Machine learning **como apoio à ataques de exfiltração de dados:** uma arquitetura integrada para diminuição de riscos de detecção

1º Ten Eng Isabelle Cecilia de Andrade
1º Ten Guilherme Resende Deus

RESUMO

O presente artigo possui como objetivo propor uma arquitetura integrada para a otimização de ações de exfiltração de dados, utilizando fontes de dados abertas e técnicas de *machine learning*. O fluxo de aplicação prevê a detecção de artefatos de maior interesse para o agente executor por meio de seleção de alvos e consequente redução do volume de dados envolvidos na ação de exfiltração de dados, permitindo assim acelerar o ritmo e reduzir o risco de detecção do ataque. A arquitetura proposta é composta por três componentes: um *crawler* que utiliza motores de busca conhecidos na *web* para a pesquisa e coleta de arquivos no formato PDF; um componente de modelagem de tópico que classifica os arquivos coletados; e um componente de aprendizado de máquina que utiliza os documentos classificados para treinar um algoritmo a identificar documentos semelhantes. Com a implementação de uma prova de conceito, este artigo demonstra ser possível atingir os objetivos pretendidos, apresentando como resultado uma redução de 90% do volume de dados envolvidos em uma ação de exfiltração de dados com a arquitetura proposta, reduzindo o tempo de execução e os riscos de detecção da ação.

Palavras-chave: Exfiltração de dados; Aprendizado de máquina; Defesa cibernética ativa.

Machine learning as a support for data exfiltration attacks: an integrated architecture for reducing detection risks

ABSTRACT

This article aims to propose an integrated architecture for optimizing data exfiltration actions using open data sources and machine learning techniques. The application flow provides for the detection of artifacts of greater interest to the executing agent through target selection and the consequent reduction in the volume of data involved in the data exfiltration action, thus allowing to speed up the pace and reduce the risk of attack detection. The proposed architecture is composed of three components: a crawler that uses search engines known on the web to search and collect files in PDF format; a topic modeling component that classifies collected files; and a machine learning component that uses classified documents to train an algorithm to identify similar documents. With the implementation of a proof of concept, this article demonstrates that it is possible to achieve the intended objectives, resulting in a 90% reduction in the volume of data involved in a data exfiltration action with the proposed architecture, reducing the execution time and the risks of detecting the action.

Keywords: Data exfiltration; Machine learning; Active cyber defense.

1 INTRODUÇÃO

O reconhecido cenário de inundação de dados nas redes evidencia-se cada vez mais pela crescente quantidade de dados trafegados: só em 2020, foram mais de 51 milhões de exabytes por mês (O'DEA, 2020) em quase 6 bilhões de páginas *web* indexadas no total (KUNDER, 2021). Nesse cenário, e em consonância com a notável definição de que "dados são o novo petróleo" (JOSSEN, 2017, p. 1), encontrar dados efetivamente úteis pode ser quase tão complexo quanto literalmente encontrar um novo poço de petróleo.

Diversos desafios são consequência dessa disponibilidade massiva de dados, porém o vazamento de dados sensíveis vem sendo considerado nos últimos anos a ameaça mais temida por organizações (BIG, 2021), principalmente por impactar tanto as áreas de inteligência de negócio quanto a área de *compliance* (RAPÔSO, 2019).

No âmbito da segurança nacional, vazamentos de dados possuem um poder de destruição ainda maior. Em um episódio recente, o exército americano admitiu ter milhares de dados de transporte de militares expostos na *web* (MUNCASTER, 2019) ocasionando impactos incalculáveis no controle tático desse tipo de informação.

Visando mitigar situações similares, a Política Nacional de Segurança da Informação (BRASIL, 2018, p. 1-2) define como princípios o respeito "à proteção de dados pessoais, à proteção da privacidade e o acesso à informação" e a "articulação entre as ações de segurança cibernética, de defesa cibernética e de proteção de dados e ativos da informação".

No contexto específico da área de defesa cibernética, a exploração de ações de defesa ativa permitem, proativamente, detectar, analisar, e mitigar falhas de segurança em tempo real com o uso de contramedidas agressivas implantadas fora de uma rede alvo (DEWAR, 2014), aumentando assim o nível de consciência situacional das organizações.

Diversas técnicas de defesa cibernética ativa são utilizadas hoje no contexto integrado de geração de informação e de investigação de vulnerabilidades, como *Open Source Intelligence* (OSINT) (TABATABAEI; WELLS, 2016; LEE; SHON, 2016) e aprendizado de máquina (TRUONG; ZELINKA, 2019; HEINL, 2014), sendo essas comumente empregadas em ações de coleta e filtragem de dados.

Dessa forma, com vistas a explorar possíveis formas de ataques e aliado ao objetivo de "estimular o desenvolvimento e a inovação de soluções de segurança cibernética nas tecnologias emergentes" previsto na Estratégia Nacional de Segurança Cibernética (BRASIL, 2020, p. 7), este artigo propõe uma arquitetura integrada com o objetivo de otimizar ações de exfiltração de dados, utilizando como meio a coleta dados de fontes abertas e o treinamento de um algoritmo com técnicas de *machine learning*, permitindo assim a seleção de artefatos de maior relevância em ataques de exfiltração de dados com um menor risco de detecção.

As próximas seções foram estruturadas de forma a contextualizar as principais técnicas envolvidas, com base em referências bibliográficas, e detalhar a arquitetura proposta. Por fim, os resultados de uma prova de conceito são apresentados demonstrando a efetividade da arquitetura proposta ao atingir seus objetivos.

2 CONTEXTUALIZAÇÃO

Considerando a necessidade de busca de informações em um contexto de abundância de dados, é interessante que técnicas de coleta de dados abertos, conhecidas também como técnicas *OSINT*, sejam aplicadas para a filtragem e seleção daqueles dados mais relevantes ao tema proposto. A coleta pode utilizar fontes abertas de dados como, por exemplo, sites na Internet, informações de jornais e revistas, informações públicas governamentais ou trabalhos acadêmicos.

Entende-se *OSINT* como a coleta de dados em fontes abertas, sendo descrita em detalhes por CEPIK (2003):

> *OSINT consiste na obtenção legal de documentos oficiais sem restrição de segurança, da observação direta e não clandestina dos aspectos políticos, militares e econômicos da vida interna de outros países ou alvos, do monitoramento da mídia (jornais, rádio e televisão), da aquisição legal de livros e revistas especializadas de caráter técnico-científico, enfim, de um leque mais ou menos amplo de fontes disponíveis cujo acesso é permitido sem restrições especiais de segurança. (CEPIK, 2003, p. 51)*

Conforme mencionado, a rede mundial de computadores é hoje uma fonte valiosa no que tange à coleta de dados abertos, principalmente por possuir um volume crescente e disponível de dados. Uma vez selecionados os domínios de Internet de interesse a partir da definição de um contexto e da utilização de técnicas *OSINT*, um coletor de dados *web*, também conhecido pelo termo *crawler*, pode ser utilizado para a efetiva coleta dos dados.

O mecanismo básico de funcionamento de um *web crawler* é efetivamente descrito por HEYDON e NAJORK (1999):

> *O algoritmo básico utilizado por qualquer web crawler, recebe como parâmetro de entrada uma lista de URLs iniciais, e repetidamente executa as etapas a seguir. Remove uma URL da lista de URLs, determina o endereço IP correspondente ao nome do servidor hospedeiro, faz o download do documento correspondente, e extrai qualquer link ali contido. Para cada link extraído, assegura que é uma URL em formato absoluto, e a adiciona à lista de URLs para download, certificando de que não se trata de uma URL já encontrada anteriormente*. (HEYDON; NAJORK, 1999, p. 220, tradução nossa).

Além de um mecanismo básico de funcionamento, HEYDON e NAJORK (1999) definem que um algoritmo do tipo *web crawler* demanda cinco componentes básicos: um componente para armazenar e baixar as URLs; um componente para resolver nomes de domínio para endereços IP; um componente para realizar o *download* dos documentos; um componente para extrair os *links* de um documento HTML baixado; e um componente para determinar se uma URL já foi encontrada anteriormente.

De acordo com essa definição de *web crawler*, reforça-se que os parâmetros iniciais de seu algoritmo devem ser uma lista de URLs de interesse. Essas URLs podem ser obtidas de diversas formas, sendo destaque sua obtenção a partir de um trabalho de pesquisa em fontes abertas com a aplicação de técnicas e ferramentas *OSINT*.

Mesmo após uma criteriosa seleção de domínios de interesse, utilizando os métodos de coleta de informações a partir de fontes abertas, o volume de dados coletados por um *web crawler* pode ainda ser enorme. Isso ocorre pois o algoritmo processa recursivamente cada URL e os *links* referenciados em cada página HTML processada.

Todo esse volume de dados cria a necessidade de aplicação de outros métodos para seleção e filtragem daqueles mais relevantes ao tema de interesse. A aplicação de técnicas de aprendizado de máquina podem auxiliar nessa tarefa.

O aprendizado de máquina (*machine learning*) pode ser definido como:

> *[...] um conjunto de métodos que podem detectar automaticamente padrões nos dados e, em seguida, usar os padrões descobertos para prever dados futuros, ou para realizar outros tipos de tomada de decisão sob incerteza. (MURPHY, 2012, p. 1, tradução nossa).*

Algoritmos de aprendizado de máquina geralmente são divididos em dois tipos principais: algoritmo supervisionado e algoritmo não supervisionado. Os supervisionados tem como objetivo predizer um rótulo para um conjunto de dados novos, considerando uma base anterior de dados rotulados. Já os não supervisionados geralmente são conhecidos como algoritmos de descobrimento de padrões. Seu objetivo é principalmente encontrar "padrões interessantes" em um conjunto de dados não rotulados (MURPHY, 2012, p. 2, tradução nossa).

Assim, os dados coletados por meio de técnicas e ferramentas *OSINT* em integração com *web crawlers* podem ser utilizados em conjunto com a tecnologia de aprendizado de máquina para a detecção de padrões e seleção de dados de interesse, por exemplo em uma atividade de exfiltração de dados.

Sobre exfiltração de dados:

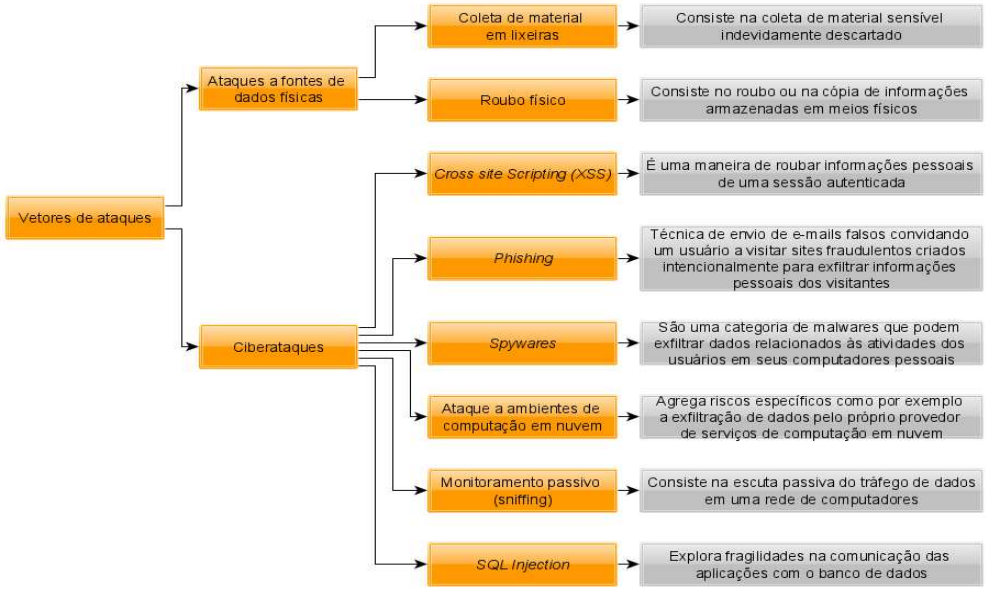
> *O roubo de dados (formalmente denominado exfiltração de dados) é um dos principais motivadores dos ataques cibernéticos, independentemente de serem realizados pelo crime organizado, concorrentes comerciais, atores estatais ou mesmo "hacktivistas". Evitar a exfiltração de dados está se tornando cada vez mais uma tarefa desafiadora por dois motivos principais. Em primeiro lugar, ao longo dos últimos anos, o crime cibernético passou de um ato individual para um ato organizacional. Essa transformação forneceu aos invasores (ou frequentemente chamados de hackers) alto orçamento, recursos e metodologias sofisticadas para se tornarem mais profissionais na exfiltração de dados. Em segundo lugar, a infraestrutura de dados existente contém vários meios que são originalmente projetados para a troca legítima de dados, mas que podem ser usados para exfiltração de dados. (ULLAH; FAHEEM, 2017, p. 1, tradução nossa).*

As fontes passíveis de sofrerem ataques objetivando a exfiltração de dados são diversas, podem ser servidores de arquivos, servidores de e-mail, dispositivos móveis, bancos de dados, servidores FTP, ou qualquer aplicação disponível na Internet que possa conter dados de interesse do atacante. Outras fontes físicas como documentos em uma impressora, *pendrives*, e discos removíveis também estão passíveis de outras formas de ataque. No ambiente cibernético, a metodologia básica consiste em coletar e analisar informações do alvo em busca de pistas sobre possíveis vulnerabilidades que possam ser exploradas para perpetrar um ataque.

A categorização e contextualização dos vetores de ataque, pode ajudar a mapear e compreender as contra-medidas necessárias. ULLAH *et al.* (2018, p. 7-45) sugerem ainda as principais formas de ataque para exfiltração de dados e as correspondentes possíveis contramedidas, organizadas nos diagramas das Figura 1 e 2.

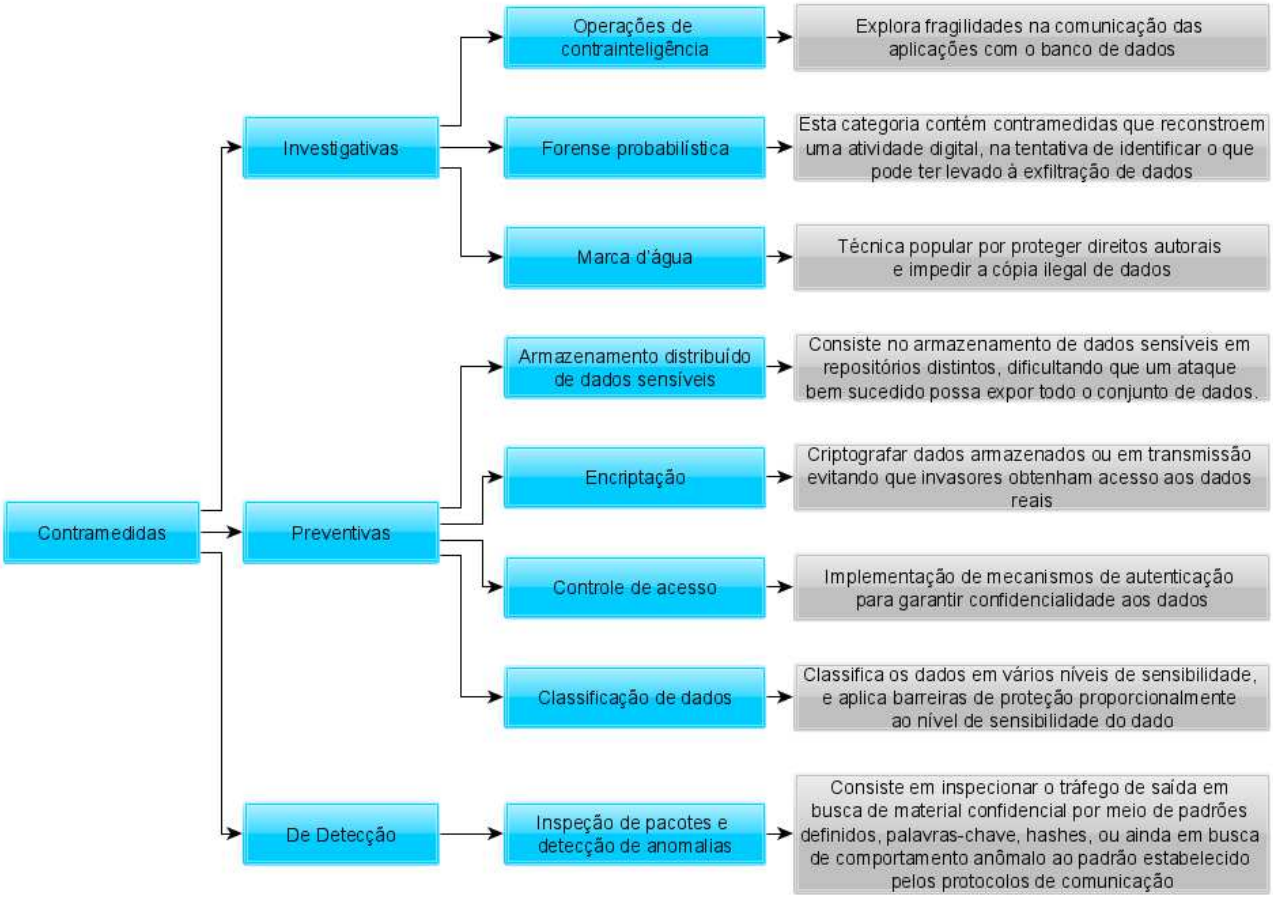
O emprego de outras contramedidas comuns, como *Intrusion Detection System* (IDS) e *Intrusion Prevention System* (IPS), pode ainda ser ressaltado pela capacidade de detectar ataques e outros tipos de anomalias na rede, podendo gerar alertas ou ainda tomar ações defensivas de forma automatizada, como por exemplo a implementação de regras de bloqueio no firewall.

Figura 1 - Vetores de ataques



Elaborada pelos autores (conteúdo adaptado de ULLAH *et al.*, 2018).

Figura 2 - Vetores de contramedidas



Elaborada pelos autores (conteúdo adaptado de ULLAH *et al*., 2018).

As fontes passíveis de sofrerem ataques objetivando a exfiltração de dados são diversas, podem ser servidores de arquivos, servidores de e-mail, dispositivos móveis, bancos de dados, servidores FTP, ou qualquer aplicação disponível na Internet que possa conter dados de interesse do atacante.

Outras fontes físicas como documentos em uma impressora, *pendrives*, e discos removíveis também estão passíveis de outras formas de ataque. No ambiente cibernético, a metodologia básica consiste em coletar e analisar informações do alvo em busca de pistas sobre possíveis vulnerabilidades que possam ser exploradas para perpetrar um ataque.

A categorização e contextualização dos vetores de ataque, pode ajudar a mapear e compreender as contra-medidas necessárias. ULLAH *et al.* (2018, p. 7-45) sugerem ainda as principais formas de ataque para exfiltração de dados e as correspondentes possíveis contramedidas, organizadas nos diagramas das Figuras 1 e 2.

O emprego de outras contramedidas comuns, como *Intrusion Detection System* (IDS) e *Intrusion Prevention System* (IPS), pode ainda ser ressaltado pela capacidade de detectar ataques e outros tipos de anomalias na rede, podendo gerar alertas ou ainda tomar ações defensivas de forma automatizada, como por exemplo a implementação de regras de bloqueio no firewall.

Em ataques de exfiltração de informações, o volume de dados trafegando para fora da rede local pode provocar a identificação da ação maliciosa por parte desse tipo de equipamento de defesa de perímetro, possibilitando a implementação de contramedidas (ULLAH *et al.*, 2018).

Deste modo, uma arquitetura que permita, de modo automatizado, a seleção de dados alvo antes do processo de exfiltração em si, poderia diminuir a chance de detecção do ataque.

3 ARQUITETURA

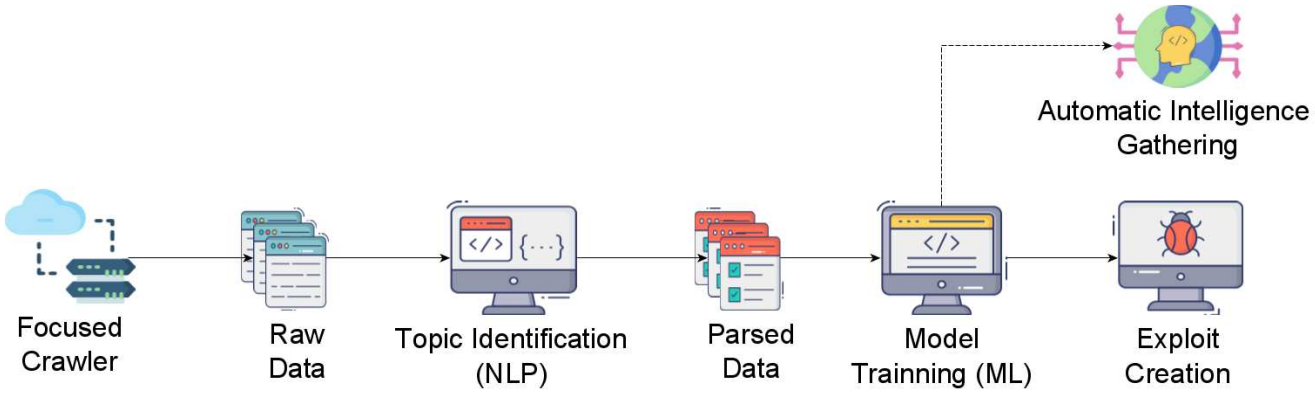
Considera-se hoje que grande parte dos ataques cibernéticos realizados enquadram-se na *Cyber Kill Chain*, *framework* de defesa cibernética orientado à inteligência formalizado pela empresa Lockheed Martin com o objetivo de melhorar a visibilidade de ataques e enriquecer a compreensão de analistas de defesa quanto às táticas utilizadas por atacantes cibernéticos (LOCKHEED MARTIN, 2021). O framework foi baseado no conceito militar de *Kill Chain* e prevê a divisão de um ataque cibernético em sete fases:

a) Reconhecimento: pesquisa, identificação, seleção e enumeração de vulnerabilidades do alvo;
b) Armamentização (*Weaponization*): criação de um *payload* e integração a um *exploit* que possa obter acesso ao alvo;
c) Entrega: utilização do *exploit* definido e transmissão do *payload* para o alvo;
d) Exploração: acionamento do *exploit* e exploração de vulnerabilidades no alvo;
e) Instalação: criação de persistência de acesso ao alvo;
f) Comando e Controle: estabelecimento de um fluxo de comunicação continua, a ser ativada pertinentemente, entre a rede do atacante e a rede alvo;
g) Ações e Objetivos: realização da ação objetivo, missão do ataque.

A arquitetura proposta neste artigo, denominada a partir daqui de Arquitetura IDEA (*Intelligent Data Exfiltration Architecture*), concentra-se, principalmente, nas fases de reconhecimento e "armamentização" da *Cyber Kill Chain*, automatizando as ações de identificação e seleção de dados alvo em um contexto pré-definido.

O objetivo é permitir a execução de um ataque canalizado e customizado, utilizando dados previamente obtidos para evitar a detecção do ataque. A Figura 3 apresenta em resumo os componentes previstos na arquitetura.

Figura 3 - Principais componentes da Arquitetura IDEA.



Fonte: Elaborada pelos autores (2021).

Os dois primeiros componentes da arquitetura relacionam-se à fase de reconhecimento da *Cyber Kill Chain* e os dois últimos à fase de armamentização.

O primeiro componente, denominado *focused crawler*, é um *web crawler* alimentado com domínios e subdomínios de interesse, sendo estes coletados a partir de pesquisas *OSINT* com o auxílio de motores de busca na *web*. Os motores de busca são utilizados para pesquisas por palavras chaves relacionadas ao alvo de interesse e por domínios obtidos por meio de sub *crawlers OSINT* automatizados.

O objetivo desse primeiro componente é coletar a maior quantidade de dados na Internet sobre o alvo definido e organizá-los em documentos de entrada para o segundo componente, nos mais diversos formatos, como por exemplo arquivos de texto em formato pdf, word, etc.

O segundo componente possui como objetivo detalhar a filtragem inicial, classificando os documentos como relevantes ou não. Cada documento coletado é analisado e seu tema chave identificado por meio de uma técnica de modelagem de tópico (*Topic Modelling*).

Essa técnica, derivada do processamento de linguagem natural (*NLP - Natural Language Processing*), utiliza conceitos de Inteligência Artificial para identificar coleções de palavras que "fazem sentido juntas" (BOYD-GRABER; HU; MIMNO, 2017, p. 4, tradução nossa), distinguindo tópicos em textos antes desconhecidos.

O terceiro componente inicia o processo de armamentização do ataque, treinando um modelo de aprendizado de máquina supervisionado com os documentos já classificados pelo modelador de tópico. O objetivo é criar um modelo com uma precisão de pelo menos 98% que consiga, após a fase de treinamento, classificar um novo documento não incluso na base de treinamento como relevante ou não, podendo assim ser utilizado pelo último componente da arquitetura.

O quarto e último componente finaliza o processo de armamentização e pode ser personalizado conforme o nível de acesso ao alvo, cuja obtenção não é escopo desta arquitetura.

Na Figura 3, exemplifica-se a utilização do modelo treinado tanto na construção de um *exploit* local específico quanto na construção de um novo *web crawler* que utilize o modelo treinado para coleta e curadoria de dados em um contexto de produção de inteligência.

4 PROVA DE CONCEITO

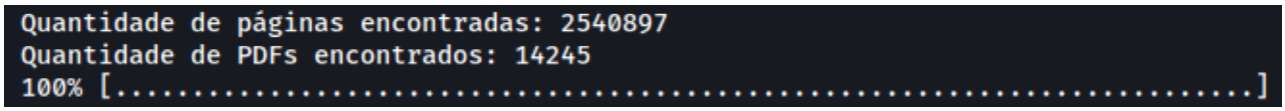
O seguinte escopo foi considerado na criação da prova de conceito de utilização da arquitetura: restrição ao formato dos dados de entrada (somente arquivos PDF) e limitação a um único servidor alvo de ataque, já existindo conexão estabelecida entre esse servidor e a máquina atacante.

O principal objetivo desta prova de conceito resumiu-se a comparar a utilização de um *exploit* de exfiltração de dados gerado a partir do fluxo da arquitetura IDEA com um suposto *exploit* sem um componente de busca por contexto.

O primeiro componente da arquitetura, denominado *focused crawler*, foi desenvolvido a partir de um script na linguagem de programação Python versão 3.8. Esse componente foi pré-alimentado com informações de domínios de interesse coletados por um script auxiliar também escrito em Python. As principais bibliotecas utilizadas na criação de ambos foram: *urllib* versões 1 e 3, *beaultifulsoup* versão 4, e *googleapiclient*.

Todos os scripts foram testados em ambiente de virtualização em nuvem (*Amazon Web Services* - AWS) e várias execuções foram realizadas para a coleta dos dados que seriam utilizados nas fases subsequentes. Algumas execuções chegaram a inspecionar mais de 2 milhões de páginas *web*, encontrando 14.245 documentos PDF de interesse para execução da prova de conceito.

Figura 4 - Resultado de execução do focused crawler.



Fonte: Elaborada pelos autores (2021).

Os componentes de modelagem de tópico e de aprendizado de máquina supervisionado foram desenvolvidos utilizando ferramentas de processamento de linguagem natural também em Python versão 3.8. Os principais módulos utilizados foram spacy (HONNIBAL *et al.*, 2021) e scikit-learn (PEDREGOSA *et al.*, 2011).

Os scripts criados foram inicialmente testados em notebooks Jupyter (PROJECT JUPYTER, 2021) no ambiente Anaconda (ANACONDA INC., 2021). O componente de modelagem de tópico utilizou a base treinada em linguagem português brasileiro do módulo spacy (*pt_core_news_lg*) para detectar, iterativamente, os principais tópicos compostos por 15 principais palavras-chave cada.

Considerando que o *focused crawler* foi o responsável por prover uma base de documentos com um filtro inicial do contexto do alvo, 10 tópicos foram definidos como parâmetro para detecção de um conjunto de arquivos. O conteúdo dos arquivos foi lido, sanitizado, tokenizado e lemantizado (SARAVIA, 2020) para que assim pudesse ser processado pelo *pipeline* de detecção de tópicos, que incluiu um elemento de extração de *features* e um elemento de decomposição.

A saída definida para o componente de modelagem de tópico foi uma base de dados contendo três colunas: identificador do arquivo, seu conteúdo processado e uma flag booleana de indicação de relevância quanto ao contexto pré-definido.

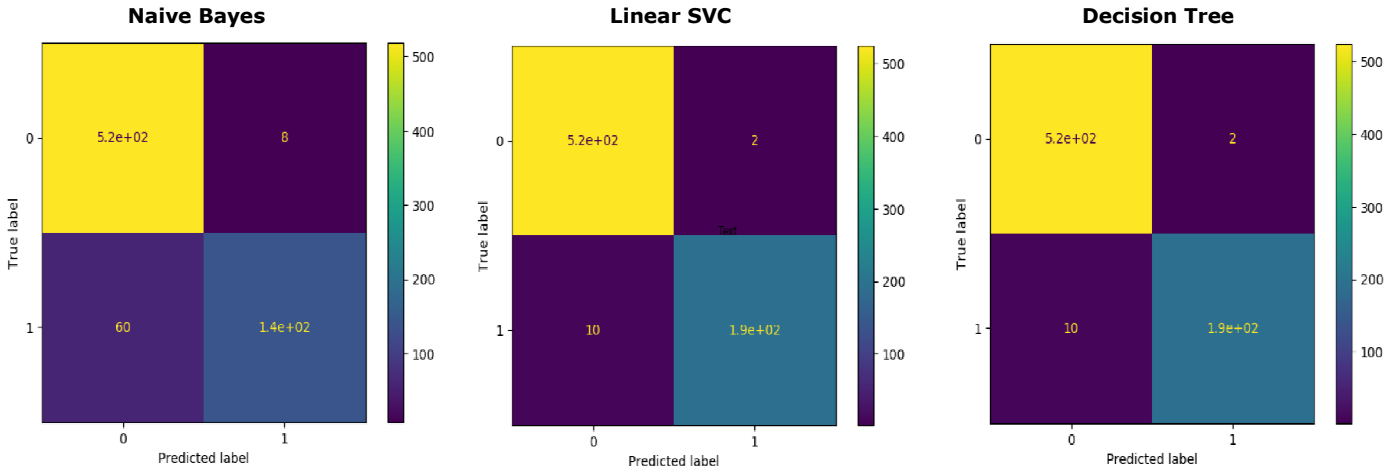
Aproximadamente 20% dos arquivos obtidos pelo componente de *web crawler* não puderam ser lidos, principalmente por possuírem um formato não suportado pelo componente, como arquivos PDF salvos como imagem, com *encoding* modificado, entre outros. Dessa forma, a base final criada e utilizada no treinamento do modelo de classificação continha 9.444 linhas.
O modelo de classificação foi definido a partir de testes com três algoritmos de classificação: *Naive Bayes*, *Linear Support Vector* e Árvore de Decisão. O processamento dos dados levou em consideração o processo de sanitização, tokenização e lemantização realizado pelo componente de identificação de tópico e o treinamento foi realizado após o balanceamento das classes de predição.

A Figura 5 apresenta a matriz de confusão obtida com cada algoritmo e a precisão de cada modelo, calculada com base na técnica de *cross validation* (BERRAR, 2019).

Observa-se na figura que o algoritmo de árvore de decisão foi o que apresentou melhor precisão de resultado. O modelo treinado foi, portanto, exportado para que pudesse ser testado como detector de arquivos relevantes em um cenário de exploração.

Conforme delimitado pelo escopo, o cenário de exploração escolhido para a prova de conceito foi a tentativa de exfiltração de dados em uma máquina alvo com conexão previamente estabelecida. Um *exploit* em Python foi criado para realizar a detecção de arquivos de interesse utilizando o modelo treinado para otimizar o ataque.

Figura 5 - Matriz de confusão e precisão de cada modelo treinado.

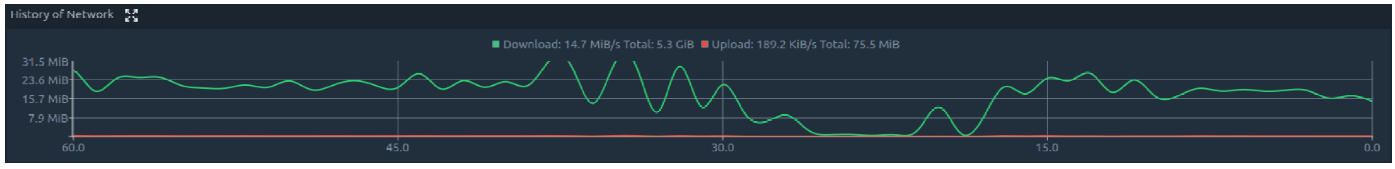


Fonte: Elaborada pelos autores (2021).

No intuito de mensurar a diferença no volume de dados exfiltrados e passíveis de serem identificados por mecanismos de segurança já discutidos em tópicos anteriores, como por exemplo sitemas do tipo IDS/IPS, foram avaliadas duas situações a partir do monitor de recursos de uma máquina atacante configurada com sistema Kali Linux 2021.

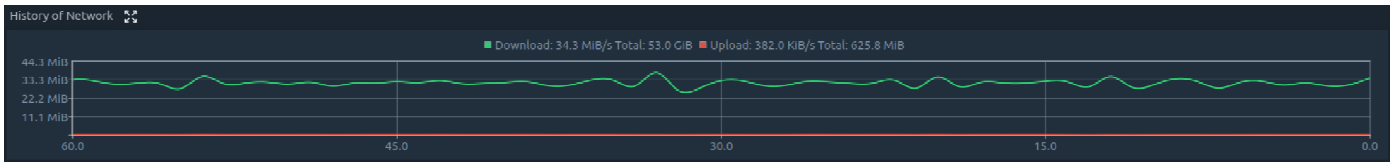
Utilizando como teste um novo conjunto de dados, a Figura 6 apresenta a simulação de um atacante que faz uso do *exploit* em Python que implementa um modelo treinado para seleção de arquivos de interesse para a realização de um ataque de exfiltração de dados otimizado. Já a Figura 7 apresenta a simulação de um atacante que não faz uso do *exploit* otimizado, mas que simplesmente coleta todo o volume de dados encontrado na máquina alvo.

Figura 6 - 5.3GB de dados coletados com o auxílio de um *exploit* otimizado.



Fonte: Elaborada pelos autores (2021).

Figura 7 - 53GB de dados coletados sem o auxílio de um *exploit* otimizado.



Fonte: Elaborada pelos autores (2021).

Nota-se que o uso do algoritmo treinado para a detecção de arquivos de interesse reduziu em 90% a quantidade de dados vazados de dentro da organização alvo, sendo mais efetivo e cumprindo seu objetivo de reduzir o tempo total para a conclusão do ataque além de evitar sua identificação por parte dos administradores da rede alvo ou por mecanismos de segurança automatizados.

A utilização de mecanismos de segurança comuns com regras específicas para detecção de saída de volume de dados acima de um padrão esperado, como o IDS/IPS Snort (ROESCH, 2021), poderiam, assim, ser facilmente contornados realizando a exfiltração somente dos dados de interesse.

Todos os scripts criados foram disponibilizados em um repositório público no GitHub (ANDRADE; DEUS, 2021).

5 CONCLUSÃO

Os resultados obtidos com a prova de conceito demonstram, portanto, ser relevante a utilização da arquitetura IDEA em alvos pré-determinados, já que, considerando a diminuição de 90% no volume de dados não-relevantes exfiltrados na prova de conceito, os objetivos propostos inicialmente foram alcançados quanto a diminuição dos riscos de detecção e o aumento de agilidade no processo de exploração.

A arquitetura proposta muito se adequa ao valor de utilização de técnicas de inteligência artificial nas fases iniciais da *Cyber Kill Chain* destacado por HEINL (2014). De acordo com a autora, técnicas de inteligência artificial podem "auxiliar na detecção precoce de vulnerabilidades e prover consciência situacional", além de auxiliar na "coleta de informações e suporte à decisão" (HEINL, 2014, p. 60, tradução nossa), evidenciando-se assim sua aplicação *dual*.

A utilização de técnicas de *OSINT* aliadas às técnicas de inteligência artificial permitem, no contexto de exploração cibernética, ir além e criar um perfil canalizado, customizado e de alto valor de alvos de interesse. Esse perfil pode ser o ponto de partida para construção de um ataque avançado de persistência, mais conhecido como *Advanced Persistent Threat* (APT). Um ataque APT baseia-se no conceito de uma ameaça cibernética persistente, sem prazo, cujo objetivo é obter informações restritas específicas de um alvo da forma mais furtiva possível (COLE, 2013).

Vislumbrando a evolução da aplicabilidade e eficácia da arquitetura IDEA, diversas melhorias podem ser objeto de trabalhos futuros, como: expansão do formato de entrada de dados, permitindo arquivos de formatos e idiomas diversos; a utilização de motores de busca alternativos no *focused crawler*, como Yandex, Bing, etc; a melhoria de performance de execução de todos os componentes, permitindo execuções paralelas e distribuídas; testes com mais algoritmos de modelagem de tópico e de classificação, incluindo um comparativo com a utilização de técnicas de *Deep Learning*; a construção de um *crawler* na camada de armamentização que permita a utilização do modelo treinado para coleta de inteligência diretamente na Internet; entre outros.

Por fim, reforça-se que um dos principais objetivos de qualquer ação de exploração cibernética é o de que essa não seja detectada pelo alvo (TRUONG; ZELINKA, 2019). Dessa forma, a arquitetura proposta nesse artigo auxilia no cumprimento deste objetivo ao demonstrar que sua utilização, mesmo que simplificada na prova de conceito, permite diminuir os riscos de detecção de um ataque de exfiltração de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

ANACONDA INC. (org.). **Anaconda**: the world's most popular data science platform. Disponível em: https://www.anaconda.com/. Acesso em: 20 maio 2021.

ANDRADE, I. C.; DEUS, G. R. **Intelligence gathering architecture**. 2021. Disponível em: https://github.com/isabellecda/intgl-gathering-arch. Acesso em: 20 maio 2021.

COLE, E. **Advanced Persistent Threat**: understanding the danger and how to protect your organization. Massachusetts, USA: Elsevier, 2013.

EXPLOSION SOFTWARE COMPANY. **Spacy**: industrial-strength natural language processing. Disponível em: https://spacy.io/. Acesso em: 20 maio 2021.

BERRAR, D. Cross-Validation. **Encyclopedia Of Bioinformatics And Computational Biology**, Oxford, v. 1, p. 542-545, 2019.

BIG Data and Information Security: Most Feared Cyber-threats. **Business Application Research Center**, 2021. Disponível em: https://bi-survey.com/cyber-threats-types. Acesso em: 04 abr. 2021.

BOYD-GRABER, J.; HU, Y.; MIMNO, D. Applications of Topic Models. **Foundations And Trends® In Information Retrieval**, [S.L.], v. 11, n. 2-3, p. 143-296, 2017. Now Publishers. http://dx.doi.org/10.1561/1500000030.

BRASIL. Decreto nº 9.637, de 26 de dezembro de 2018. Institui a Política Nacional de Segurança da Informação. **Diário Oficial da União**: seção 1, Brasília, DF, n. 248, p. 23, 27 dez. 2018.

BRASIL. Decreto nº 10.222, de 5 de fevereiro de 2020. Aprova a Estratégia Nacional de Segurança Cibernética. **Diário Oficial da União**: seção 1, Brasília, DF, n. 26, p. 6, 6 fev. 2020.

CEPIK, M. A. C. **Espionagem e Democracia**. Rio de Janeiro: Editora FGV, 2003.

DEWAR, R. S. The "triptych of cyber security": a classification of active cyber defence. In: 2014 6TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT (CYCON), 6., 2014, Tallinn, Estonia. **Proceedings [...] .** Tallinn, Estonia: IEEE, 2014. p. 7-21.

JOSSEN, S. The world's most valuable resource is no longer oil, but data. **The Economist**. 06 maio. 2017. Disponível em: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. Acesso em: 04 abr. 2021.

HEINL, C. H. Artificial (intelligent) agents and active cyber defence: policy implications. In: 2014 6TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT (CYCON), 6., 2014, Tallinn, Estonia. **Proceedings [...] .** Tallinn, Estonia: IEEE, 2014. p. 53-66.

HEYDON, A.; NAJORK, M. M.: A scalable, extensible Web crawler. **Compaq Systems Research Center**, Palo Alto, p. 220, dez. 1999.

HONNIBAL, M. *et al*. **spaCy**: Industrial-strength Natural Language Processing in Python. Versão 3.0.6. [S. l.],2016. Disponível em: https://spacy.io/. Acesso em: 20 maio 2021.

KUNDER, M. **The size of the World Wide Web (The Internet)**. Disponível em: https://www.worldwidewebsize.com/. Acesso em: 04 abr. 2021.

LEE, S.; SHON, T. Open source intelligence base cyber threat inspection framework for critical infrastructures. In: 2016 FUTURE TECHNOLOGIES CONFERENCE (FTC), 1., 2016, San Francisco, CA, USA. **Proceedings [...]** . San Francisco, CA, USA: IEEE, 2016. p. 1030-1033.

LOCKHEED MARTIN. **Cyber Kill Chain**. 2021. Disponível em: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html. Acesso em: 30 abr. 2021.

MUNCASTER, P. US Military Personnel Exposed in Latest Cloud Data Leak. **Info Security Magazine**, 22 out. 2019. Disponível em: https://www.infosecurity-magazine.com/news/military-personnel-exposed-latest/. Acesso em: 07 abr. 2021.

MURPHY, K. P. **Machine Learning:** a probabilistic perspective. Massachusetts: Massachusetts Institute of Technology, 2012.

O'DEA, S. Estimated internet traffic in the United States from 2018 to 2023. **STATISTA**, 09 jun. 2020. Disponível em: https://www.statista.com/statistics/216335/data-usage-per-month-in-the-us-by-age/. Acesso em: 04 abr. 2021.

PEDREGOSA, F. *et al*. Scikit-learn: machine learning in python. **Journal Of Machine Learning Research**, [S. L.], v. 12, p. 2825-2830, 2011.

PROJECT JUPYTER (org.). **Jupyter**. Disponível em: https://jupyter.org/. Acesso em: 20 maio 2021.

RAPÔSO, C. F. L. *et al*. LGPD-Lei Geral de Proteção de Dados Pessoais em Tecnologia da Informação: Revisão Sistemática. **RACE-Revista de Administração do Cesmac**, v. 4, p. 58-67, 2019.

ROESCH, M. **Snort**: network intrusion detection & prevention system. Network Intrusion Detection & Prevention System. Disponível em: https://www.snort.org/. Acesso em: 22 maio 2021.

SARAVIA, E. **Fundamentals of NLP**: tokenization, lemmatization, stemming, and sentence segmentation. 2020. Disponível em: https://dair.ai/notebooks/nlp/2020/03/19/nlp_basics_tokenization_segmentation.html. Acesso em: 19 jun. 2021.

TABATABAEI, F.; WELLS, D. OSINT in the Context of Cyber-Security. In: AKHGAR, Babak *et al* (ed.). **Open Source Intelligence Investigation**: from strategy to implementation. [S.l.]: Springer, 2016. p. 213-231.

TRUONG, C. T., ZELINKA, I. A Survey on Artificial Intelligence in Malware as Next-Generation Threats. **MENDEL**, v. 25, n. 2, p. 27-34, 20 dez. 2019.

ULLAH, F. *et al*. Data exfiltration: a review of external attack vectors and countermeasures. **Journal Of Network And Computer Applications**, [S.l.], v. 101, p. 18-54, 1 jan. 2018. Elsevier BV.

*Artigo realizado a partir do trabalho de conclusão do Curso de Especialização em Guerra Cibernética do Centro de Instrução de Guerra Eletrônica – CIGE pelos Tenentes Isabelle Cecilia de Andrade e Guilherme Resende Deus.

Endereço postal: DF-001, 5, Lago Norte. Brasília, Distrito Federal – DF, CEP: 71559-902. email: isabelleica@fab.mil.br, guilhermegrd@fab.mil.br.