelo computacional inspirado na composição e atividade do cérebro humano e composto por interconectores ("*artificial synapses*") de neurónios artificiais, denominados Nodes. Estes são capazes de realizar certos cálculos nos dados introduzidos (inputs) [11]. Caracterizados pela adaptação própria a experiências, capacidade de aprendizagem, organização de dados, tolerância a falhas, armazenamento distribuído, etc., estes têm a capacidade de criar modelos não-lineares com o objetivo de obter certas relações entre os inputs introduzidos e a classificação de rótulos [12].

- **Decision Tree** (DT): este tipo de modelos é eficaz, estável e de fácil interpretação. Representados em arvores de decisão, são capazes de apresentar relações não lineares para a resolução de certos problemas. Têm uma grande precisão e são bastante elaborados, o que os faz serem dos modelos mais importantes nesta área. Por um lado, e comparado com os outros modelos, estes são capazes de criar previsões mais precisas, mas por outro têm uma performance mais limitada. O CART, CHAID ou ID3 são alguns dos mais conhecidos [13] [14] [15].

De seguida, apresentam-se os modelos de ML supervisionados em classificação:

- **Support Vector Machine** (SVM): caracterizado pela identificação dum Hiper plano entre duas classes de dados rotulados num conjunto de dados treinados. Usa diversos tipos de métodos (não lineares, kernels, separação e minimização de margens ou riscos) e permite a transformação de um problema não linear num linear, o que facilita a obtenção final de resultados [14] [13] [16].

- **Bayesian Networks** (BN): modelos gráficos probabilísticos usados para descrever e analisar distribuições de variáveis. Estas variáveis podem ser discretas ou continuas, no entanto, quando todas são discretas, a notação é representada como séries de produtos e somas. Na sua representação gráfica, os Nós simbolizam estados ou variáveis observáveis, e os bordes, dependências condicionais entre os nós [13] [17].

- **k-nearest neighbour** (k-NN): modelo usado tanto para problemas de classificação como para problemas de regressão. Pode ser usado para identificar o "vizinho" mais próximo num conjunto de dados baseado na medição da sua distância, que pode ser definido como

uma distância Euclidiana entre dois pontos. A finalidade é que elementos similares estejam mais próximos um do outro. Neste caso, a decisão da classificação pode vir influenciada pela sensibilidade de "k", especialmente em conjuntos pequenos de dados de valores atípicos. Existem outros tipos de métodos para a medição entre dois pontos como Minkowski, produto interno, acorde quadrado, entropia de Shannon, e Vicissitude [18].

- **Modelo Hidden Markov** (HMM): modelo probabilístico estocástico de eventos discretos e uma variação da corrente Markov, uma corrente de estados/eventos interligados, onde o próximo estado depende unicamente do estado atual do sistema. Este modelo é usado para analisar características ou observações com o objetivo de prever o próximo estado da sequência [13] [19].

Por outro lado, a **aprendizagem não supervisionada** não contém um conjunto de dados treinados. Alguns dados não rotulados são apresentados e o modelo, por si só, deve aprender com os mesmos para prever resultados à posteriori. Este é o tipo de modelo mais apropriado quando o conjunto de dados apresentado é maior e os mesmos não se encontram rotulados de forma alguma. Tem como finalidade encontrar padrões não identificados ou estruturas de dados que ainda não tenham sido determinadas no seu conjunto. Este modelo usa os seguintes algoritmos:

- **Análise do Componente Principal** (PCA): procedimento de redução de dimensão. Este método estatístico é apropriado quando existem grandes quantidades de variáveis, onde cada uma delas tem maior ou menor importância. O PCA gera uma "*matriz T*", onde a correlação entre variáveis é apresentada em duas ou três dimensões, como máximo. Este procedimento é realizado com o objetivo de obter um conjunto menor de variáveis correlacionadas de forma não linear [20].

- **K-means** é um algoritmo de cluster. Esta técnica caracteriza-se por selecionar dados de input em "k clusters" para grupos "k" predefinidos. Cada dado no conjunto de inputs é um dado não rotulado. A interpretação para cada grupo "k" é que o valor médio de cada um deles representa os dados do grupo no seu conjunto. Alternativamente, cada grupo "k" pode representar um tipo de dado nos inputs. Este algoritmo usufrui de medições computacionais para encontrar a distância entre dois pontos (distancia. Euclidiana). Pode ser aplicado em Sistemas de Deteção de Intrusão (IDS) [21].

- **Fuzzy c-means**: Algoritmo de cluster caracterizado por escolher de forma aleatória o número dos clusters e, de seguida, a cada dado é-lhe associado uma filiação de cluster. Tem como objetivo primário a redução da distância e o grau da filiação dos clusters [22].

- **Cluster hierárquico**: usado para agrupar dados não rotulados. Podendo ser de dois tipos, *divisivo e aglomerativo*; no primeiro tipo, os "*data points*" são reconhecidos com um só cluster de dados e, à posteriori, dividido em clusters menores; no segundo caso, cada "*data point*" é considerado um elemento individual e, só depois, são adicionados a um cluster [23].

MACHINE LEARNING APLICADO À CIBER SEGURANÇA NA DETEÇÃO DE APTs

Devido à existência de um incremento de ciberataques no meio cibernético, maior é a necessidade da aplicação e implementação de novos métodos e ferramentas para evitar, prever e deter este tipo de ameaças. Alguns destes métodos envolvem a aplicação de ML para a proteção no ciberespaço, mas por outro lado, este tipo de ataques também usufrui deste tipo de aplicações e métodos para obter o maior sigilo possível na implementação e desenvolvimento do ataque sobre organizações.

As medidas de prevenção requerem uma maior capacidade computacional para a análise e tomada de decisão no menor espaço temporal possível, devido à grande quantidade de dados digitais e à evolução de ameaças neste âmbito. As técnicas de ML são uma ferramenta útil no âmbito da cibersegurança, como por exemplo, a criação de modelos de deteção no tráfego interno da rede para detetar anomalias na sua

atividade, a redução de falsos positivos nos alarmes, a deteção de ameaças em tempo real, etc. [24] A classificação destes métodos/ferramentas na cibersegurança pode ser determinada e subdividida da seguinte forma [25]:

- **Deteção**: ferramentas apropriadas para a deteção anormal de certos acontecimentos numa rede. Desta forma criam-se alarmes e processos automatizados de decisão (AI);

- **Proteção**: deteção de vulnerabilidades no sistema de forma a instalar medidas de segurança automaticamente;

- **Predição**: técnicas e algoritmos que prevejam ataques e desenvolvam técnicas contra potencias malwares;

- **Eliminação**: eliminação automática de ameaças.

O Machine Learning pode complementar a atividade humana no setor da cibersegurança de forma a evitar o comprometimento de sistemas de redes contra ciberataques como é o caso dos APTs. Alguns pontos essenciais para a deteção de APTs são:

- **Observação de padrões de alerta** anormais para detetar malwares com intenções maliciosas (execução de código malicioso, controlo remoto não autorizado, etc.);

- **Monitorização do "outbound traffic"** na rede, podendo representar parâmetros significantes como computadores infetados, intervenção de centros C&C e exfiltração de dados;

- **Monitorização de tráfego na rede** inesperado pode revelar o escalamento de privilégios, movimentação lateral e propagação de malware.

Algumas das aplicações que o Machine Learning tem no âmbito da cibersegurança são:

- **Deteção de Phishing e Spam**, sendo um dos vetores mais usados e explorados pelos atacantes, caracteriza-se pelo recebimento de emails não solicitados que possam conter conteúdo malicioso e estabelecer um ponto de entrada e ligação no sistema para o atacante. Diversas técnicas de Machine Learning podem ser aplicadas neste âmbito com o objetivo de identificar emails fraudulentos e compará-los com os emails acessíveis. A classificação entre os dois tipos de email é necessária para definir os critérios que os distinguem permitindo que o algoritmo aprenda a identificá-los [26] [27].

- **Deteção de Malware**, sendo um problema ao tratar-se de ficheiros executáveis que permitem a propagação do mesmo pela rede ou o dano no sistema comprometendo toda a informação que nele se encontra. Normalmente, o malware usa uma comunicação com um servidor C&C a través de um endereço IP ou URL aleatórios, o que dificulta a sua deteção. Com o uso do ML, é possível detetar padrões de comunicação que este tipo de software estabelece através dos "web browsers" ou até mesmo através da análise do tráfego malicioso no DNS [28].

- **Deteção de Intrusão**, permitindo desta forma o monitoramento do tráfego da rede para analisar comportamentos anormais na mesma e detetar padrões. Podem ser usadas também assinaturas (*hash*) que identifiquem um possível ataque [29]. No estudo [30] é proposto também a deteção de movimentos laterias em sessões de Windows de conexões remotas maliciosas (RDP) não permitidas, através do estudo e análise dos *Event Logs*.

Desenvolvimento

De seguida, são propostas 5 abordagens de possíveis soluções na aplicação do Machine Learning na deteção de APTs. Vários estudos foram analisados para a descrição das mesmas com o objetivo final de demonstrar a eficácia de cada uma das abordagens.

DETEÇÃO DE APTs - ML DE CLASSIFICAÇÃO FUNDAMENTADO EM DIMENSÕES FRACTAIS

Segundo o estudo [31], classificam-se os padrões anormais do tráfego fundamentados em APTs a través do processamento de vetores obtidos pela informação das sessões de TCP/IP com o objetivo de obter a maior precisão e veracidade possível, seguido da proposta da criação de um novo algoritmo de correlação de dimensão fractal. Os resultados experimentais, comparados com o *k-NN*, apresentam uma **melhoria de 8%** no que toca à precisão, sensibilidade e veracidade. Por outro lado, demonstra-se também uma **melhoria de 12%** em relação à *medição-F*, o que indica uma redução nos falsos positivos por parte do método fractal, dando lugar a uma melhor performance. Mesmo quando o conjunto de dados é desequilibrado, a performance do algoritmo continua a ser melhor que o k-NN.

DETEÇÃO DE APTs – GAN-LSTM

Este modelo de Machine Learning visa responder ao problema dos detetores tradicionais de ataques APT. Enquanto um detetor tradicional baseia-se apenas na comparação de um só tipo de ataque (o que diminui a probabilidade de detetar ataques APT que possam ter variações em comparação com ataques passados), neste modelo [32] propõe-se um método de deteção para ataques APT no qual se combinam o *Generative Advesarial Networks* (GAN) e o *Long Short-term Memory* (LSTM). Este tipo de deteção determina-se pelo módulo de geração de dados de um ataque e o próprio ataque em si.

Este modelo caracteriza-se pela sua composição e processos para o resultado final da deteção de APTs. Em primeiro lugar, é gerada uma grande quantidade de dados de possíveis ataques na simulação GAN para posteriormente serem analisados de forma a aumentar a precisão do modelo. Por outro lado, a unidade de memoria e a estrutura baseada no modelo LSTM garantem a correlação e o tempo na sequência do ataque APT e a memoria de recursos entre os fragmentos de larga sequência resolve o problema de intervalos de tempo compridos.

A geração de dados de ataques simulados através do modelo para otimizar o modelo discriminativo aumenta a precisão do modelo original em **2.84%.** Comparado também com o modelo *Recurrent Neural Network* (RNN), este tem uma maior precisão em **0.99%** valores percentuais.

Os resultados experimentais deste estudo demonstram que os algoritmos baseados no GAN-LSTM podem realmente aumentar a probabilidade de deteção de APTs com a introdução de modelos generativos, aumentar a sua precisão e diminuir os falsos positivos.

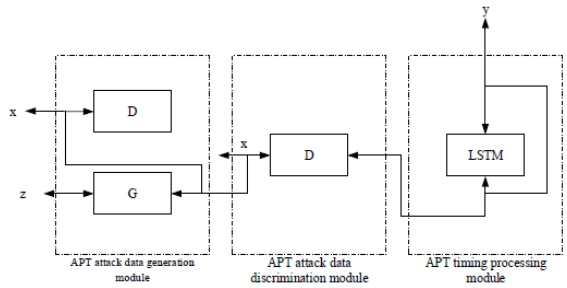

Figure 2. The basic structure of the model.

DETEÇÃO DE APTs – ANÁLISE DE CORRELAÇÃO DE ML

Neste modelo [33] propõe-se um novo sistema MLAPT capaz de detetar de forma precisa e rápida este tipo de ataques. Está composto pelas seguintes etapas:

1. **Deteção da ameaça**: oito métodos são desenvolvidos para detetar diferentes tecnologias que possam ser usadas pelos ataques APT;

2. **Aviso de correlação prévia**: definição dum framework de correlação, a ligação dos outputs de métodos de deteção e a identificação de avisos prévios que possam ser relevantes no cenário dum APT;

3. **Predição do ataque**: objetivo de procura dum ataque APT completo em base da probabilidade determinada pelos avisos prévios e do resultado dos outputs do framework de correlação.

Desta forma, comprova-se que com este sistema um APT pode ser previsto com uma **precisão de 84.8%.**

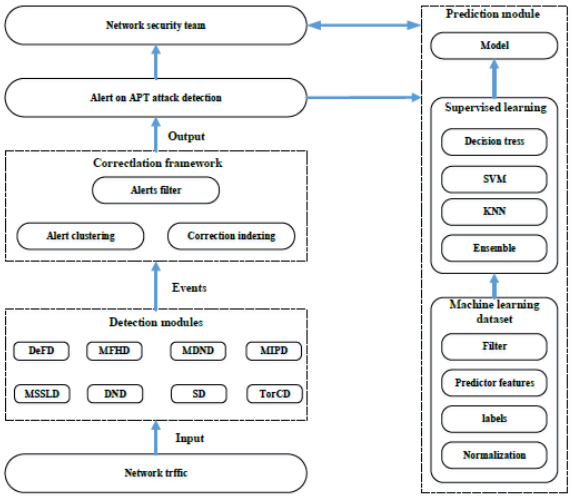

Figure 3. The framework of the system.

DETEÇÃO DE APTs – DEEP LEARNING STACK

Segundo [34] é descrito um método composto por blocos e processos de Deep Learning baseados numa abordagem teórica. Neste modelo, o ataque APT é considerado um ataque multi-media dividido em etapas de forma estratégica. Para a deteção destes ataques, toda o fluxo da rede deve ser usado como input para o processo de deteção com o objetivo principal de decompor o "grande problema" em tarefas menores e, assim, poder analisá-las para obter resultados conclusivos sobre as mesmas.

DETEÇÃO DE APTs – APRENDIZAGEM SEMI-SUPERVISIONADA E CARACTERÍSTICA DE REDE COMPLEXA

Os APTs são ciberataques complexos compostos por várias etapas cujo objetivo pode diferir consoante o autor e a intenção final. Neste tipo de ataques são usadas técnicas e métodos sofisticados que também podem usufruir de tecnologias como o Machine Learning, o que requer uma maior atenção por parte da cibersegurança no meio cibernético. Este estudo [35] visa responder às necessidades de proteção dentro de uma rede e às limitações existentes contra este tipo de

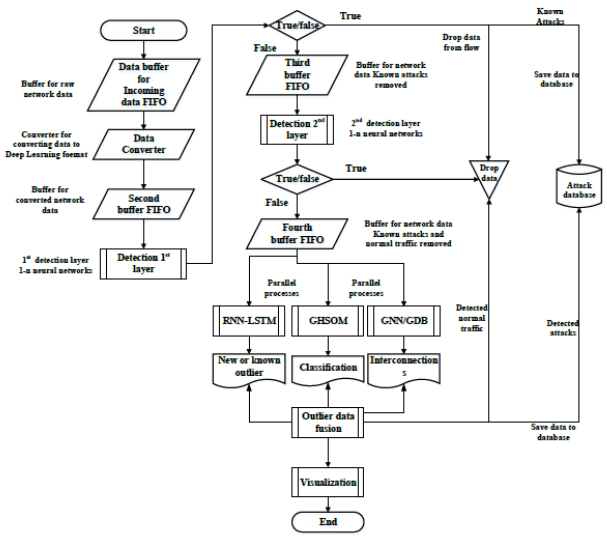

Figure 4. Data flow through the detection process.

ataques.

O modelo fundamenta-se em métodos *semi-supervisionados* e *características de redes complexas*. A rede alvo é modelada como uma pequena rede enquanto a *APT-Attack Network* (APT-AN) é estruturada de norma não escalável. O estado finito da máquina é usado para definir o estado das transições de cada Nó no domínio do tempo para caracterizar a mudança dos estados durante um ataque APT. O modelo proposto tem a capacidade de analisar um grande conjunto de dados para definir as características dum ataque APT entre o centro de comando e host da vítima.

Os resultados demonstram uma **precisão de 90.5%,** verificando-se a capacidade de detetar anomalias correspondes às diferentes etapas dum ataque APT.

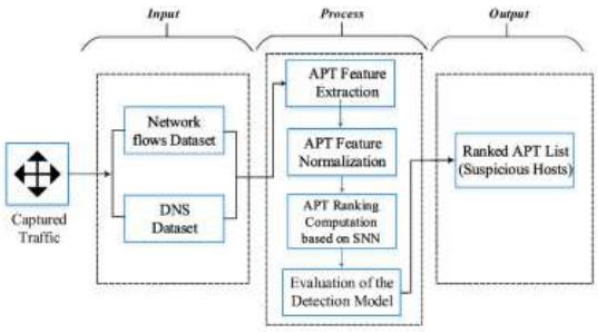

Figure 5. The framework of the system.

Discussão

Neste artigo fez-se referência a uma série de estudos e investigações que visam contribuir para o desenvolvimento de métodos de Machine Learning dentro do âmbito da cibersegurança. Para a sua concretização, pode-se determinar os algoritmos implementados em cada um deles, que pretendem **detetar, analisar e mitigar** qualquer tipo de sinal que seja resultado (output) de cada um dos modelos e de todos os processos que os envolvem. Cada um deles apresenta objetivos concretos com a finalidade de apresentar vantagens sobre os métodos tradicionais de deteção dos ataques APT, seja através de experimentos, indicadores, previsões e precisões sobre a deteção e a identificação de falsos positivos. Desta forma, podem-se apresentar na figura a baixo as vantagens e desvantagens de cada método aplicado.

Table 2. Disadvantages of mentioned methods.

Methods/Ideas	Shortage
Classification based on fractal dimensions	Hard to deal with malicious data classification
GAN-LSTM	Unable to deal with long sequence
Machine Learning Correlation Analysis	Lack of publicly available datasets
A new Deep Learning stack	Uncomputed time complexity
Semi-supervised Learning and complex network feature	Computational overhead

Podendo verificar-se as vantagens e as desvantagens de cada um dos métodos de ML, pode-se deduzir (consoante o valor percentual referente à precisão do mesmo) que o modelo de "Aprendizagem semi-supervisionada e característica de rede complexa" apresenta uma maior eficácia, o que é representado por 90.5% valores percentuais referentes à sua precisão na deteção de APTs.

Reflexão e aprendizagem

A abordagem à tecnologia do Machine Learning aplicada na deteção dos ciberataques APTs representa toda a evolução tecnológica no âmbito da cibersegurança. Desta forma e ao longo deste artigo, representaram-se vários métodos e algoritmos capazes de detetar, analisar e mitigar possíveis anomalias dentro das redes, que até aos dias de hoje se encontram em constante evolução. A definição dum ataque APT, a apresentação e descrição de cada algoritmo/método de ML (aplicado ao ramo da cibersegurança e não só) e a aplicação que cada um deles pode ter neste âmbito foram de suma importância para o entendimento geral do tema proposto.

Ao longo de todo o artigo fui aprendendo várias metodologias do ML e vários algoritmos que desconhecia, tendo em conta a complexidade de cada um deles e de que forma podem ajudar a detetar anomalias numa rede. Graças à investigação realizada, pude conhecer os perigos reais que um APT apresenta e os processos/etapas que o compõem de forma a alcançar os seus objetivos finais; a estrutura de diversos modelos de ML e os algoritmos implementados; formas de abordagem para a deteção de APT, etc.

Assim, é de destacar as aportações e o contributo de que cada método de ML, tendo em conta as suas vantagens e desvantagens. com o objetivo de aumentar a segurança no meio cibernético.

Conclusão

Este artigo visa demonstrar ferramentas e modelos de Machine Learning na deteção de APTs que se desenvolveram ao longo dos anos como forma de contributo para o âmbito da cibersegurança. Alguns dos métodos explicados neste artigo encontram-se em desenvolvimento e em constante evolução, o que demonstra também o acompanhamento por parte da cibersegurança no estudo de qualquer tipo de ameaça que possa por em perigo organizações ou até mesmo indivíduos no meio cibernético. Propõe-se, assim, a continuação desta investigação de forma a demonstrar novos métodos de ML aplicados à deteção, prevenção e/ou mitigação de APTs.

Referencias

[1] Fireeye, “Fireeye Mandiant Services Special Report,” Fireeye, Milpitas, CA, 2019.

[2] J. C. F. M. J. M. F. Antoine Lemay, “A survey of publicly available reports on advanced persistent threat actors,” em *Computers & Security*, 2018, pp. 26-59.

[3] Swisscom, “Targeted Attacks Cyber Security Report,” Ltd. Group Security, Bern, Switzerland, 2019.

[4] B. T., B. H., D. A.A, S. M.A., L. N. e B. R., “A Machine Learning Aproach for RDP-based Lateral Movement Detection,” em *IEEE 44th Conference Local Computer Networks*, Osnabueck, Germany.

[5] Y. H. J. L. F. W. Ru Zhang, “Constructing APT Attack Scenarios Based on Intrusion Kill Chain and Fuzzy Clustering,” *Hidawi*, pp. 1-9, december 2017.

[6] D. L. H. C. Chen P., “Lecture Notes in Computer Science,” *A Study in Advance Threats, Springer*, pp. 63-72, 2014.

[7] L. Y. W. D. Jeun I., A Practical Study on Advanced Persistent Threats, 2012.

[8] M. L. C. E. Falliere N., White Pap., 2011.

[9] J. D. C. F. M. C. Ussath M., “Advanced Persistent Threats: Behind the scenes,” em *Annual Conference on Information Science and Systems (CISS)*, Princeton, NJ, USA, 2016.

[10] H. H. M. H. E. W. Katharina Krombholz, “Advanced social engineering attacks,” *SicenceDirect*, pp. 113-122, 2015.

[11] I. S. Sara Kaviani, “Influence of random topology in artificial neural networks: A survey,” *ScienceDirect*, 2020.

[12] D. S. R. F. L. L. S. A. Ivan Silva, “Artificial Neural Networks,” *Springer*, pp. 1-307, 2017.

[13] A. Joshi, “Machine Learning and Artificial Intelligence,” *Springer*, pp. 49A-60A, 2020.

[14] C. J. L. K. N. C. Wen Lin Chu, “Detection and Classification of Advanced Persistent Threats and Attacks Using the Support Vector Machine,” *MDPI*, 2019.

[15] D. A.-J. A. H. J. M. T. B. A. A. Mohamed Alloghani, “Implementation of Machine Learning and Data Mining to Improve Cybersecurity and Limit Vulnerabilities to Cyber Attacks,” *Springer*, pp. 47-76, 2019.

[16] C. C. P. N. Javier Torres, “Review: machine learning techniques applied to cybersecurity,” *Springer*, pp. 2823-2836, 2019.

[17] Z. A. Cleophas T.J., “Modern Bayesian Statistics in Clinical Research,” *Springer*, 2018.

[18] A. H. O. L. A. T. M. A. H. S. S. P. Haneen Alfeilat, “Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review,” *Liebertpub*, pp. 221-248, december 2019.

[19] R. K. Mariette Awad, “Hidden Markov Model,” *Springer*, pp. 81-104, 2015.

[20] A. Olivieri, “Principal Component Analysis,” *Springer*, pp. 57-71, 2018.

[21] I. G. P. V. J. K. Ansam Khraisat, “Survey of intrusion detection systems: techniques, datasets and challenges,” *Springer*, p. 20, 2019.

[22] M. D. Lihua Yang, “Based on k-Means and Fuzzy k-Means Algorithm Classification of Precipitation,” *IEEE*, pp. 218-221, 2010.

[23] A. C. S. G. P. A. S. K. Ravinder Ahuja, “Classification and Clustering Algorithms of Machine Learning with their applications,” *Springer*, pp. 225-248, 2019.

[24] L. B. T. S. J. L. Zhenyu Guan, “When Machine Learning meets Security Issues: A survey,” *IEEE*, pp. 158-165, 2018.

[25] S. V. Soumendra Mohanty, “Cybersecurity and AI,” *Springer*, pp. 143-153, 2018.

[26] OWASP, "OWASP cheat sheet series," [Online]. Available: https://cheatsheetseries.owasp.org/cheatsheets/Unvalidated_Redirects_and_Forwards_Cheat_Sheet.html. [Acedido em 2019].

[27] P. Paganini, "Security Affairs," Security Affairs, [Online]. Available: https://securityaffairs.co/wordpress/91877/cyber-crime/adobe-google-open-redirects.html. [Acedido em 2019].

[28] K. L. B. G. Zhao, "Detecting APT Malware Infection Based on Malicioues DNS and Traffic Analysis," *IEEE,* pp. 1132-1142.

[29] E. G. Anna Buczak, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE,* pp. 1153-1176, 2016.

[30] B. H. D. A. S. M. L. N. B. R. Bai T., "A machine Learning Approach for RDP-based Lateral Movement Detection," em *44th Conference Local Computer Networks*.

[31] M. S. K. F. W. K. Sana Siddiqui, "Detecting Advanced Persistent Threats using Fractal Dimension based Machine Learning Classification," em *ACM on INternational Workshop on Security and Privacy Analytics*, 2016.

[32] W. T. b. S. J. S. C. T. Liu Hai bo, "Advanced ersisten Threat Detection Based on Generative Adversarial Networks and Long Short-term Memory".*Computer Science.*

[33] H. M. P. V. Ghafir I., "Detection of advanced persistent threat using machine-learning correlation analysis," *Future Generation Computer Science,* pp. 349-359.

[34] T. H. Tero Bodström, "A Novel Deep Leraning Stack for APT Detection," 2019.

[35] H. C. Z. W. M. C. Aaron Zimba, "Modeling and detection of multi-stages of Advanced Persisten Threats attacks based on semi-supervised learning and complex networks characteristics," *Future Generation Computer Systems,* 2020.