UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Pedro Henrique Silva Santana

Ferramenta para Tratamento e Análise Visual de Dados Capturados por Honeypots IoT

Uberlândia, Brasil

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Pedro Henrique Silva Santana

Ferramenta para Tratamento e Análise Visual de Dados Capturados por Honeypots IoT

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Rodrigo Sanches Miani

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2025

Pedro Henrique Silva Santana

Ferramenta para Tratamento e Análise Visual de Dados Capturados por Honeypots IoT

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 13 de maio de 2025:

Prof. Dr. Rodrigo Sanches Miani
Orientador

Prof. Dr. Ivan da Silva Sendin
Convidado 1

Prof. Dr. Paulo Henrique Ribeiro Gabriel
Convidado 2

Uberlândia, Brasil
2025

Dedico este trabalho a todos aqueles que, mesmo com suas mentes sobrecarregadas de pensamentos, nunca desistiram de pensar.

Agradecimentos

Agradeço à minha família — em especial à minha mãe, Maria José, que sempre investiu no meu aprendizado e nunca mediu esforços para que eu tivesse todos os recursos necessários para crescer e aprender.

À minha amada Heloísa Benatt, que esteve ao meu lado desde o início da minha jornada, o meu mais sincero agradecimento por todo o apoio, pelas palavras de incentivo e por acreditar em mim e em um futuro brilhante ao meu lado.

Aos amigos que fiz ao longo dessa jornada, em especial André Oliveira, Guilherme Dias, Gustavo Ramos e Victor Hugo, minha gratidão por tornarem essa caminhada mais leve, divertida e inspiradora. Aprender e evoluir com vocês tornou essa fase inesquecível.

Agradeço, em especial, ao meu orientador Rodrigo Miani, por sua contribuição, paciência e valiosas instruções ao longo de todo o curso. Seu apoio e incentivo foram essenciais para minha formação e crescimento acadêmico.

Por fim, agradeço à minha instituição de ensino e a todos os professores que, com devoção, compartilharam seus conhecimentos e contribuíram imensamente para minha formação. Carregarei o conhecimento adquirido ao longo da vida e da carreira, com o objetivo de retribuir à sociedade tudo aquilo que recebi.

Resumo

A evolução da tecnologia e o investimento em dispositivos IoT conectados à internet impõem desafios significativos à segurança cibernética, incentivando o uso de estratégias como os *honeypots* para investigar ataques e comportamentos de invasores. Este trabalho apresenta o desenvolvimento de uma ferramenta em Python voltada para o tratamento, organização e análise visual de dados coletados por *honeypots*. A ferramenta realiza a transformação de arquivos de log em estruturas de *DataFrames*, corrigindo inconsistências e explorando métricas relevantes, permitindo a visualização gráfica. Foram implementadas funções básicas para análise e comparação entre diferentes conjuntos de dados, possibilitando a identificação de padrões de comportamento entre os atacantes. Os testes foram conduzidos com base nos dados obtidos a partir de dois experimentos executados pelo grupo de pesquisa, nomeados de "Experimento - Roteadores" e "Experimento - Câmeras", ambos consistindo na disponibilização virtual de *honeypots* que emulam interfaces de acesso com o objetivo de atrair e registrar atividades de atacantes, evidenciando a eficácia da exploração desses dados. Como resultado, a ferramenta oferece soluções que auxiliam na interpretação dos dados coletados, com potencial para apoiar estudos futuros e servir como base sólida para investigações sobre tentativas de ataque.

Palavras-chave: *Honeypot*, IoT, Análise de dados, Ferramenta, Python.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

CIA	*Confidentiality, Integrity, Availability*
DDoS	*Distributed Denial of Service*
ETL	*Extract, Transform, Load*
HTTP	*HyperText Transfer Protocol*
IDS	*Intrusion Detection System*
IP	*Internet Protocol*
IoT	*Internet of Things*
MQTT	*Message Queue Telemetry Transport*
NFC	*Near Field Communication*
NuSec	Núcleo de Segurança
RFID	*Radio-Frequency Identification*
SSH	*Secure Shell*
TCP	*Transmission Control Protocol*
URL	*Uniform Resource Locator*
WSN	*Wireless Sensor Network*
XSS	*Cross-Site Scripting*

Sumário

1 Introdução

Nos últimos anos, com o desenvolvimento da tecnologia e a popularização de dispositivos auxiliares com conexão à rede de Internet, a temática de segurança vem sendo cada vez mais abordada tanto por usuários comuns dessas ferramentas quanto por grandes companhias a fim de assegurar seus dados sensíveis (CANONGIA; JUNIOR, 2009). Para garantir a segurança destes usuários, várias estratégias são utilizadas em dispositivos para que seja possível evitar sua invasão e proteger seus dados. Uma dessas estratégias a serem exploradas são as ferramentas *honeypots* que simulam dispositivos e sistemas, atraindo possíveis invasores e para que seja possível avaliar seus comportamentos ao comprometer um dispositivo (PERKINS; HOWELL, 2021) e, assim, relatando os passos tomados,conteúdos acessados, possíveis vulnerabilidades, comandos e padrões (TABARI; OU; SINGHAL, 2021).

A implementação de estruturas de *honeypot* é tema de pesquisa em diversos trabalhos no campo da segurança da informação. Mendes (2023) aborda o processo de construção, coleta e análise de *honeypots* que simulam interfaces de roteadores, disponibilizados durante um período de 15 dias, com o objetivo de avaliar a eficiência da ferramenta. O "Experimento - Roteadores" conduzido por Mendes (2023) obteve resultados que superaram as expectativas, possibilitando a identificação de comportamentos fora do padrão esperado. Seguindo essa metodologia, outro trabalho desenvolvido dentro do Núcleo de Segurança (NuSec) investigou conceitos semelhantes, utilizando *honeypots* que simulam câmeras de vigilância. O "Experimento - Câmeras" conduzido por Costa (2025) apresentou resultados semelhantes aos do estudo anterior, comprovando a eficácia da estratégia em atrair invasores e capturar padrões de comportamento em ambientes simulados.

Dispositivos conectados à Internet, como assistentes virtuais que interagem entre si, auxiliam usuários em atividades cotidianas e gerenciam rotinas programadas. Estes dispositivos passaram a ser cada vez mais comuns no mercado de tecnologia e se associam ao conceito da internet das coisas ou *Internet of the Things (IoTs)* (MAGRANI, 2018). Eletrodomésticos, *smartphones*, assistentes pessoais ou qualquer dispositivo interativo com acesso à rede de internet passaram a ser uma realidade para várias pessoas e a se destacar no mercado mundial de tecnologia.

Durante os períodos de 2020 e 2021, o mundo passou pela pandemia do Corona vírus e, apesar de todos os problemas resultantes deste período, o crescimento dos setores tecnológicos obteve grande expressividade, com destaque aos setores voltados à saúde que obtiveram um crescimento próximo a 20% em 2020 comparado com o ano anterior. Dispositivos vestíveis (*weareables*) que mensuram a frequência cardíaca, oxigenação sanguínea

e qualidade de sono do usuário tiveram uma maior demanda, assim como dispositivos que auxiliam e otimizam procedimentos laboratoriais (UMAIR et al., 2021). Além disso,com o novo cenário que a pandemia do Covid-19 proporcionou, diversas empresas tiveram que se adaptar a um novo panorama organizacional quanto ao quadro de funcionários e à execução de suas atividades remotamente (*home office*) (PEREIRA, 2022), e mesmo que não imediatamente, para atender às novas expectativas e ao conforto para trabalho remoto, criando um ambiente interativo e conectado, o mercado de *smart homes* é outro segmento que tende a crescer nos próximos anos (UMAIR et al., 2021).

Com essa nova rotina e a facilidade do acesso às redes, usuários permanecem conectados por mais tempo. Além da preocupação que o Covid-19 gerou a respeito de informações sobre o cenário, os mesmos usuários passaram a assumir maiores riscos de tentativas de golpes e invasões de seus dispositivos (ANDRADE; ORTIZ-GARCéS; CAZARES, 2020). Dados são gerados e armazenados todos os dias e, muitas vezes, são alvos de criminosos que buscam se beneficiar ao se apossarem dessas informações, seja para coagir vítimas, raptar e vender seus dados, cometer roubo ou espionar suas rotinas (ABOMHARA; KØIEN, 2015).

1.1 Justificativa

A possibilidade de invasores se apossarem de dados, sejam eles pessoais ou informações sensíveis de empresas, é um problema que aflige todos que utilizam e confiam seus dados a dispositivos conectados à rede de internet. Para reduzir esse risco, estratégias de segurança são adotadas para evitar o sequestro de informações ou seu uso indevido. Os *honeypots* desempenham uma função fundamental na segurança de sistemas integrados com dispositivos IoT, atraindo invasores e coletando dados dos métodos de acesso e comportamentos dentro de um ambiente controlado.

Os resultados obtidos a partir da coleta desses dados consistem em análises e relatórios que auxiliam na identificação de vulnerabilidades e na compreensão dos principais objetivos dos agentes invasores. Diante dessa necessidade, é fundamental interpretar essas informações para a implementação de estratégias e metodologias eficazes, garantindo assim a segurança dos dispositivos IoT.

1.2 Objetivos e desafios da pesquisa

O objetivo deste trabalho é desenvolver uma ferramenta em Python para realizar o processo de ETL (Extração, Transformação e Carga) dos dados coletados por *honeypots* em diferentes experimentos. O intuito é analisar o comportamento de usuários mal-intencionados ao interagir com os campos de URL dessas ferramentas e comparar

padrões entre distintos *honeypots* ao longo do período em que estiveram ativos.

Como principal desafio, busca-se identificar e implementar um método eficiente para o tratamento e a estruturação desses dados, garantindo que possam ser explorados de forma otimizada. Além disso, é necessário abordar os dados com clareza e de forma dinâmica e objetiva durante a execução de pesquisas e a construção de relatórios sobre o tema.

1.3 Organização da monografia

Este trabalho foi organizado respeitando a seguinte estrutura: O Capítulo 2 explora conceitos presentes em bibliografias que fundamentam os principais tópicos e tecnologias para a evolução da pesquisa. São abordados conceitos de segurança cibernética , dispositivos IoT e *honeypot*, além dos trabalhos correlatos. O Capítulo 3 refere-se ao processo de desenvolvimento da ferramenta, passando pelas etapas de coleta e organização de arquivos, avaliação dos recursos necessários, a construção do método de investigação dos dados e definição da ferramenta para conversão e visualização gráfica dos dados. No Capítulo 4, os resultados obtidos durante a construção da ferramenta são apontados a partir dos avanços em relação às funções definidas. Finalmente, o Capítulo 5 encerra a discussão do trabalho com base nos resultados obtidos, destacando a relevância e a escalabilidade da ferramenta, além de disponibilizar uma base para estudos e trabalhos futuros.

2 Revisão Bibliográfica

Esta Seção tem como motivação listar e explicar os conceitos fundamentais que envolvem os objetos de estudo abordados neste trabalho, além de reunir os trabalhos correlatos à temática de segurança de dispositivos IoT e métodos para rastrear comportamentos de invasores, descrevê-los e relacionar seus conteúdos pertinentes à pesquisa.

O ambiente de segurança para dispositivos IoT é alvo de discussão de diversos pesquisadores e entusiastas de tecnologia. Os trabalhos desenvolvidos com base neste assunto têm como motivação avaliar ataques e invasões em sistemas e dispositivos IoT, definir padrões de acesso e quais métodos foram abordados por invasores e, assim, planejar defesas para esses propósitos e relatar possíveis brechas para invasão.

2.1 Fundamentação teórica

Nos tópicos subsequentes, serão abordados os conceitos fundamentais que regem a temática de *honeypots* para IoT com o intuito de garantir a compreensão das tecnologias e ferramentas relevantes ao trabalho.

2.1.1 Segurança cibernética

Com o processo de inovação da comunicação e informação, houve a necessidade do aprofundamento dos conceitos de segurança de dados. A evolução do setor tecnológico trouxe novas vulnerabilidades a serem exploradas por atacantes (NGUYEN; REDDI, 2019). Ferramentas de segurança assumiram um espaço de destaque para companhias, pessoas e governos, sendo alvo de investimento para garantir a segurança na rede (SOLMS; NIEKERK, 2013).

O campo da segurança da informação é fundamentado por três importantes elementos, sendo eles: confidencialidade, integridade e disponibilidade dos dados. Estes conceitos compõem o que Whitman e Mattord (2021) traduz como triângulo de CIA (Confidencialidade, Integridade e Disponibilidade, do inglês: *Confidentiality, Integrity, Availability*). Entretanto, o conceito atual de segurança cibernética abrange novas perspectivas que não estão contempladas nessa abordagem tradicional, como autenticidade, responsabilidade, privacidade e resiliência, refletindo as necessidades de um ambiente digital cada vez mais complexo:

- **Confidencialidade**: Aborda o princípio de que dados devem ser assegurados contra a divulgação sem autorização e o descumprimento deste causa a quebra de confi-

dencialidade.

- **Integridade**: Se baseia no conceito da completude do dado. As informações estão sujeitas a se corromperem durante o transporte ou armazenamento, neste cenário acontece a corrupção dos dados.
- **Disponibilidade**: Garante que o acesso à recursos aconteça sem obstrução e no modelo requisitado pelo usuário ou sistema.

O conteúdo que classifica a segurança da informação pode ser reaproveitado para a segurança cibernética para a maior parte dos riscos em rede (SOLMS; NIEKERK, 2013). De acordo com Li e Liu (2021), as redes de computadores são sujeitas tanto a funções ativas do usuário quanto ao roubo de informações por um invasor.

O mundo cibernético se tornou um espaço integrado para comunicação e controle do mundo físico (LI; LIU, 2021) e este tipo de situação coloca os elementos do mundo real em perigo. Diversas informações circulam pela internet todos os dias, sejam elas fotos, vídeos, cadastros em novos sites e aplicativos; estes conteúdos alimentam diversos bancos de dados e se perdem entre os processos da grande rede de computadores. Não são incomuns notícias de grandes vazamentos de informação de empresas com alto valor de mercado, como foi o caso do Twitter em 2023 onde dados de mais de 200 milhões de usuários foram alvos de um ataque cibernético (Forbes, 2023).

A segurança cibernética tem como papel principal adotar medidas práticas para assegurar a proteção de informação, dados e rede (LI; LIU, 2021) além de assegurar a proteção de usuários e instituições sujeitas a práticas criminosas e de violência contra indivíduos, assim como apontado por (SOLMS; NIEKERK, 2013). Para cumprir com essas ponderações, alguns serviços são abordados e aplicados no âmbito virtual:

- **Segurança de Redes**: Possui a premissa de garantir a segurança da conexão contra elementos diruptores adotando uma série de soluções tomadas para se afastar de hackers em potencial (LI; LIU, 2021).
- **Segurança de Aplicações**: Garante a segurança a partir de ferramentas externas ou internas contra operações indesejadas (LI; LIU, 2021), como por exemplo sistemas de login, autenticadores, criptografia de mensagens e senhas, entre outros.
- **Segurança de Informações**: Protege dados contra acesso não autorizado, corrupção ou exclusão de conteúdo (WHITMAN; MATTORD, 2021).
- **Segurança Operacional**: Trabalha com as operações e funções que gerenciam os acessos de conteúdo (LI; LIU, 2021).

- **Segurança em Nuvem**: Aplicações em nuvem são hospedadas em ambientes virtuais e distribuídas para diversos usuários simultaneamente, assim agregando maiores riscos e possíveis vulnerabilidades (SINGH; CHATTERJEE, 2017).

Além destes aspectos, a capacitação de usuários para identificar e lidar com conteúdos suspeitos inseridos no cotidiano é de extrema importância (FURMAN et al., 2012; LI; LIU, 2021). Existem iniciativas dentro de empresas para o treinamento de funcionários com o intuito de garantir uma navegação mais segura e reduzir a chance de comprometer suas informações sensíveis; porém Furman et al. (2012), em seu trabalho, identifica uma baixa aderência de usuários.

No contexto de IoTs, a proteção destes dispositivos assume maiores proporções, principalmente pelo fato de sua disponibilidade e conectividade. A invasão por meio desse tipo de dispositivo, associando-se aos cenários de casas inteligentes onde há uma maior facilidade de integração e conexão desses dispositivos, tem maiores capacidades de causar dano às suas vítimas (LI; LIU, 2021).

2.1.2 Dispositivos IoT

De acordo com Madakam, Ramaswamy e Tripathi (2015), devido à grande diversidade de áreas que abordam os conceitos de Internet das Coisas, não há um consenso para a definição de IoT; sendo assim, o autor trabalha com um conceito geral, onde os dispositivos atuam como uma rede aberta autorreguladora, onde seus elementos estabelecem comunicação entre si, compartilhando dados e recursos, e estão sujeitos a alterações do ambiente em que estão inseridos.

Conforme abordado por Ashton (2009) em uma de suas apresentações, o termo "*Internet of Things*" tomou nota em 1999 ao relacionar as noções de internet ao uso de tecnologia de identificação por Rádio-Frequência (RFID) na cadeia de suprimentos da companhia Procter & Gamble (P&G). A abordagem da comunicação entre máquinas não é um conceito novo, o IoT representa a evolução dessa interação, alinhando as interconexões de dispositivos através da internet. Essas tecnologias combinam o uso de sensores com programas e serviços para melhor interação com o usuário a partir da união de hardwares, softwares, suas arquiteturas e aplicações (WHITMORE; AGARWAL; XU, 2015):

- **Hardware**: A construção de um dispositivo IoT deve ser esquematizada com base na função que o mesmo deverá cumprir e a partir destes propósitos, sensores e peças são definidos para melhor atende-las. Diversos elementos e marcas constituem o mercado de hardware e com isso, estão sujeitos a agregar vários projetos e esquemas de dispositivos IoT. Dentre as tecnologias mais comuns, o **RFID** se destaca devido à sua aplicabilidade em diversas tecnologias populares. Seu uso esta relacionado a identificação de objetos, gravação de metadados e controle de dispositivos

via ondas de rádio devido a possibilidade de comunicação de curta distância (JIA et al., 2012). Similar a este, a tecnologia de **Comunicação por Proximidade de Campo** (NFC) possui o mesmo propósito. Utilizado em *smartphones* para o sistema de pagamentos, impressoras e sistemas de empréstimo de bibliotecas, sua tecnologia é atrelada a uma identificação única associada ao seu rótulo (WHITMORE; AGARWAL; XU, 2015). **Micro-controladores** também fazem parte deste conjunto de hardware; são responsáveis por conectar e atribuir a programação para o restantes dos sensores; sendo o Arduino e Rasberry Pi os dispositivos predominantes dessa categoria (AMAN et al., 2020; AL-FUQAHA et al., 2015). Sensores de modo geral são utilizados para avaliar condições de ambientes e ao serem associados diretamente, são referenciados como **Rede de Sensores Sem Fio** (WSN). Por meio dessas redes de sensores, a comunicação entre usuários e dispositivos passam a ser mais dinâmicas e permite que ações sejam tomadas com base em situações identificadas por parte do próprio dispositivo, como por exemplo a ativação de um sistema de incêndio onde a ferramenta emite um alerta sonoro ou o monitoramento de pacientes em hospitais com sistema inteligente onde emite alertas caso os indicativos de pressão e temperatura sofram alterações significativas (WHITMORE; AGARWAL; XU, 2015; AL-FUQAHA et al., 2015):

- **Software**: Para suportar novas ferramentas atuando de forma individual ou em conjunto, é necessário se atentar à programação dos dispositivos. Com a diversidade de sensores, métodos de comunicação, coleta de dados e linguagens de programação, o programa deve ser sempre revisado e atualizado. O conceito de software está inserido dentro da classificação de *middleware*; este atua entre as camadas do *hardware*, dados e aplicações, auxiliando na gestão dos dados e a criação e entrega de novos serviços suportados pelo IoT sem a necessidade de criar novos códigos (WHITMORE; AGARWAL; XU, 2015). Dentro do sistema de IoT, o software é de importante destaque pois é a partir dele que o processamento de dados, funcionalidades do dispositivo e segurança são abordados , atuando nessa gestão e se conectando as fontes externas de nuvem (MOCRII; CHEN; MUSILEK, 2018; AMAN et al., 2020).

- **Arquitetura**: De acordo com Aman et al. (2020), aparelhos IoT podem assumir configurações complexas e amplas devido à sua heterogeneidade e escalabilidade. Para providenciar uma melhor otimização, a arquitetura é trabalhada em camadas distintas para suas tecnologias e protocolos de comunicação. Diversos trabalhos utilizam conceitos diferentes para abordar as camadas, porém é possível agrupá-las em três principais categorias, elaboradas a partir dos protocolos e funções que regem cada uma delas. A **Camada Alta** é composta por funções de requisição do próprio usuário e lida com a comunicação direta entre o dispositivo e o usuário. Protocolos como o *Message Queue Telemetry Transport* (MQTT) são comuns nesta

camada. A **Camada Média** está relacionada à rede, sendo responsável por prover conectividade e serviços às outras duas camadas. Por fim, a **Camada Baixa** compreende as funções de hardware, atuando como uma ponte entre os sensores físicos e a disponibilização digital dos dados mensurados (AMAN et al., 2020).

- **Aplicação**: A percepção de que ferramentas inteligentes fazem parte da realidade e a possibilidade da conexão entre elas permitem a elaboração de diversas aplicações (ABDMEZIEM; TANDJAOUI, 2014; FAROOQ et al., 2015). Estes recursos, somados à capacidade de garantir a autonomia, são elementos importantes para o desenvolvimento de novas ferramentas e para o aprimoramento da qualidade de vida (FAROOQ et al., 2015). IoTs podem ser divididos em categorias de aplicação que atendem a propósitos específicos e interagem entre si, como ilustrado na Figura 1 (AMAN et al., 2020). Entre essas categorias destacam-se: o **auto cuidado**, com dispositivos voltados à saúde, como relógios inteligentes e sistemas de monitoramento de pacientes com a função de avaliar as condições relacionadas à saúde do usuário; o **transporte**, com aplicações embarcadas que auxiliam motoristas com condução inteligente, ajudando em manobras e no sistema de rotas; o **ambiente**, com dispositivos para monitoramento climático e gestão de recursos, garantindo melhor eficiência e redução de gastos na manutenção de fazendas inteligentes; as **cidades inteligentes**, com o objetivo de conectar as funcionalidades recorrentes das rotinas de usuários, auxiliando nas interações entre seus pares e serviços; e a **indústria inteligente**, com ferramentas voltadas à gestão de fluxos extensos de atividades com entregas programadas com o objetivo de um melhor controle de qualidade, automação de processos e gestão eficiente da produção.

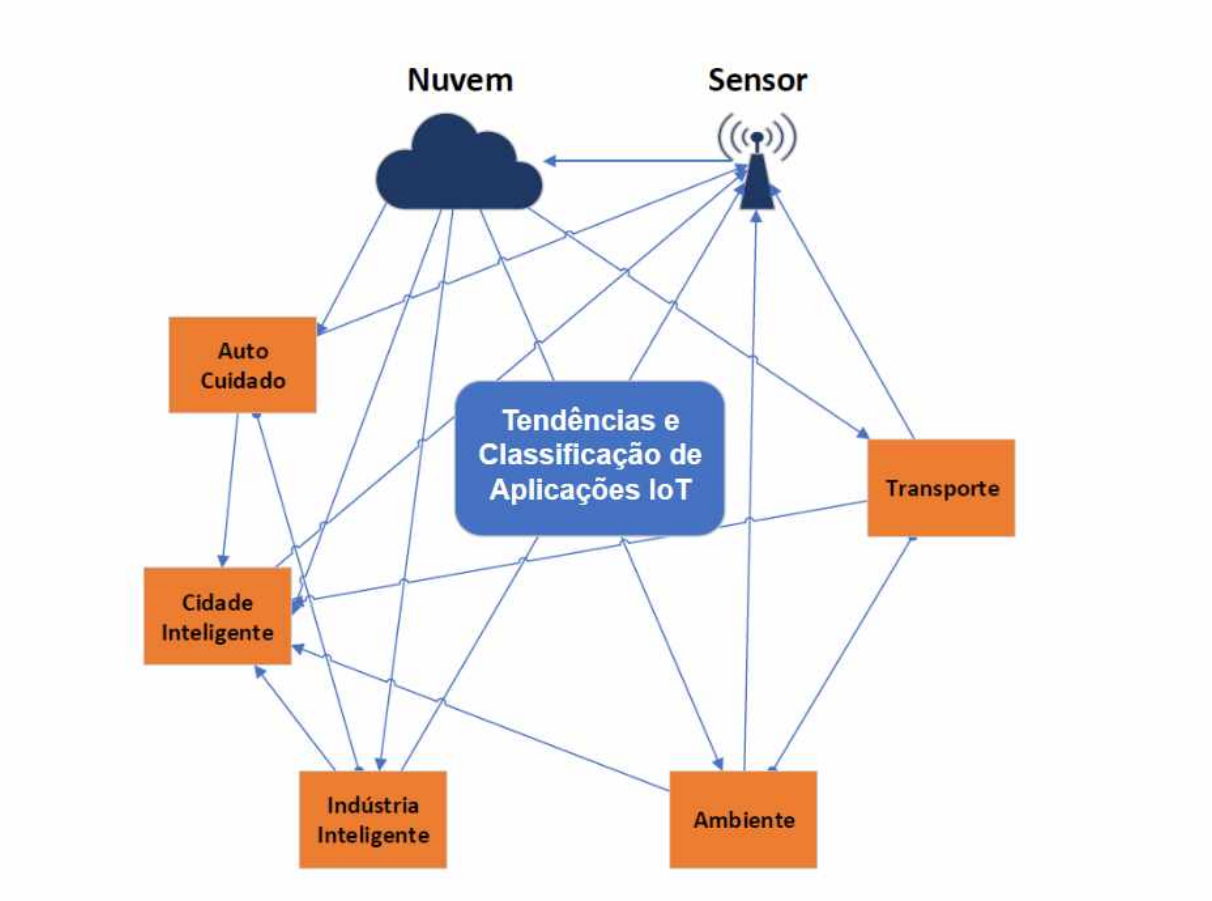

Figura 1 – Interações entre categorias IoT. Fonte: Adaptada de Aman et al. (2020)

2.1.3 Honeypot

Stoll (2005) foi um dos precursores do conceito de utilizar ferramentas para investigar acessos indevidos em sistemas de computadores. Seu livro apresenta o processo de identificação e análise de uma invasão de um atacante russo a um sistema do Laboratório Lawrence Berkeley durante o período da Guerra Fria. Além disso, Cheswick (1992) apresenta outro caso semelhante, no qual interpelou um ataque cibernético em um ambiente controlado, rastreando todas as ações dos invasores. Essas ações foram registradas e analisadas, permitindo um estudo detalhado do ataque e possibilitando seu controle.

Honeypots são *Deception Tools*, ou "Ferramentas Enganosas", que simulam sistemas operacionais ou sistemas de dispositivos IoT conforme ilustrado de forma simplificada na Figura 2. Seu principal objetivo é atrair invasores para avaliar seus comportamentos (SRINIVASA; PEDERSEN; VASILOMANOLAKIS, 2022). Para que a ferramenta cumpra seu papel de forma eficaz e atrativa, é essencial que, no momento de sua implementação, sua exposição em ferramentas de busca seja evitada (ACIEN et al., 2018).

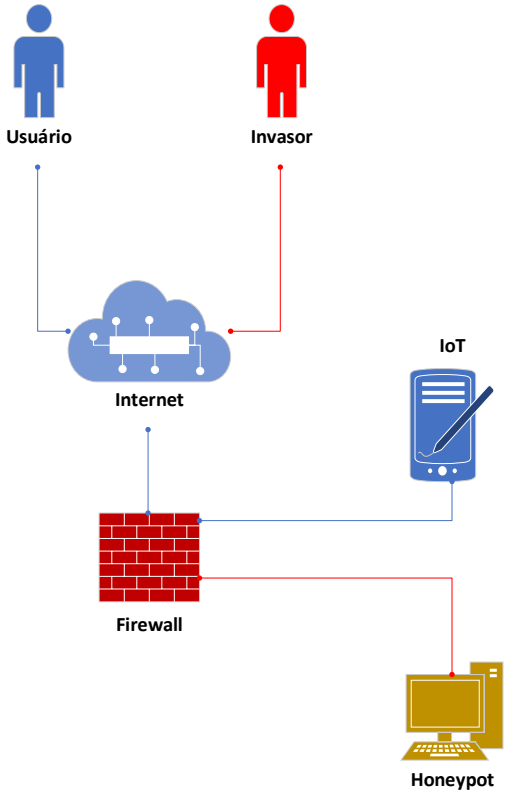

Figura 2 – Arquitetura simplificada de *honeypot*. Fonte: Adaptada de Franco et al. (2021)

Com base no trabalho de Oza, Kumar e Khorajiya (2018), Franco et al. (2021), os *honeypots* podem ser divididos em três gerações:

- **Geração I**: Desenvolvidos em 1999, estes *honeypots* possuem uma arquitetura simples porem com uma grande capacidade de aprofundar, coletar informações e identificar novos ataques, no entanto, os mesmos são fáceis de serem identificados pelos atacantes (OZA; KUMAR; KHORAJIYA, 2018; FRANCO et al., 2021).
- **Geração II**:Como uma evolução da geração anterior e com o objetivo de corrigir problemas anteriores, desenvolvida em 2002, a geração II implementa sensores que combinam as funcionalidades de sistemas de detecção de intrusão (IDS) com o firewall da primeira geração (OZA; KUMAR; KHORAJIYA, 2018). Além disso, os sensores atuam como uma ponte, dificultando a detecção do *honeypot* pelo lado do atacante (OZA; KUMAR; KHORAJIYA, 2018; FRANCO et al., 2021).
- **Geração III**: em 2004 a geração III foi desenvolvida. Sua arquitetura é exatamente a mesma da geração anterior, porem com melhorias nos conceitos de implementação e gestão da ferramenta (OZA; KUMAR; KHORAJIYA, 2018; FRANCO et al., 2021).

O processo de avaliação de invasões em sistemas *honeypots* é fundamental para gerenciar os tipos de ataques aos quais dispositivos reais estão sujeitos. Nesse contexto, conforme destacado por Franco et al. (2021), o controle, a captura e o armazenamento de dados são elementos primordiais de qualquer ferramenta *honeypot*. A partir desses aspectos, é possível monitorar o progresso das invasões, garantindo que o atacante não perceba a emulação, que todos os dados gerados pelo ataque sejam coletados e que essas informações sejam transferidas de forma segura para um banco centralizado.

2.2 Estado da arte

Este tópico reúne diversos trabalhos relacionados ao desenvolvimento de ferramentas *honeypot* e ao processo de análise dos dados coletados. Com a constante evolução das técnicas de invasão, investir em estratégias para identificar e corrigir vulnerabilidades nesses dispositivos tornou-se essencial para seu bom funcionamento. A análise dos dados coletados possui um papel fundamental nesse processo, identificando padrões de ataques e a criação de perfis de comportamento dos invasores. Desta forma, é possível aprimorar medidas e estratégias de segurança contra ameaças, reforçando a segurança destes dispositivos.

Durante a década de 2010, com todos os avanços da tecnologia e com a disposição de novos dispositivos IoT conectados à rede, foi identificada a necessidade de se trabalhar com a segurança destes aparelhos baseando-se em seus diversos protocolos e serviços. Como proposto por Pa et al. (2016)e seu time através do IoTPOT, ferramenta de baixa-interação que atua em paralelo ao IoTBOX que executa o papel de alta-interação simulando o serviço Telnet. A partir deste trabalho foi identificado um crescimento de ataques no ano de 2014, além de apontar 5 famílias de *malware* utilizados ativamente para ataques DDoS (PA et al., 2016).

Srinivasa, Pedersen e Vasilomanolakis (2022) aborda a discussão sobre os níveis de interação em ferramentas *honeypot* e propõe o RIoTPot, um *honeypot* de interação híbrida que oferece maior facilidade de implementação e gerenciamento para os moderadores. O estudo avalia os diferentes níveis de interação e observa que, em ambientes de baixa interação, há uma redução gradual no número de ataques em determinados protocolos, enquanto os *honeypots* de alta interação tendem a atrair ataques mais sofisticados.

Além disso, o RIoTPot configurado no modo de alta-interação foi comparado diretamente com uma ferramenta de média-interação, e os resultados demonstraram que o *honeypot* desenvolvido apresentou um desempenho superior em relação à alternativa analisada (SRINIVASA; PEDERSEN; VASILOMANOLAKIS, 2022). O estudo também apresenta métricas que correlacionam a distribuição percentual dos tipos de ataques com os protocolos explorados no RIoTPot, identificando que a maioria das ameaças corres-

ponde a ataques de força bruta, *port scans* e requisições suspeitas.Os protocolos mais frequentemente explorados por *malwares*, segundo os dados analisados, foram Telnet e SSH.

O uso de aprendizado de máquina pode ser incorporado às ferramentas *honeypot*, como observado por Luo et al. (2017). Relacionando essa abordagem ao uso de um *chatbot*, Mfogo et al. (2023) desenvolveram o AIIPot, uma ferramenta projetada para administrar a etapa de checagem, na qual o atacante envia suas requisições de forma indireta ao *chatbot*. Simulando um ambiente de comando e gerenciando as interações realizadas pelo invasor, o sistema busca coordenar o tipo de dispositivo simulado e tenta manter o atacante na sessão. Após a avaliação dos resultados, observou-se uma redução no número de requisições ao longo do tempo, especialmente em interações com sessões de curta duração (inferiores a sete interações). A partir dessas métricas, levantaram-se suspeitas sobre a possível detecção da ferramenta pelos atacantes, fazendo com que abandonassem suas investidas, ou um possível comprometimento do banco de dados utilizado para o treinamento do modelo, afetando a assertividade das respostas geradas pela ferramenta.

Focando na etapa de análise, Thakar, Varma e Ramani (2005) desenvolveu uma ferramenta para auxiliar administradores na definição de assinaturas precisas de tráfegos maliciosos, utilizando um *honeypot* de baixa-interação chamado Honeyd. A solução baseia-se na coleta de *logs* extraídos do *honeypot*, somados a registros de tráfego de rede capturados pela ferramenta de monitoramento de rede `Tcpdump`. Esses dados são então armazenados em uma base unificada, enquanto uma interface web os consome para exibi-los em gráficos, permitindo uma leitura mais intuitiva e detalhada dos padrões de tráfego malicioso, facilitando a identificação de ameaças. Além disso, ao comparar a HoneyAnalyzer com o Honeycomb, outra ferramenta que utiliza metodologia semelhante, Thakar, Varma e Ramani (2005) demonstraram que sua solução apresentou resultados de maior qualidade e um processo de detecção de invasão mais preciso, reduzindo as redundâncias e os falsos positivos/negativos.

Com o objetivo de processar grandes volumes de dados, Sobesto et al. (2011) desenvolveu a ferramenta DarkNOC a partir de uma *honeynet* com uma rede de 2.000 *honeypots*. A interface da ferramenta organiza e correlaciona diferentes atividades das ferramentas em rede, utilizando múltiplas fontes de dados, como registros de fluxo de rede (*NetFlow*), *malwares* coletados pelo Nepenthes (*honeypots* de baixa-interação) e ataques detectados pelo sistema de detecção de intrusão Snort. Como resultado, a DarkNOC se destaca por sua capacidade de fornecer informações detalhadas sobre redes e dispositivos *honeypots* comprometidos.

3 Desenvolvimento

Neste Capítulo serão apresentados os detalhes das etapas de desenvolvimento da ferramenta para o tratamento e análise dos *logs* gerados pelos dispositivos *honeypots*. A Figura 3 apresenta o fluxo do processo que abrange a coleta e organização dos arquivos pelos *honeypots*, bem como os recursos e bibliotecas utilizados para a gestão dos dados e sua representação gráfica, demonstrando a evolução do método de investigação.

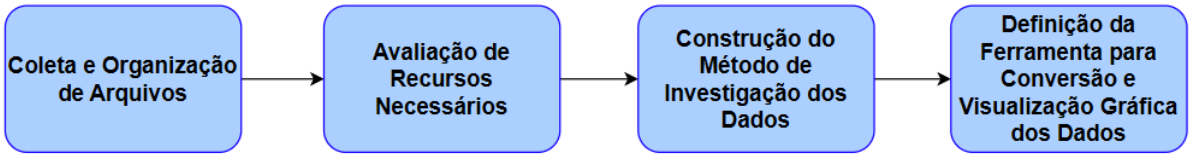

Figura 3 – Fluxograma de desenvolvimento da ferramenta. Fonte: Do Autor

3.1 Coleta e organização de arquivos

O processo de coleta de dados se iniciou em 2023 dentro do NuSec com o "Experimento - Roteadores" , que, em seu desenvolvimento, foram simuladas seis diferentes interfaces web de roteadores *honeypots*, totalizando 12 dispositivos em diferentes localidades e, durante um período de 15 dias, foram coletados dados gerados pelos acessos de usuários atacantes. Os dados são organizados por data pelo próprio servidor Apache, sendo assim, os mesmos são divididos pelos dias de experimento em arquivos *.log* e compactados, como abordado na Figura 4. Cada linha de informação registra uma requisição e seu conteúdo possui informações de endereço IP, data e hora, método HTTP, URL explorada, resposta do servidor e o *user agents* que ajuda a identificar o dispositivo ou navegador utilizado no acesso. Abaixo, um exemplo de requisição obtida do "Experimento - Roteadores":

```
192.227.173.18 - - [09/Sep/2023:14:58:03 +0000] "GET /.env
HTTP/1.1"404 437 -Python/3.7 aiohttp/3.7.4.post0"
```

Posteriormente, outro trabalho desenvolvido no NuSec teve como objetivo replicar a metodologia abordada por Mendes (2023), porém utilizando interfaces de câmeras para investigar o comportamento de atacantes ao interagir com esses dispositivos emulados. Os dados obtidos do "Experimento - Câmeras" (COSTA, 2025) seguem a mesma estrutura do experimento anterior e, em ambos os estudos, os arquivos possuem o mesmo padrão

organizacional, sendo armazenados em arquivos *.logs* e compactados pelo próprio servidor Apache, onde essas ferramentas foram hospedadas.

O diretório destinado ao gerenciamento dos *logs* armazena os diferentes arquivos compactados, conforme apresentado na Figura 4, sendo necessária sua extração e organização para se administrar de forma eficiente. Com o objetivo de otimizar essa atividade, é essencial uma pré-organização, garantindo um único padrão no diretório dos *honeypots* assim como na Figura 5. A ferramenta desenvolvida possui funções que realizam a leitura deste diretório, a extração dos arquivos necessários e os padronizam em uma nova subpasta, garantindo a integridade dos arquivos originais para eventuais auditorias, caso sejam necessárias. Dessa forma, a primeira etapa do processo de construção e análise usa como base a cópia corrigida dos arquivos *logs* disponibilizados.

Nome	Tipo	Tamanho Compacta...	Protegido ...	Tamanho	Razão	Data de modificação
access.log	Documento de Texto	4 KB	Não	23 KB	86%	28/10/2023 16:36
access.log.1	Arquivo 1	6 KB	Não	74 KB	93%	27/10/2023 23:58
access.log.2.gz	Pasta de Arquivo Compac...	4 KB	Não	4 KB	0%	26/10/2023 23:43
access.log.3.gz	Pasta de Arquivo Compac...	6 KB	Não	6 KB	0%	25/10/2023 23:42
access.log.4.gz	Pasta de Arquivo Compac...	4 KB	Não	4 KB	0%	24/10/2023 23:57
access.log.5.gz	Pasta de Arquivo Compac...	5 KB	Não	5 KB	0%	23/10/2023 23:59
access.log.6.gz	Pasta de Arquivo Compac...	5 KB	Não	5 KB	0%	22/10/2023 23:50
access.log.7.gz	Pasta de Arquivo Compac...	6 KB	Não	6 KB	0%	21/10/2023 23:49
access.log.8.gz	Pasta de Arquivo Compac...	5 KB	Não	5 KB	0%	20/10/2023 23:58
access.log.9.gz	Pasta de Arquivo Compac...	4 KB	Não	4 KB	0%	19/10/2023 23:49
access.log.10.gz	Pasta de Arquivo Compac...	5 KB	Não	5 KB	0%	18/10/2023 23:32
access.log.11.gz	Pasta de Arquivo Compac...	5 KB	Não	5 KB	0%	17/10/2023 23:14
access.log.12.gz	Pasta de Arquivo Compac...	10 KB	Não	10 KB	0%	17/10/2023 00:00
access.log.13.gz	Pasta de Arquivo Compac...	6 KB	Não	6 KB	0%	15/10/2023 23:59
access.log.14.gz	Pasta de Arquivo Compac...	7 KB	Não	7 KB	0%	14/10/2023 23:26
error.log	Documento de Texto	1 KB	Não	1 KB	50%	28/10/2023 11:29

Figura 4 – Exemplo de arquivos *log* resgatados do servidor apache. Fonte: Do Autor

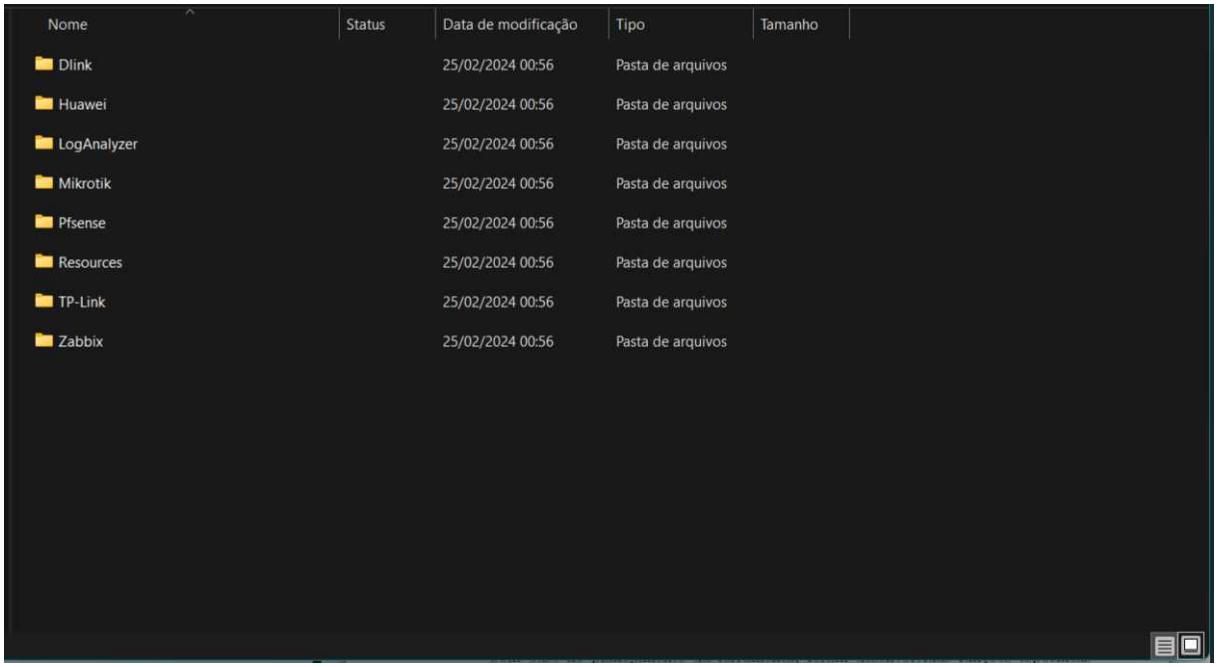

Figura 5 – Pré-organização das pastas de *logs* dos *honeypots*. Fonte: Do Autor

3.2 Avaliação de recursos necessários

Após a organização dos arquivos, a etapa seguinte consistiu na definição dos recursos a serem utilizados para a construção e execução dos métodos de análise. Considerando o extenso conjunto de dados e a possibilidade de crescimento da ferramenta, optou-se pelo uso da linguagem de programação Python, devido ao seu melhor suporte para bibliotecas especializadas e à sua facilidade de desenvolvimento. Além disso, o Python conta com uma ampla comunidade, sendo uma das linguagens mais populares para análise de dados (MCKINNEY, 2018).

Para garantir o bom funcionamento das funções desenvolvidas e otimizar a gestão dos dados, diversas bibliotecas foram exploradas ao longo do trabalho. Entre elas, recursos da biblioteca padrão do Python, como os módulos "*os*", "*re*" e "*gzip*", foram utilizados com o propósito de acessar e manipular arquivos e diretórios. O módulo "*os*" facilita a navegação e gerenciamento de diretórios, interagindo diretamente com o sistema operacional. Já o módulo "*gzip*" oferece suporte para abrir e ler arquivos no formato GZIP. Em conjunto, esses dois módulos desempenham um papel crucial no processo inicial de organização e transformação dos arquivos *log*, tornando as operações subsequentes mais eficientes.

Simultaneamente, o módulo "*re*" desempenha um papel menor, porém fundamental, na etapa inicial de gestão dos arquivos. Este módulo possibilita o trabalho com expressões regulares e a manipulação de texto presente nos *logs*, permitindo a procura por

padrões dentro dos arquivos e a construção de um dicionário estruturado com seus atributos nomeados (IP, Data, Método, URL, Status, Tamanho, *User-Agent*). Esse dicionário pode, então, ser convertido em um *DataFrame*, que se torna o principal elemento nas análises subsequentes.

Acerca das funcionalidades essenciais para a evolução da ferramenta, destacam-se os pacotes responsáveis por administrar diretamente os dados, como a biblioteca "*pandas*" na versão 2.1.0 que, de maneira rápida e flexível, permite a manipulação de conjuntos de informações extraídas dos *logs*. Adicionalmente, o pacote "*Matplotlib*", na versão 3.8.0, foi escolhido como um método simples, porém eficaz, para explorar esses dados em gráficos e facilitar a compreensão do observador.

3.3 Construção do método de investigação dos dados

Como abordado na Seção 3.2, a ferramenta opera com a biblioteca "*Pandas*" para proporcionar uma abordagem mais dinâmica e intuitiva no tratamento dos dados por meio de *DataFrames*. Dessa forma, foram desenvolvidas funções específicas para realizar diferentes tipos de leitura, permitindo tanto a análise de métricas individuais quanto o cruzamento de informações com outros elementos, possibilitando a identificação de padrões de comportamento adotados por atacantes ao longo do período de coleta dos dados pelos *honeypots*.

Os dados foram organizados em oito atributos distintos, sendo sete extraídos diretamente dos *logs* — *IP, data, método, URL, status, tamanho* e *User-Agent* — e um adicional, *fonte*, incorporado ao conjunto, permitindo assim a identificação do *honeypot* de origem de cada entrada e garantindo uma gestão mais eficiente das informações.

Cada um desses atributos apresenta características qualitativas que descrevem as interações entre os atacantes e os *honeypots*, permitindo a análise de padrões de comportamento. Além disso, para as análises quantitativas, foram consideradas métricas como a frequência de acessos, a repetição de ataques, o volume de interações ao longo do período de atividade das ferramentas, entre outros fatores que auxiliam na identificação de tendências e padrões de ataque. A seguir, os atributos presentes no *DataFrame* são detalhados:

- **IP**: Endereço de origem do usuário que realizou a requisição ao *honeypot*.
- **Data**: Registro do momento da interação, incluindo dia e hora.
- **Método**: Método HTTP utilizado na requisição, como *GET*, *POST*, *PUT* ou *DELETE*.

- **URL**: Caminho solicitado pelo atacante na requisição, indicando a página ou recurso acessado.
- **Status**: Código de resposta HTTP retornado pelo servidor, representando o resultado da requisição.
- **Tamanho**: Quantidade de dados (em bytes) transferida na resposta do servidor.
- ***User-Agent***: Informação de registro que identifica o cliente que acessa a ferramenta.
- **Fonte**: Nome do *honeypot* responsável por capturar a entrada, definido com base no diretório onde os *logs* estão armazenados.

A partir desses atributos, foram definidas funções que atuam como filtros para organizar os dados de acordo com as necessidades de sua representação gráfica, facilitando a visualização e análise. Com o objetivo de investigar quais elementos os atacantes buscavam acessar e quais comandos tentaram executar, o campo URL recebe destaque, pois permite entender qual era o verdadeiro objetivo do agente invasor.

3.4 Definição da ferramenta para conversão e visualização gráfica dos dados

A fim de garantir uma análise eficiente dos dados por meio de gráficos, a ferramenta foi projetada a partir de um escopo inicial focado na tradução dos dados extraídos dos *honeypots* e na exibição de métricas fundamentais, como a quantidade de acessos diários, o número de entradas por *honeypot* e os principais IPs de atacantes. Para proporcionar maior dinamismo e flexibilidade, tanto as etapas de extração e tratamento dos dados quanto a execução das análises são abordadas de forma simultânea, reduzindo a necessidade de múltiplos códigos para a realização dessas atividades.

Durante o planejamento da ferramenta, algumas funções básicas foram definidas para possibilitar uma exploração objetiva de métricas importantes na compreensão das metas dos atacantes. Este conjunto de funções participa diretamente do fluxo de execução da ferramenta, como ilustrado na Figura 6. Parte dessas funções foi desenvolvida para converter os *logs* em *DataFrames*, com o propósito de permitir que esses dados sejam carregados individualmente ou simultaneamente para comparações pelas funções de visualização. Os diagramas da Figura 7 abordam as atividades presentes no processo de tratamento dos arquivos.

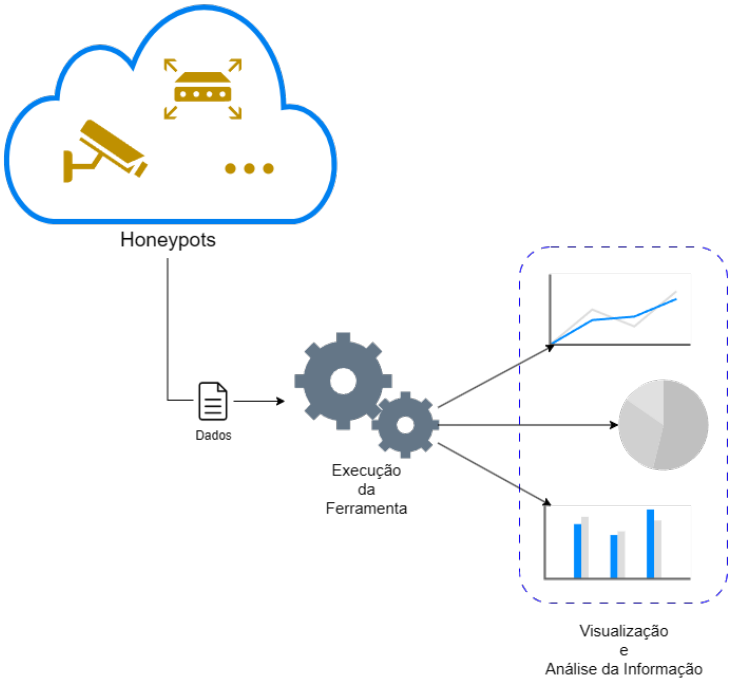

Figura 6 – Diagrama de funcionamento da ferramenta. Fonte: Do Autor

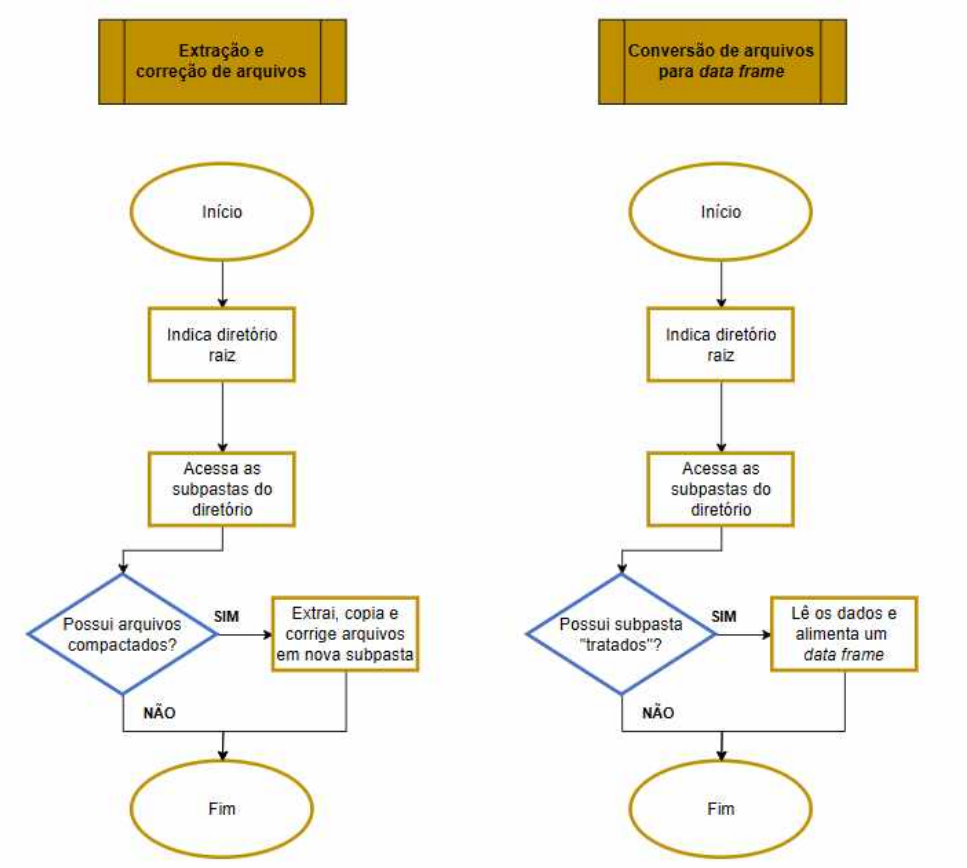

Figura 7 – Diagramas de atividade para processo de extração, correção e conversão de arquivos. Fonte: Do Autor

O segundo grupo de funções é responsável pela exibição visual dos dados. A biblioteca *Matplotlib* foi incorporada à ferramenta, viabilizando a construção e exposição dos gráficos diretamente na estrutura dos *notebooks* Python, onde as funções são carregadas e executadas. Dessa forma, os visuais podem ser facilmente ajustados aos propósitos e necessidades do usuário e, a partir dessa estrutura, a ferramenta pode ser expandida para atender a requisitos específicos. De modo geral, os casos de uso das funções seguem um mesmo padrão, diferenciando-se apenas pelos filtros aplicados e pelo tipo de gráfico gerado. A seguir, é apresentado o fluxo básico e alternativo de execução:

- **Descrição geral**
  Funções responsáveis por gerar gráficos com base nos dados coletados e organizados em *DataFrames*. O fluxo básico refere-se à execução com apenas um *DataFrame*; já o fluxo alternativo utiliza dois conjuntos de dados distintos.

- **Fluxo básico de execução**

  1. O usuário indica o *DataFrame* como parâmetro;
  2. A função filtra os dados de acordo com sua descrição (por exemplo: *acesso_dia*, *acesso_fonte*, *top_ips*, *hp_dia* e *ip_dia*);
  3. Os dados são agrupados, classificados e organizados para visualização;
  4. O gráfico é gerado e exibido ao usuário.

- **Fluxo alternativo de execução**

  1. O usuário indica dois *DataFrames* como parâmetro;
  2. A função filtra e ajusta os dados com base na necessidade dos conjuntos (por exemplo: identificação de IPs em comum);
  3. Os dados de ambas as fontes são agrupados, classificados e organizados para visualização conjunta;
  4. Os gráficos são gerados de forma concorrente ou sobrepostos, e exibidos ao usuário.

4 Resultados

Este Capítulo tem como propósito apresentar as funções da ferramenta e a aplicação do método de tratamento dos dados coletados no "Experimento - Roteadores" e no "Experimento - Câmeras", além de realizar um comparativo entre as duas bases de dados e descrever seu estado atual de desenvolvimento. A partir da execução da ferramenta, podemos visualizar os resultados esperados durante a etapa de desenvolvimento do trabalho, sendo favorável à identificação de possíveis melhorias que serão alvo de discussão do Capítulo 5. A ferramenta, assim como as orientações para sua execução, está disponível em um repositório online [1].

4.1 Extração e tratamento dos dados

Conforme abordado na Seção 3.1, os dados utilizados como base para a construção da ferramenta são resultado do "Experimento - Roteadores" e do "Experimento - Câmeras". Em ambos os experimentos, os *honeypots* foram instanciados em localidades distintas e, inicialmente, no caso do "Experimento - Roteadores", observou-se que os invasores demonstraram interesse pelo campo de URL das ferramentas — comportamento que também foi identificado no "Experimento - Câmeras". A Tabela 1 apresenta métricas comuns entre os dois experimentos.

Categoria	**Experimento - Roteadores**	**Experimento - Câmeras**
Duração	15 dias	15 dias
Número de Máquinas Virtuais	12	12
Número de Interfaces	6	6
Total de Requisições	48.719	48.494

Tabela 1 – Comparação entre os experimentos com roteadores e câmeras. Fonte: Extraído de (MENDES, 2023)

Com base no planejamento da ferramenta, foram desenvolvidas funções específicas que operam a partir do endereço do diretório como parâmetro, onde estão armazenadas as pastas dos *honeypots* com seus respectivos *logs*. A função principal, "*log_df*", inicia o processo corrigindo os tipos de arquivos, extraindo os necessários e armazenando-os no formato *.log* em uma nova subpasta, garantindo a padronização e organização dos dados. Essa função contém várias outras funções auxiliares que atuam em diferentes processos durante as etapas de execução. A partir dessa configuração, cada um desses *logs* é lido e

[1] Disponível em: https://github.com/Pe-Santana/Analise-Honeypot

seus dados são adicionados a um *DataFrame*, conforme a estrutura definida na Seção 3.3 do capítulo anterior.

Durante o processo de gestão das informações, os atributos da base de dados são ajustados para garantir que cada um adote o tipo de dado mais adequado para as análises exploratórias. Dentre eles, o campo *Data* recebeu atenção especial, devido à sua diversidade de formatos que podem ser registrados. Como os *honeypots* capturam entradas continuamente enquanto estão ativos, foi definido por manter detalhes como horas e minutos, uma vez que essas informações possuem grande valor para análises mais precisas e específicas, assim permitindo o estudo de padrões temporais no comportamento dos atacantes.

4.2 Visualização e exploração dos dados na ferramenta

A etapa de exploração dos dados possibilitou a experimentação de diferentes métricas para responder a questionamentos como o número de acessos por dia dos experimentos, a divisão dos acessos para cada *honeypot*, os principais endereços IP que interagiram com os *honeypots*, o comportamento dos *honeypots* baseado no total de acessos por dia em cada um deles e o comportamento de um endereço IP especificado pelo usuário para estudar seu comportamento, que podem ser observados nas Figuras 8, 9, 10, 11, 12, 13, 14 e 15. Para abordar cada um desses questionamentos, foram desenvolvidas funções específicas com base nas estruturas discutidas na Seção 3.4, as quais, de modo geral, recebem como parâmetro os *DataFrames*, filtram e ajustam os dados e, por fim, disponibilizam-nos como gráficos. A seguir, são detalhadas as funções desenvolvidas até o momento da publicação deste trabalho:

- **acesso_dia**: Calcula a frequência de requisições para cada dia do experimento. Os dados são exibidos em um gráfico de barras, no qual o eixo X representa os dias e o eixo Y exibe a quantidade de requisições.
- **acesso_fonte**: Explora a distribuição das entradas entre diferentes fontes. Os dados são apresentados por meio de um gráfico de pizza, indicando o percentual de entradas referentes a cada fonte do experimento.
- **top_ips**: Filtra os cinco endereços IP com mais registros no *DataFrame* e os representa em um gráfico de barras. O eixo X contendo os IPs filtrados, enquanto o eixo Y exibe a quantidade de entradas, ordenados do IP mais frequente para o menos frequente.
- **hp_dia**: Apresenta a frequência de requisições por dia para cada um dos *honeypots*. A solução para essa visualização, foi utilizar um gráfico de linhas, no qual cada

honeypot é representado individualmente, permitindo observar o comportamento durante o período do experimento.

- **ip_dia**: Exibe o comportamento de um endereço IP específico, definido como parâmetro. Os dados são representados em um gráfico de linhas, destacando apenas os *honeypots* que possuem interação com o endereço IP informado.

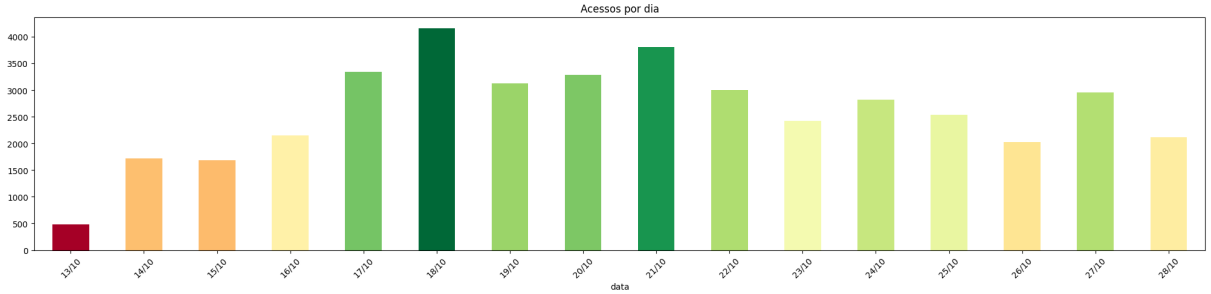

Figura 8 – Exemplo de resultado para a função *acesso_dia* utilizando a base de dados "Experimento - Roteadores". Fonte: Do Autor

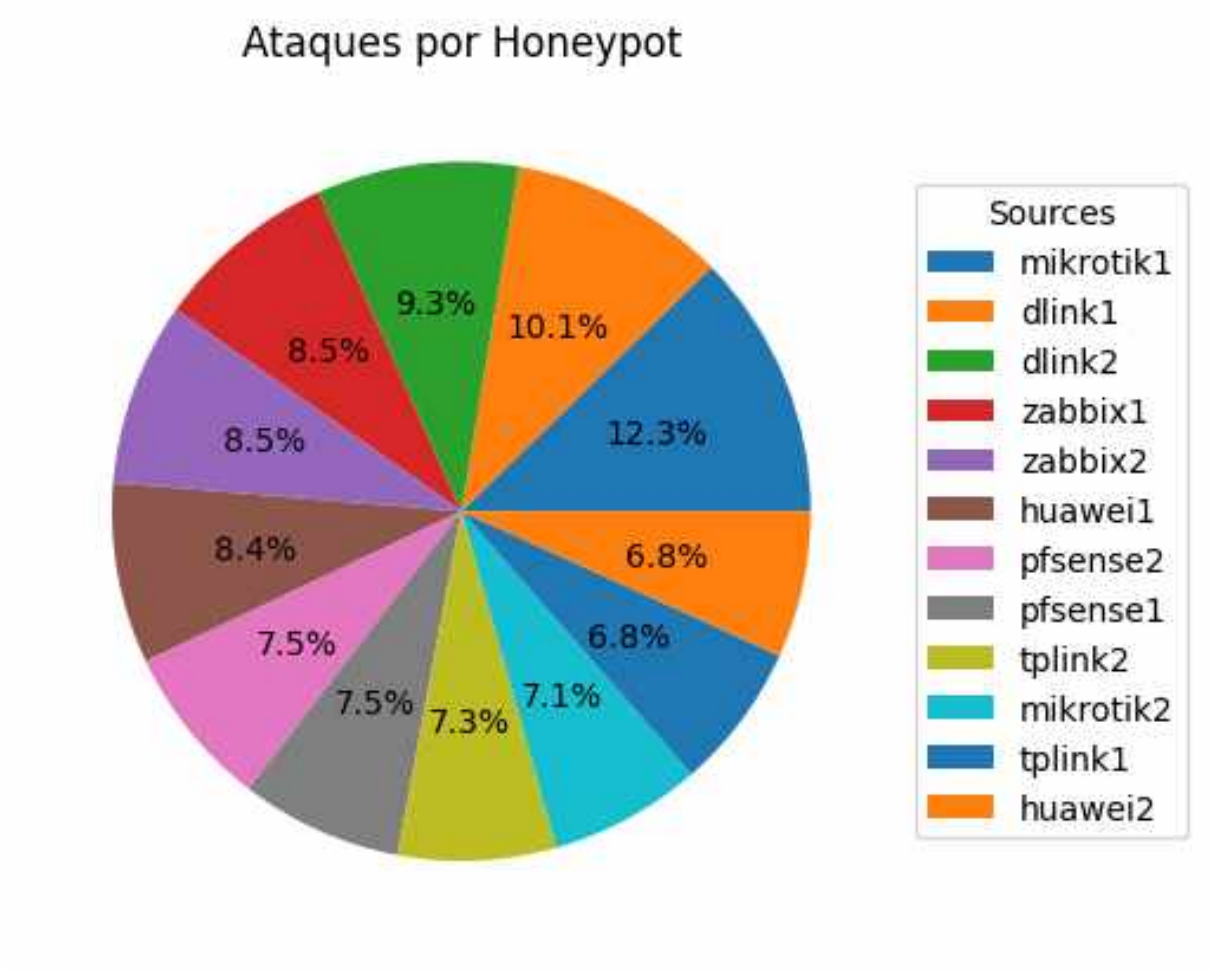

Figura 9 – Exemplo de resultado para a função *acesso_fonte* utilizando a base de dados do "Experimento - Roteadores". Fonte: Do Autor

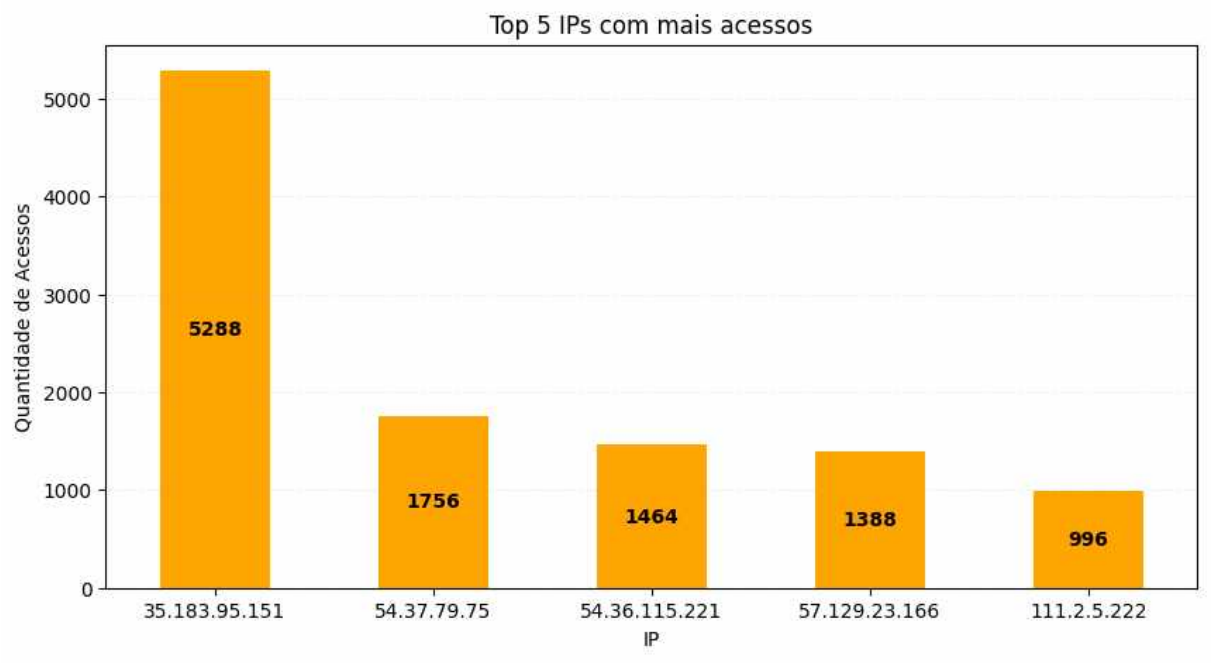

Figura 10 – Exemplo de resultado para a função *top_ips* utilizando a base de dados do "Experimento - Roteadores". Fonte: Do Autor

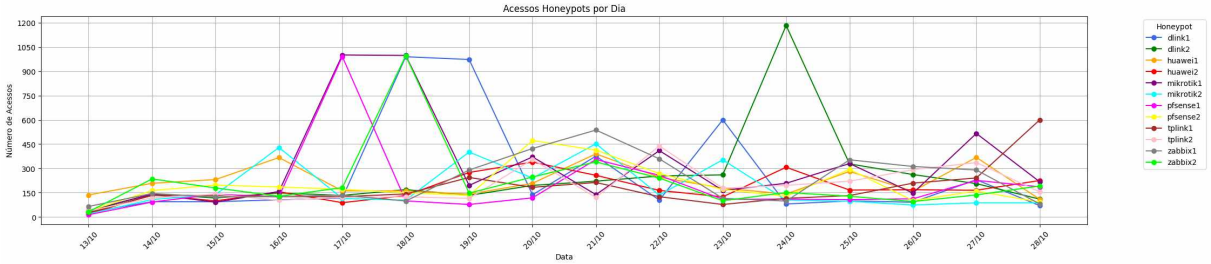

Figura 11 – Exemplo de resultado para a função *hp_dia* utilizando a base de dados do "Experimento - Roteadores". Fonte: Do Autor

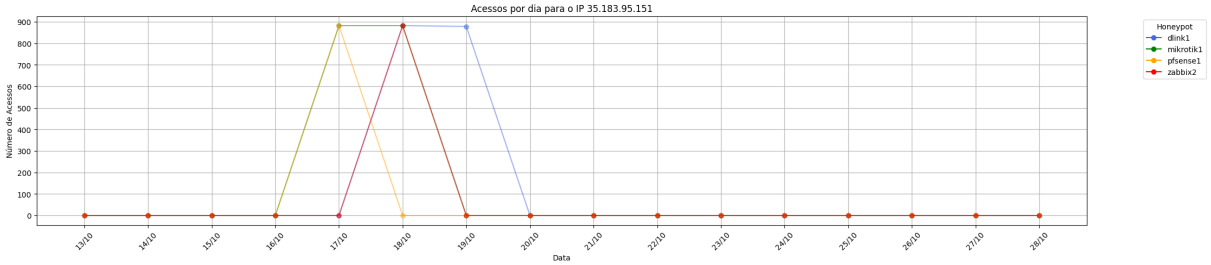

Figura 12 – Exemplo de resultado para a função *ip_dia* para o IP "35.183.95.151" utilizando a base de dados do "Experimento - Roteadores". Fonte: Do Autor

4.3 Comparação entre os conjuntos de dados

Para o desenvolvimento desta etapa, foram utilizados dois conjuntos de dados, com o objetivo de experimentar e acompanhar a evolução da construção da ferramenta. Esses dados são resultados do "Experimento - Roteadores" e do "Experimento - Câmeras", conforme detalhado na Seção 3.1 e ambos os experimentos foram desenvolvidos partindo de uma mesma metodologia. Dessa forma, foi definido um período comum de análise, correspondente aos primeiros 15 dias de coleta, para fins comparativos.

Devido às diferenças entre os tipos de dispositivos emulados, às tarefas atribuídas a cada trabalho e à janela de tempo entre as coletas dos *honeypots*, os comportamentos dos atacantes estão sujeitos a variações. Entretanto, o objetivo desta Seção é demonstrar como essas comparações podem ser conduzidas de forma estruturada. As Figuras 13, 14 e 15 exploram esse processo, abordando visualizações semelhantes às apresentadas na Seção 4.2, facilitando a análise comparativa entre os contextos.

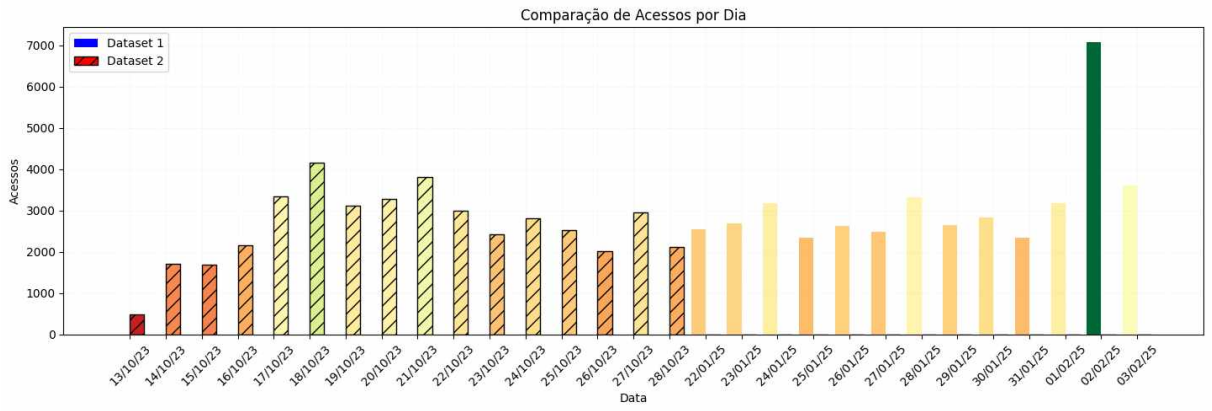

Figura 13 – Comparação de acessos por dia entre as bases de dados do "Experimento - Roteadores" e o "Experimento - Câmeras". Fonte: Do Autor

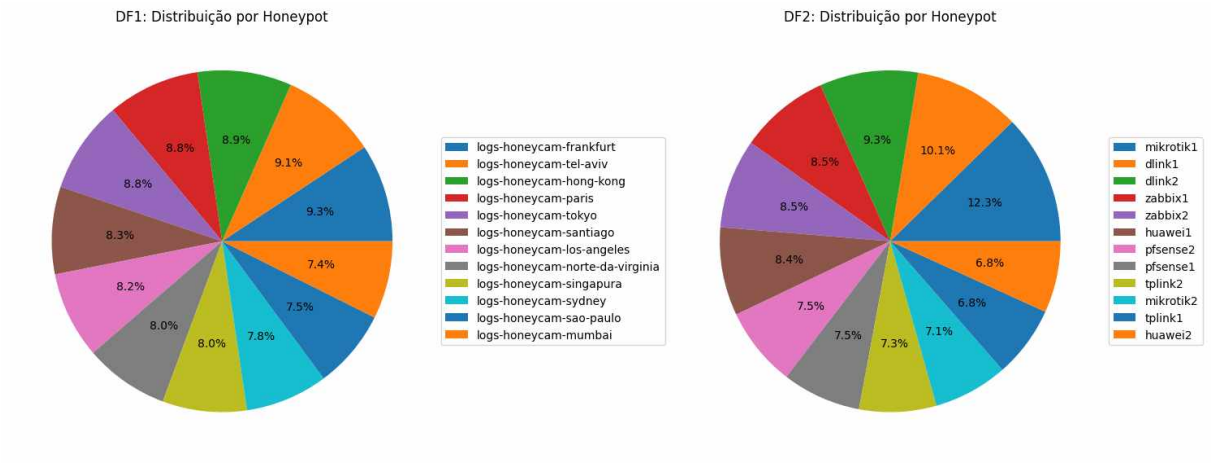

Figura 14 – Comparação de distribuição por *honeypots* entre as bases de dados do "Experimento - Roteadores" e o "Experimento - Câmeras". Fonte: Do Autor

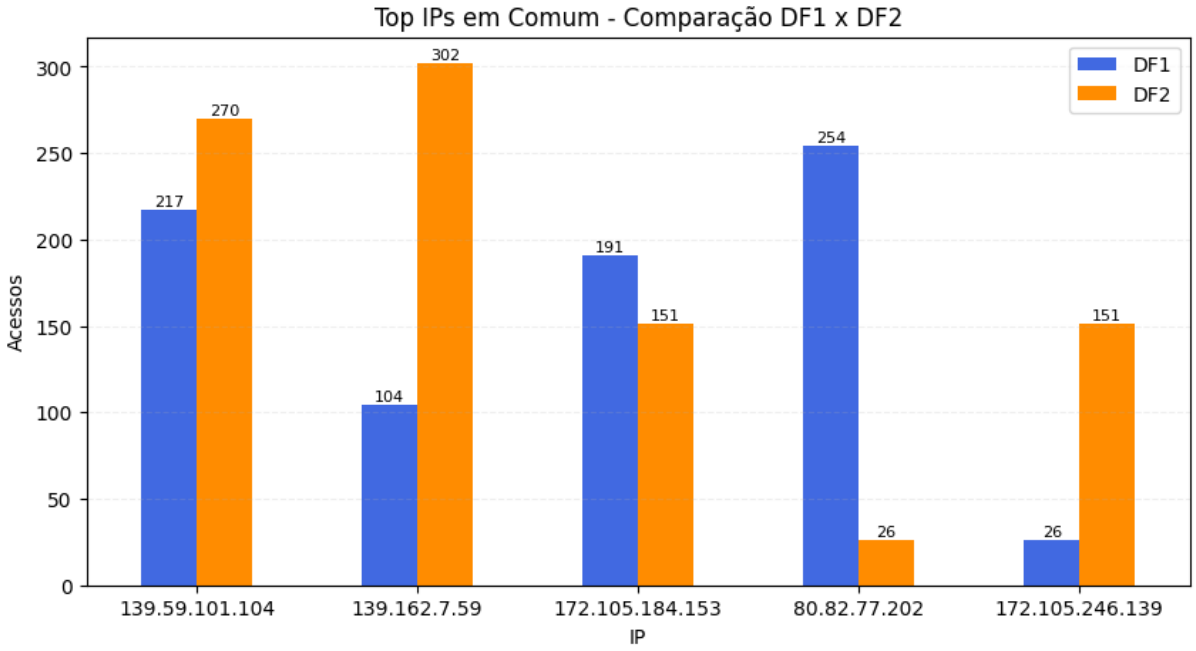

Figura 15 – Comparação dos endereços IP mais frequentes e comuns entre as bases de dados do "Experimento - Roteadores" e o"Experimento - Câmeras". Fonte: Do Autor

4.4 Status da ferramenta

A ferramenta, até o momento da confecção deste trabalho, possui seu escopo inicial desenvolvido, incorporando as funções abordadas ao longo deste Capítulo e garantindo a execução eficiente das principais etapas de processamento e análise dos dados. Os gráficos são carregados diretamente nos *notebooks* onde as funções são executadas, conforme

ilustrado nas Figuras 16 e 17. No entanto, existe um espaço para a implementação de painéis interativos, que possibilitem uma apresentação mais dinâmica e acessível dos dados. Com isso, é possível integrar o uso da ferramenta a projetos e estudos de terceiros, sendo necessário seguir o processo de adequação dos arquivos e ajustar as funções conforme as necessidades do usuário. Além disso, é possível explorar suas funcionalidades em outros escopos, desde que a estrutura básica definida anteriormente seja respeitada; caso contrário, será necessário alterar as regras de execução das funções estabelecidas.

Figura 16 – Exemplo de execução da função *acesso_dia* em *notebook* Python com o conjunto de dados do "Experimento - Câmeras". Fonte: Do Autor

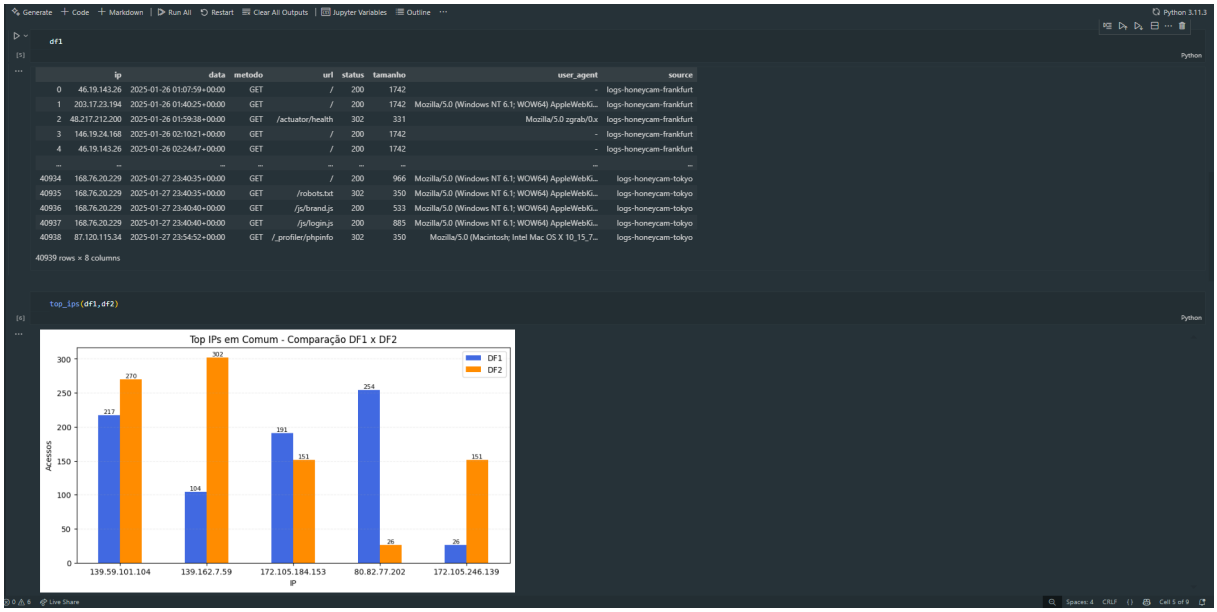

Figura 17 – Exemplo de execução da função *top_ips* em *notebook* Python com os conjuntos de dados do "Experimento - Câmeras" e do "Experimento - Roteadores". Fonte: Do Autor

Além disso, foram realizados testes para identificar e classificar possíveis tipos de ataques, com foco em ameaças do tipo *Cross-Site Scripting* (XSS); entretanto, ainda não foram definidos filtros precisos para identificar todas as entradas que se enquadram nessa categoria. Com o objetivo de agregar mais valor à ferramenta, foi considerada a possibilidade de disponibilizar os gráficos em uma interface web; contudo, a ideia não foi implementada nesta primeira versão.

5 Conclusão

A proposta de desenvolver uma ferramenta para processar os dados coletados de *honeypots* IoT foi o objetivo principal deste trabalho. A necessidade de organizar e transformar um conjunto de dados irregulares para um formato estruturado, convertendo os arquivos *log* em um *DataFrame* de estrutura simples, impulsionou a criação de funções para tratamento, análise e visualização dos dados, garantindo o funcionamento e a escalabilidade da ferramenta.

Durante a etapa de planejamento e construção, priorizou-se a organização e a leitura dos arquivos, garantindo a gestão eficiente de um grande volume de dados. Foram abordadas soluções que facilitam a exploração desses arquivos e a conversão dos *logs* para a estrutura de *DataFrames*, permitindo que a ferramenta se adapte às necessidades e tendências de seus usuários. A visualização gráfica também foi implementada com o intuito de auxiliar esses usuários no processo de investigação dos dados, fornecendo um melhor entendimento dos comportamentos dos invasores nos ambientes emulados.

Como resultado, a ferramenta agregou valor às análises de segurança de dispositivos IoT, facilitando a visualização das métricas mais relevantes nos conjuntos de dados utilizados para teste. Todavia, mesmo com os esforços na construção do método investigativo, alguns tópicos ainda necessitam ser refinados para oferecer ao usuário maior eficiência e personalização em suas análises. Para isso, é essencial dedicar uma maior atenção ao desenvolvimento de novos visuais e filtros, tornando o processo de exploração dos dados mais dinâmico e intuitivo.

Trabalhos futuros poderão utilizar a base desenvolvida neste trabalho para aprofundar conceitos mais específicos e dinâmicos, como a criação de consultas avançadas e o aprimoramento de filtros para uma classificação mais precisa dos tipos de ataques.Outro tópico relevante a ser explorado em trabalhos futuros é o desenvolvimento de um painel interativo, com foco no desenvolvimento da interface gráfica que permita agrupar e organizar os gráficos gerados de forma dinâmica. O painel possibilitaria a comparação visual entre conjuntos de dados de diferentes *honeypots*, viabilizando uma melhor usabilidade e eficiência da ferramenta para a análise e interpretação de resultados.

Referências

ABDMEZIEM, R.; TANDJAOUI, D. **Internet of Things: Concept, Building blocks, Applications and Challenges**. 2014. ArXiv preprint arXiv:1401.6877. Disponível em: <https://arxiv.org/abs/1401.6877>. Citado na página 17.

ABOMHARA, M.; KØIEN, G. M. Cyber security and the internet of things: Vulnerabilities, threats, intruders and attacks. **Journal of Cyber Security and Mobility**, v. 4, n. 1, p. 65–88, 2015. Citado na página 11.

ACIEN, A.; NIETO, A.; FERNANDEZ, G.; LOPEZ, J. A comprehensive methodology for deploying IoT honeypots. In: SPRINGER. **Trust, Privacy and Security in Digital Business: 15th International Conference, TrustBus**. Regensburg, Germany, 2018. p. 229–243. Citado na página 18.

AL-FUQAHA, A.; GUIZANI, M.; MOHAMMADI, M.; ALEDHARI, M.; AYYASH, M. Internet of things: A comprehensive review of enabling technologies, security and privacy, and applications. **IEEE Communications Surveys & Tutorials**, v. 17, n. 4, p. 2347–2376, 2015. Citado na página 16.

AMAN, A. H. M.; YADEGARIDEHKORDI, E.; ATTARBASHI, Z. S.; HASSAN, R.; PARK, Y.-J. A survey on trend and classification of internet of things reviews. **IEEE Access**, IEEE, v. 8, p. 111763–111782, 2020. Citado 4 vezes nas páginas 6, 16, 17 e 18.

ANDRADE, R. O.; ORTIZ-GARCéS, I.; CAZARES, M. Cybersecurity attacks on smart home during covid-19 pandemic. In: **2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)**. United States: Institute of Electrical and Electronics Engineers Inc., 2020. p. 398–404. Citado na página 11.

ASHTON, K. That 'internet of things' thing. **RFID Journal**, Hauppauge, New York, v. 22, n. 7, p. 97–114, 2009. Citado na página 15.

CANONGIA, C.; JUNIOR, R. M. Segurança cibernética: o desafio da nova sociedade da informação. **Parcerias Estratégicas**, v. 14, n. 29, p. 65–88, 2009. Citado na página 10.

CHESWICK, B. An evening with berferd in which a cracker is lured, endured, and studied. In: **Proceedings of the Winter USENIX Conference**. San Francisco, CA: USENIX Association, 1992. p. 163–174. Citado na página 18.

COSTA, H. S. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação), **Construção de Honeypot para câmeras IoT usando computação em nuvem**. Uberlândia, Brasil: [s.n.], 2025. Citado 2 vezes nas páginas 10 e 22.

FAROOQ, M. U.; WASEEM, M.; MAZHAR, S.; KHAIRI, A.; KAMAL, T. A review on internet of things (iot). **International Journal of Computer Applications**, Foundation of Computer Science, v. 113, n. 1, p. 1–7, 2015. Citado na página 17.

Forbes. **Twitter foi hackeado e endereços de e-mail de 200 milhões de usuários vazaram**. 2023. Disponível em: <https://forbes.com.br/forbes-tech/2023/

01/twitter-foi-hackeado-e-enderecos-de-email-de-200-milhoes-de-usuarios-vazaram/>. Citado na página 14.

FRANCO, J.; ARIS, A.; CANBERK, B.; ULUAGAC, A. S. A survey of honeypots and honeynets for internet of things, industrial internet of things, and cyber-physical systems. **IEEE Communications Surveys & Tutorials**, v. 23, n. 4, p. 2351–2383, 2021. Citado 3 vezes nas páginas 6, 19 e 20.

FURMAN, S.; THEOFANOS, M. F.; CHOONG, Y.-Y.; STANTON, B. Basing cybersecurity training on user perceptions. **IEEE Security & Privacy**, IEEE, v. 10, n. 2, p. 40–49, 2012. Citado na página 15.

JIA, X.; FENG, Q.; FAN, T.; LEI, Q. Rfid technology and its applications in internet of things (iot). In: IEEE. **2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)**. [S.l.], 2012. p. 1282–1285. Citado na página 16.

LI, Y.; LIU, Q. A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments. **Energy Reports**, Elsevier, v. 7, p. 8176–8186, 2021. Citado 2 vezes nas páginas 14 e 15.

LUO, T.; XU, Z.; JIN, X.; JIA, Y.; OUYANG, X. Iotcandyjar: Towards an intelligent-interaction honeypot for iot devices. In: **Black Hat USA**. [S.l.: s.n.], 2017. p. 1–11. Citado na página 21.

MADAKAM, S.; RAMASWAMY, R.; TRIPATHI, S. Internet of things (iot): A literature review. **Journal of Computer and Communications**, v. 3, n. 5, p. 164–173, 2015. Citado na página 15.

MAGRANI, E. **A Internet das Coisas**. 1. ed. Rio de Janeiro: Editora FGV, 2018. ISBN 978-85-225-2005-3. Citado na página 10.

MCKINNEY, W. **Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython**. 1. ed. São Paulo: Novatec Editora, 2018. Citado na página 24.

MENDES, L. G. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação), **Construção de infraestrutura de Honeypots IoT usando computação em nuvem**. Uberlândia, Brasil: [s.n.], 2023. Citado 4 vezes nas páginas 7, 10, 22 e 29.

MFOGO, V. S.; ZEMKOHO, A.; NJILLA, L.; NKENLIFACK, M.; KAMHOUA, C. Aiipot: Adaptive intelligent-interaction honeypot for iot devices. **arXiv preprint arXiv:2303.12367**, 2023. Citado na página 21.

MOCRII, D.; CHEN, Y.; MUSILEK, P. Iot-based smart homes: A review of system architecture, software, communications, privacy and security. **Internet of Things**, Elsevier, v. 1, p. 81–98, 2018. Citado na página 16.

NGUYEN, T. T.; REDDI, V. J. Deep reinforcement learning for cyber security. **arXiv preprint arXiv:1906.05799**, 2019. Citado na página 13.

OZA, A. D.; KUMAR, G. N.; KHORAJIYA, M. Survey of snaring cyber attacks on iot devices with honeypots and honeynets. In: **2018 3rd International Conference for Convergence in Technology (I2CT)**. [S.l.: s.n.], 2018. p. 1–6. Citado na página 19.

PA, Y. M.; SUZUKI, S.; YOSHIOKA, K.; MATSUMOTO, T.; KASAMA, T.; ROSSOW, C. Iotpot: A novel honeypot for revealing current iot threats. **Journal of Information Processing**, v. 24, n. 3, p. 522–533, 2016. Citado na página 20.

PEREIRA, M. M. d. A. **Impacto causado pela pandemia de Covid-19 nas empresas da IBOVESPA (B3) do setor tecnológico: uma análise através das notas explicativas**. 2022. Trabalho de Conclusão de Curso (Graduação) — Universidade Federal do Rio Grande do Norte. Disponível em: <https://repositorio.ufrn.br/handle/123456789/46625>. Citado na página 11.

PERKINS, R. C.; HOWELL, C. J. Honeypots for cybercrime research. In: **Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches**. [S.l.]: Springer, 2021. p. 233–261. Citado na página 10.

SINGH, A.; CHATTERJEE, K. Cloud security issues and challenges: A survey. **Journal of Network and Computer Applications**, Elsevier, v. 79, p. 88–115, 2017. Citado na página 15.

SOBESTO, B.; ANCIAUX, N.; BOUET, M.; DRAGONI, N.; SALLES, M. DarkNOC: Dashboard for honeypot management. In: USENIX ASSOCIATION. **25th Large Installation System Administration Conference (LISA 11)**. Boston, MA, USA, 2011. p. 1–6. Citado na página 21.

SOLMS, R. V.; NIEKERK, J. V. From information security to cyber security. **Computers & Security**, Elsevier, v. 38, p. 97–102, 2013. Citado 2 vezes nas páginas 13 e 14.

SRINIVASA, S.; PEDERSEN, J. M.; VASILOMANOLAKIS, E. Interaction matters: a comprehensive analysis and a dataset of hybrid iot/ot honeypots. In: **Proceedings of the 38th Annual Computer Security Applications Conference**. United States: Association for Computing Machinery, 2022. p. 742–755. Citado 2 vezes nas páginas 18 e 20.

STOLL, C. **The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage**. New York: Simon and Schuster, 2005. Citado na página 18.

TABARI, A. Z.; OU, X.; SINGHAL, A. What are attackers after on iot devices? an approach based on a multi-phased multi-faceted iot honeypot ecosystem and data clustering. **arXiv preprint arXiv:2112.10974**, 2021. Citado na página 10.

THAKAR, U.; VARMA, S.; RAMANI, A. K. Honeyanalyzer – analysis and extraction of intrusion detection patterns & signatures using honeypot. In: **Proceedings of the Second International Conference on Innovations in Information Technology**. Dubai, UAE: [s.n.], 2005. p. 1–7. Citado na página 21.

UMAIR, M.; CHEEMA, M. A.; CHEEMA, O.; LI, H.; LU, H. Impact of covid-19 on iot adoption in healthcare, smart homes, smart buildings, smart cities, transportation and industrial iot. **Sensors**, v. 21, n. 11, p. 3838, 2021. Citado na página 11.

WHITMAN, M. E.; MATTORD, H. J. **Principles of information security**. [S.l.]: Cengage Learning, 2021. Citado 2 vezes nas páginas 13 e 14.

WHITMORE, A.; AGARWAL, A.; XU, L. D. The internet of things—a survey of topics and trends. **Information Systems Frontiers**, Springer, v. 17, p. 261–274, 2015. Citado 2 vezes nas páginas 15 e 16.