UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gabriel Oliveira Souza

Ransomset: Um conjunto de dados baseado em análise dinâmica para ransomwares

Uberlândia, Brasil

2024

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gabriel Oliveira Souza

Ransomset: Um conjunto de dados baseado em análise dinâmica para ransomwares

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 S. Miani

Coorientador: Prof. Dr. Silvio E. Quincozes

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2024

Gabriel Oliveira Souza

Ransomset: Um conjunto de dados baseado em análise dinâmica para ransomwares

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.

Prof. Dr. Rodrigo S. Miani
Orientador

Prof. Dr. Silvio E. Quincozes
Co-orientador

Professor

Uberlândia, Brasil
2024

Agradecimentos

Primeiramente, expresso minha profunda gratidão aos meus pais, pelo apoio incondicional, paciência e incentivo ao longo de toda esta jornada. A dedicação e o amor que recebi deles foram essenciais para que eu alcançasse esta etapa, e sou eternamente grato por toda a motivação e pela confiança depositada em meu potencial.

Aos meus irmãos, agradeço pela força transmitida, pela amizade e pelas palavras de encorajamento. A parceria de vocês tornou esse caminho mais leve e inspirador.

Aos meus amigos da faculdade, que estiveram presentes em todos os momentos intensos, felizes, tristes e desafiadores, tanto no curso quanto em minha vida pessoal, meu sincero agradecimento. A amizade pura e leve de cada um de vocês, e o apoio inestimável, foram essenciais para que eu superasse as dificuldades.

Agradeço de coração ao Prof. Dr. Rodrigo S. Miani, que, apesar de todos os meus desafios, sempre me apoiou e me encorajou a ser o meu melhor. Sua compreensão nos momentos difíceis foi um divisor de águas em minha vida. Obrigado pela paciência e pela amizade exemplar, que foram fundamentais para mim.

Ao Prof. Dr. Silvio E. Quincozes, agradeço por acreditar em meu potencial. Mesmo diante de erros, atrasos e dúvidas, seu incentivo constante me motivou a seguir em frente. Sua coragem e determinação para fazer as coisas acontecerem me inspiraram a ser uma pessoa melhor.

Por fim, agradeço a mim mesmo. Apesar de contar com apoio, o caminho até aqui foi, para mim, marcado por momentos de solidão. Enfrentei um vazio difícil de compreender, mas o processo de sair desse lugar e me conhecer melhor foi, ao mesmo tempo, assustador e reconfortante. Hoje, me sinto mais feliz com minhas conquistas e comigo mesmo. Agradeço a cada pessoa que fez parte da minha trajetória, aos que encontrei, aos que perdi pelo caminho, a todos que contribuíram para esta história, que hoje me inspira a continuar.

Com carinho, Gabriel Oliveira Souza.

Resumo

Ransomware é um tipo de software malicioso que criptografa arquivos e exige pagamento para a restauração do acesso. Como os *ransomwares* são frequentemente detectados tardiamente, representam uma ameaça crescente para indivíduos e organizações, causando perdas financeiras substanciais e comprometendo dados confidenciais. Este trabalho tem como objetivo a criação e análise de um conjunto de dados para a detecção de 6 tipos conhecidos de *ransomware*: *WannaCry*, *Ryuk*, *LockBit*, *CryptoLocker*, *Sodinokibi* e *Conti*. Para isso, foi utilizado o ambiente *Cuckoo Sandbox*, no qual as atividades desses *ransomwares* são registradas. A partir desses registros, foi criado um conjunto de dados contendo amostras normais e de ataque para cada tipo de *ransomware* estudado. Por fim, foram aplicados métodos de ranqueamento de atributos nesses conjuntos de dados para identificar os 25 atributos mais relevantes para a detecção dos *ransomwares* analisados, um dos atributos reportados foi o *NtCLose*, o atributo tem sua função de fechar um objeto no Sistema Operacional (SO), esse forma de ranqueamento de atributos contribui para o aprimoramento dos mecanismos de detecção de *ransomware*. Conclui-se que o RansomSet demonstrou ser uma ferramenta eficaz para a análise e detecção de ransomware, fornecendo uma base sólida para o desenvolvimento de modelos de segurança cibernética mais robustos. Além disso, sua estrutura flexível permite futuras expansões e adaptações, contribuindo para o avanço no combate a ameaças cibernéticas emergentes.

Palavras-chave: *Ransomwares*, Atributos, Seleção de atributos, Detecção.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

IG	*Information Gain* (Ganho de Informação)
JSON	*JavaScript Object Notation* (Notação de Objetos JavaScript)
CSV	*Comma-Separated Values* (Valores Separados por Vírgula)
EDA	*Exploratory Data Analysis* (Análise Exploratória de Dados)
KARA	Relatório de Ameaças da Kaspersky
CTI	*Cyber Threat Intelligence* (Inteligência de Ameaças Cibernéticas)
C&C	*Command and Control* (Comando e Controle)
RaaS	*ransomware-as-a-service* (Ransomware como Serviço)
VM	*Virtual Machine* (Máquina Virtual)

Sumário

1 Introdução

A pandemia de COVID-19 impactou severamente a forma de trabalho, intensificando o uso de computadores e ambientes virtuais. Pesquisas em mais de 200 empresas indicam que 81% delas adotaram o modelo de trabalho remoto (CURRAN, 2020). Como consequência, a ocorrência de ataques por meio de *malwares* aumentou, atingindo um percentual superior a 75% no primeiro semestre de 2022[1]. Entre os *malwares* mais devastadores, destacam-se os *ransomwares*, que bloqueiam o acesso aos dados das vítimas, onde o atacante por trás do ataque acaba praticando a extorsão para o pagamento de um resgate de dados criptografados (ABBASI et al., 2022)(BREWER, 2016). Embora recente, o surgimento do *ransomware* já causa sérios prejuízos financeiros e operacionais, especialmente com variantes populares como *WannaCry*, *Ryuk*, *LockBit*, *CryptoLocker*, *Sodinokibi* e *Conti* (BEAMAN et al., 2021) (HYUNA, 2024).

O *ransomware* bloqueia e criptografa dados, exigindo pagamento para restaurar o acesso. A perda temporária de controle é crítica, especialmente em Sistemas de Controle Industrial (ICSs) e dispositivos associados, como HMIs, PLCs e IEDs (MICRO, 2020). O modelo *ransomware-as-a-service* (*RaaS*), onde desenvolvedores fornecem ferramentas de *ransomware* a afiliados mediante pagamento, facilita sua propagação e torna a defesa digital mais difícil. Relatórios indicam aumento na atividade cibernética maliciosa, reforçando a necessidade de métodos eficazes de identificação e defesa (MICRO, 2022a) (RAZAULLA et al., 2023).

Estudos sobre *ransomware* investigam métodos de detecção, focando na análise dinâmica de seu comportamento em ambientes controlados, como máquinas virtuais ou *sandboxes* (DAMODARAN et al., 2017). Entretanto, limitações na seleção de amostras benignas afetam a replicação dos resultados. Um estudo recente (SOUZA et al., 2023) identificou e classificou atributos de detecção, considerando fatores como frequências de chamadas de sistema e padrões comportamentais de *ransomwares* específicos, incluindo *WannaCry*, *Ryuk* e *CryptoLocker*.

Embora tenha contribuído para a identificação de características relevantes, essa análise inicial não explorou a combinação de atributos em um contexto mais amplo, nem estendeu a abordagem para outros tipos de *ransomware*. O estudo serviu como base para o desenvolvimento de um conjunto de dados abrangente e multiclasse, permitindo uma análise comparativa entre famílias de *ransomware* e comportamentos benignos.

1.1 Objetivos

Este trabalho tem como principal objetivo o desenvolver do RansomSet, um conjunto de dados multiclasse focado na análise comportamental de *ransomwares* e de amostras benignas.

[1] Fonte: <https://canaltech.com.br/seguranca>.

O RansomSet visa aprimorar a caracterização, classificação e seleção de atributos para detecção e análise de diferentes classes de *ransomware*. Os objetivos específicos do trabalho incluem:

- Desenvolver de um conjunto de dados multiclasse para análise de *ransomware*, incluindo informações detalhadas e resumos históricos das amostras;
- Caracterizar o comportamento dos *ransomwares* por meio da análise de chamadas de sistema, permitindo a identificação de padrões e comportamentos específicos;
- Aperfeiçoar a classificação e seleção de atributos em diferentes classes e amostras, visando a melhoria na detecção e identificação de ameaças;
- Criar uma base de dados representativa que permita comparações entre comportamentos maliciosos e benignos, fornecendo suporte para o treinamento de modelos de detecção de ameaças cibernéticas.

1.2 Monografia

Esta monografia está organizada da seguinte forma:

- Referencial Teórico: Este capítulo apresenta os conceitos fundamentais e revisa estudos sobre *ransomware* e métodos de detecção, abordando comportamento, técnicas de análise e desafios na identificação de ataques, além das contribuições de pesquisas recentes.
- Conjunto de Dados: Este capítulo detalha o processo de criação do conjunto de dados usado na pesquisa. O fluxo de coleta de atributos, realizado com o *Cuckoo Sandbox*, monitora o comportamento dos *ransomwares* em ambiente controlado, registrando interações como chamadas de sistema e acessos a arquivos. Esse conjunto multiclasse visa capturar padrões para detectar atividades maliciosas.
- Conclusão: A conclusão resume os principais resultados e contribuições do trabalho, enfatizando o desenvolvimento do conjunto de dados para análise de *ransomware*. Destacam-se, ainda, possíveis aplicações futuras e sugestões de aprimoramentos para pesquisas subsequentes, além de uma seção dedicada às produções bibliográficas relacionadas ao estudo.

2 Revisão Bibliográfica

Este capítulo apresenta os fundamentos teóricos necessários para a compreensão deste estudo e situa a pesquisa no contexto de trabalhos realizados na área. Serão discutidos conceitos essenciais para o entendimento do problema, incluindo a evolução histórica dos *ransomwares* na Seção 2.1.1, os tipos de *ransomware* comumente estudados na Seção 2.1.2 e as principais abordagens de análise aplicadas a essas ameaças na Seção 2.1.3. A Seção 2.2 abordará os trabalhos relacionados, destacando as contribuições e limitações de estudos prévios que investigaram conjuntos de dados para análise de *ransomware*, servindo como base e referência para o desenvolvimento do RansomSet.

2.1 Fundamentação Teórica

Para a compreensão do trabalho, é necessário abordar conceitos fundamentais, como a análise de *ransomwares* e estratégias para sua realização, além do conhecimento da linguagem de programação *Python*. Esses conceitos serão apresentados a seguir.

2.1.1 Evolução dos *Ransomwares*

Ao longo da história, os *ransomwares* emergiram como uma ameaça constante, evoluindo em suas estratégias de propagação e impacto. Sua origem remonta à década de 1980, com os primeiros indícios de binários maliciosos. Uma das versões iniciais é o *AIDS Trojan*, de 1989, que exibia uma mensagem no sistema alvo, alegando que o sistema havia sido bloqueado devido a atividades ilegais e exigindo um resgate para a liberação dos dados. Nos anos 2000, a evolução da tecnologia possibilitou uma maior sofisticação do *ransomware*. O *GPcode*, em 2004, foi um dos primeiros a incorporar criptografia avançada para bloquear arquivos e exigir resgate. O *CryptoLocker*, em 2013, representou um marco, ao utilizar criptografia avançada para comprometer sistemas (RICHARDSON; NORTH, 2017) (HUMAYUN et al., 2021).

A contínua evolução do *ransomware* destaca a importância de estudos aprofundados para compreender o funcionamento e a execução desses binários. Diante da sofisticação crescente dos ataques, torna-se crucial investigar os *ransomwares* mais proeminentes, que representam uma ameaça recente ao setor industrial, como o *Ryuk*, *CryptoLocker*, *Sodinokibi*, *LockBit* e *Conti*, como apresentado na Figura 1 (MICRO, 2022b) (MICRO, 2024) (MICRO, 2022a). Justifica-se, ainda, a análise do *WannaCry*, que, apesar de não ser recente, teve impacto significativo enquanto ativo. A investigação dessas variantes possibilita um entendimento melhor de suas táticas, vulnerabilidades exploradas e métodos eficazes de prevenção e mitigação (MICRO, 2020).

Figura 1 – Distribuição geográfica das regiões mais afetadas por ataques de *ransomware*, destacando áreas com maior incidência.

2.1.2 Famílias de *ransomwares* estudadas

Os *ransomwares* estudados apresentam características únicas, como classe, método de propagação e forma de infecção do alvo. O primeiro deles é o *WannaCry*, que afetou mais de 300 mil sistemas em mais de 150 países. A complexidade da defesa contra o *WannaCry* decorre de sua capacidade de se disseminar para outros sistemas, característica de um *worm*, o que aumenta a probabilidade de um ataque bem-sucedido, demandando mecanismos de defesa capazes de reagir em tempo real (AKBANOV; VASSILAKIS; LOGOTHETIS, 2019).

De acordo com a Trend Micro, o *ransomware Ryuk*[1] é conhecido por atingir empresas e instituições relevantes, espalhando-se por meio de e-mails de *phishing* ou exploração de vulnerabilidades do sistema. Após a invasão, *Ryuk* criptografa os arquivos e exige resgate em criptomoeda. Sua sofisticação e altos valores de resgate são características associadas a grupos avançados de crimes cibernéticos.

O *CryptoLocker*, por sua vez, utiliza métodos de criptografia que tornam o ataque difícil de ser revertido. Esse *ransomware* utiliza uma chave pública para criptografia e uma chave privada para descriptografia. Durante sua execução, o *CryptoLocker* se instala no perfil do usuário e escaneia discos e arquivos de redes (SHARMA; SHANKER, 2022).

O *Sodinokibi*, também conhecido como *REvil*, explora vulnerabilidades em softwares de servidor e possui habilidades evasivas que dificultam sua detecção por antivírus tradicionais, caracterizando-se por estratégias de ataque sofisticadas e resgates elevados (UMAR; RIADI; KUSUMA, 2021).

Por fim, o *ransomware LockBit*, originado em setembro de 2019, opera como um serviço

[1] Disponível em: <https://www.trendmicro.com/en_us/what-is/ransomware/ryuk-ransomware.html>

(*RaaS*) e ganhou destaque após o ataque à Accenture (ADVISOR, 2021) em 2021. Focado em empresas e órgãos governamentais, criptografa os primeiros 4 KB de cada arquivo para otimizar a velocidade de execução, utilizando *AES* e *RSA*. A versão 2.0, lançada em junho de 2021, introduz uma fase simultânea de ataque e propagação, desativa avisos de segurança e adota dupla extorsão. Durante a propagação, verifica redes locais para rápida disseminação. O *LockBit* aprimora-se com técnicas anti-análise e semelhanças com *DarkSide* e *BlackMatter*, enquanto a versão 2.0 evidencia alta prevalência, segundo dados do *MalwareBazaar* (GAGULIC et al., 2023). Sua popularidade cresceu, consolidando-se como uma das principais variantes de *ransomware* em 2024, com um número elevado de ataques registrados globalmente, especialmente devido à capacidade de realizar extorsão dupla e comprometer sistemas em poucos minutos (HYUNA, 2024)(MICRO, 2024).

2.1.3 Tipos de análises de *ransomwares*

Para observar e compreender o comportamento de um binário malicioso, utilizam-se técnicas que permitem identificar seus tipos, ações, comportamentos e características. O objetivo dessas técnicas de análise é entender as atividades realizadas pelo binário em um momento específico de sua execução, coletando informações para estudo e análise dos dados gerados. As principais abordagens de análise para *ransomwares* incluem a análise estática, dinâmica e híbrida (SIHWAIL; OMAR; ARIFFIN, 2018).

A análise estática é uma técnica fundamental que permite examinar binários maliciosos sem executá-los. Essa abordagem possibilita uma análise detalhada do código-fonte de um binário, avaliando suas propriedades, estrutura e comportamento potencialmente suspeito. Durante esse processo, utilizam-se técnicas de depuração para descobrir informações importantes por meio da análise de sequências de *opcode* e do fluxo de controle no código. Essa análise, realizada com o auxílio de ferramentas e *software* especializados, permite a identificação de padrões maliciosos, assinaturas específicas e aspectos fundamentais do código, sendo uma ferramenta valiosa para a compreensão de ameaças potenciais. Além de auxiliar na compreensão da estrutura dos sistemas binários, oferece subsídios para o desenvolvimento de estratégias defensivas e medidas proativas para fortalecer a segurança cibernética de uma organização (DAMODARAN et al., 2017).

Na análise estática, há dois tipos principais de avaliação. O primeiro é o método *signature-based*, que compara padrões de um binário malicioso, utilizando esses padrões para identificar um *malware* específico. Como os códigos de diferentes binários variam, esse método permite distinguir um *malware* de outro. O segundo método é a *heuristic detection*, semelhante ao *signature-based*, mas que inclui a busca por comandos e instruções que não estão diretamente presentes no binário malicioso, possibilitando a detecção de novas variantes de *malwares* ainda não identificadas (DAMODARAN et al., 2017)(UPPAL; MEHRA; VERMA, 2014).

Conforme mencionado, a análise dinâmica, diferentemente da estática, envolve a inspe-

ção de binários maliciosos durante sua execução. Esse método executa o binário em questão e analisa o fluxo de informações e as chamadas de função enquanto ele está em operação. A análise dinâmica é realizada em máquinas virtuais, permitindo que as interações do programa em tempo real sejam observadas e analisadas. Essa abordagem fornece informações importantes sobre a atividade de códigos maliciosos, facilitando a compreensão de seu comportamento suspeito e permitindo a identificação de potenciais ameaças à segurança do sistema. Além de descrever o comportamento dos binários em execução, a análise dinâmica contribui para o aprimoramento das estratégias de defesa e das medidas de segurança cibernética (DAMODARAN et al., 2017).

A análise híbrida combina as técnicas de análise estática e dinâmica (UPPAL; MEHRA; VERMA, 2014). Essa abordagem examina padrões presentes em um binário e, em seguida, realiza a detecção de informações e instruções durante sua execução em uma *sandbox*.

2.2 Trabalhos Relacionados

A pesquisa sobre *ransomware* e suas características tem crescido nos últimos anos, acompanhando a complexidade e a frequência com que esses ataques ocorrem. Diversos estudos e conjuntos de dados foram desenvolvidos com o intuito de capturar as particularidades do comportamento de *ransomwares* em sistemas computacionais, especialmente por meio de análises dinâmicas. Esses conjuntos de dados desempenham um papel fundamental no treinamento de modelos de aprendizado de máquina para detectar e categorizar *ransomwares* de forma eficaz. Nesta seção, são revisados trabalhos relevantes, destacando seus métodos e limitações, a fim de contextualizar a contribuição do *RansomSet*, que se propõe a oferecer uma nova perspectiva e dados atualizados para análise e detecção de *ransomwares* modernos.

O estudo de (HERRERA-SILVA; HERNáNDEZ-ÁLVAREZ, 2023) emprega análise dinâmica e algoritmos de aprendizado de máquina para identificar assinaturas de *ransomwares*. Os principais objetivos incluem a realização de experimentos com *ransomwares* juntamente com *softwares* benignos, cuja interação tende a gerar arquivos *JSON* com parâmetros dinâmicos em ambiente *sandbox*. O estudo também gera um conjunto de dados e o aplica para criar modelos de aprendizado de máquina para detecção de *ransomwares* usando plataformas *Windows*. O estudo analisa 20 amostras de *ransomware* e 20 amostras de *goodware*, resultando em um conjunto de dados composto por 50 características. A pesquisa reforça a hipótese de que um conjunto dinâmico de assinaturas pode ser criado ao bloquear *ransomwares* juntamente com *softwares* benignos em diferentes plataformas *Windows*. Embora a pesquisa mencione o uso de binários de um repositório para garantir a abrangência, não fica claro o critério específico para a seleção de cada binário. É importante que a escolha dos binários seja justificada de forma clara, explicando, por exemplo, se o binário é recente, seu impacto atual e o motivo de sua inclusão na pesquisa. Sem essa justificativa, corre-se o risco de utilizar binários desatualizados, que podem ser detectados por mecanismos *antimalware* básicos, comprometendo a relevância

dos resultados obtidos.

No trabalho realizado em (SGANDURRA et al., 2016), é apresentada a abordagem *EldeRan*, baseada em aprendizado de máquina para análise e classificação dinâmica de *ransomware*. O *EldeRan* monitora um conjunto de ações realizadas por um aplicativo durante os estágios iniciais de instalação, em busca de sinais e características típicas de *ransomware*. O conjunto de dados possui amostras de binários pertencentes a 12 famílias diferentes de *ransomware*, binários benignos e trojans, totalizando 942 amostras. O *EldeRan* também visa demonstrar sua capacidade de identificar novas famílias de *ransomwares*. Entretanto, com base nas recomendações da pesquisa, os autores sugerem reduzir o tempo de análise dos recursos. Essa abordagem gera uma quantidade limitada de informações sobre a execução dos binários, como chamadas de *API* do sistema e tráfego de rede. Isso afeta a capacidade de capturar o comportamento completo do *ransomware*, limitando a detecção de padrões complexos e de atividades que ocorrem após um período mais longo de execução, como a ativação de cargas maliciosas que dependem de temporizadores ou de eventos específicos. Consequentemente, essa limitação pode comprometer a eficácia da análise e a precisão dos resultados obtidos.

Em (RIECK et al., 2011), é apresentado um estudo utilizando técnicas de aprendizado de máquina. O trabalho destaca a importância da análise dinâmica de *malware* binário em tempo de execução como uma ferramenta essencial para caracterizar e proteger sistemas contra essas ameaças. Os autores contribuem com um mapeamento do comportamento monitorado para um espaço vetorial, facilitando o acesso eficiente aos padrões comportamentais por meio do aprendizado de máquina. A funcionalidade do *framework* é demonstrada através de uma avaliação com *malwares* reais. No entanto, como citado pelo autor, o servidor que armazena o conjunto de dados ficou offline, e o diretório disponível contém apenas os dados que foram salvos ou restaurados, fragilizando a análise sem o conjunto de dados completo. As informações obtidas a partir do diretório disponível[2] indicam que o conjunto de dados possui 24 *malwares* e que suas características foram coletadas com a ferramenta *CWSandbox*, totalizando 300 amostras.

Diversos estudos têm se concentrado na detecção de *malware* utilizando análise de memória e técnicas de aprendizado de máquina, enfrentando desafios em termos de complexidade e tempo de processamento. Pesquisas como a de (CARRIER et al., 2022) propõem uma estrutura de detecção de *malware* usando métodos de engenharia de recursos de memória e modelos de aprendizado de máquina empilhados. A criação do conjunto de dados *MalMemAnalysis2022* visa permitir a avaliação da eficácia dessa solução na simulação de cenários de *malware* do mundo real. Esse conjunto de dados possui 2.916 binários benignos e 5 famílias de *ransomwares*, totalizando 57 atributos e 58.058 amostras. Embora o estudo apresente uma abordagem para a detecção de *malware* utilizando engenharia de recursos de memória e aprendizado de máquina empilhado, possui uma limitação significativa: o conjunto de dados *MalMemAnalysis2022* inclui apenas 5 famílias de *ransomware*. Além disso, os binários utilizados no estudo são de 2022

[2] Disponível em: <https://www.sec.cs.tu-bs.de/data/malheur/>

ou anteriores, o que pode limitar a eficácia da abordagem em detectar *ransomwares* mais modernos, que empregam técnicas de evasão mais sofisticadas. Em comparação, este trabalho inclui amostras de *ransomware* de 2023, refletindo melhor o cenário atual e oferecendo uma análise mais robusta frente às ameaças emergentes que podem não estar presentes no conjunto de dados utilizado no estudo mencionado.

Em (HIRANO; HODOTA; KOBAYASHI, 2022), o *RanSAP* é apresentado. Trata-se de um conjunto de dados aberto voltado para o estudo de *ransomwares*, incluindo padrões de acesso de 7 famílias de *ransomwares* e 5 tipos de *software* benigno. Para a construção do conjunto de dados proposto, os autores realizam análise dinâmica usando o *software BitVisor* com o objetivo de capturar o comportamento malicioso. Em seus experimentos, cada binário malicioso ou benigno é executado uma vez. Contudo, os binários do *RanSAP* não são atualizados. Além disso, como cada binário é executado apenas uma vez, a captura do comportamento malicioso tende a ser menos fidedigna. O trabalho exclui *ransomwares* proeminentes da atualidade, como o *LockBit*.

O trabalho (SCAIFE et al., 2016) propõe um sistema de alerta para detecção de *ransomware*, monitorando diretamente os dados do usuário em busca de alterações indicativas, em vez de inspecionar a execução do programa. O sistema visa detectar atividades suspeitas independentemente do método de execução do *malware*. De acordo com a pesquisa, o *CryptoDrop* foi projetado para complementar, e não substituir, o *software* anti-*malware* existente. O estudo identifica três indicadores primários que, em conjunto, ajudam a detectar alterações maliciosas em arquivos. Além disso, a pesquisa inclui análise de *ransomware* de criptografia, visando a perda mínima de arquivos, reduzindo a necessidade de pagamento de resgate e enfraquecendo a eficácia do ataque. As amostras foram capturadas por meio do *Cuckoo Sandbox*, totalizando 492 *malwares*, pertencentes a 14 famílias de *ransomwares*. Embora o trabalho (SCAIFE et al., 2016) apresente o *CryptoDrop* como um sistema complementar para a detecção de *ransomware*, há uma limitação importante: o conjunto de dados utilizado na pesquisa não está publicamente disponível. Isso impede que outros pesquisadores validem os resultados ou realizem comparações independentes com novas abordagens. Sem acesso ao conjunto de dados, não é possível avaliar com precisão e clareza a eficácia do sistema, nem verificar a representatividade das amostras de *ransomware* usadas no estudo, o que compromete a replicabilidade e a generalização dos resultados para cenários mais atuais.

No estudo (TANG et al., 2020) é apresentado o *RansomSpector*, que utiliza famílias de *ransomwares* para detecção, operando em uma máquina virtual e na camada do hipervisor, abaixo do sistema operacional onde o *ransomware* é executado. Esse sistema monitora as atividades de rede e o sistema de arquivos de baixo nível do *ransomware* e modela seus padrões de acesso de I/O de arquivos e atividade de rede. Com base na análise do estudo, foi utilizada uma base de 2017 da plataforma *VirusTotal* com *ransomwares* convencionais para a detecção de binários já identificados como maliciosos na própria plataforma *VirusTotal*. A amostra inicial

inclui *ransomwares* que são facilmente bloqueados pelos sistemas de antivírus mais básicos, comprometendo sua eficácia em análises de binários recentes e com novas técnicas de ataque. Embora o *RansomSpector* apresente características importantes na detecção de *ransomwares* ao utilizar introspecção em máquinas virtuais e monitorar atividades de sistema de arquivos e rede, há limitações quanto à relevância das amostras de *malware* utilizadas. O estudo avalia amostras de *ransomwares* anteriores a 2020, que são facilmente detectadas por soluções *antimalware* mais recentes. Além disso, o trabalho não inclui *ransomwares* mais sofisticados, como o *LockBit*, que tem se destacado em relatórios recentes por sua alta capacidade de evasão. A ausência da análise dessas novas ameaças reduz a aplicabilidade do *RansomSpector* frente aos *ransomwares* emergentes.

Em (BERRUETA et al., 2020), o trabalho descreve a criação de um repositório público de dados chamado *PCAP*. Ele contém operações de acesso a arquivos realizadas por mais de 32 famílias de *ransomwares* durante a criptografia de dados em diretórios compartilhados em rede. No artigo, os autores detalham a captura dos dados e como eles podem ser empregados na avaliação e comparação de resultados de diversas técnicas de detecção de *ransomware*. O trabalho estuda 94 amostras de *ransomwares*, utilizando servidores *C&C* para realizar a análise desses binários. No entanto, 73% dessas amostras são obsoletas, tendo como ano de lançamento 2019 ou anterior. Além disso, o número de atributos do conjunto de dados é baixo, especialmente considerando que é comum a aplicação de algoritmos de seleção para considerar apenas as atributos mais relevantes no processo de classificação.

Em (CONTINELLA et al., 2016), é apresentada uma abordagem para proteger sistemas operacionais contra ataques de *ransomware*. Em vez de depender exclusivamente de técnicas de detecção, como *sandboxes*, o *ShieldFS* atua como um driver adicional para o sistema de arquivos nativo do *Windows*, visando ser eficaz contra ataques de *ransomwares*. O sistema utiliza um mecanismo de análise dinâmica que cria cópias dos arquivos ao detectar atividades maliciosas no sistema de arquivos, permitindo que as operações prejudiciais sejam revertidas automaticamente. A partir de uma análise de requisições de I/O de sistemas, o *ShieldFS* constrói modelos adaptativos que monitoram a atividade do sistema e identificam comportamentos maliciosos em tempo real. Ele trabalha com binários benignos, totalizando 2.245, e 688 *malwares*, abrangendo 11 famílias de *ransomwares*. Embora o *ShieldFS* ofereça uma solução para combater ataques de *ransomware* ao criar modelos de detecção baseados na atividade do sistema de arquivos e implementar um sistema de autorrecuperação, há limitações devido à sua data de publicação. Como o estudo é de 2017, muitos dos *ransomwares* analisados podem ser facilmente detectados por mecanismos *antimalware* mais recentes. Além disso, o cenário de *ransomware* evoluiu significativamente desde então, com novas variantes e técnicas de evasão que não foram abordadas no estudo, limitando sua aplicabilidade frente às ameaças modernas[3].

No trabalho (KHARAZ et al., 2016), é proposto um sistema de análise dinâmica para de-

[3] Disponível em: <https://www.crowdstrike.com.br/recursos/white-papers/a-evolucao-do-ransomware/>

tectar *ransomwares*. O sistema simula um ambiente artificial de usuário e monitora as interações do *ransomware* com os arquivos e a área de trabalho, identificando comportamentos maliciosos. A avaliação do *UNVEIL* visa detectar *ransomwares* evasivos, anteriormente desconhecidos, que passaram despercebidos por soluções *antimalware* existentes, contribuindo para o avanço da detecção de *ransomware*. O sistema utiliza 7 binários benignos para a realização da análise e 1.926 *malwares*, abrangendo 11 famílias de *ransomwares*. As análises dinâmicas e capturas de informações de cada binário malicioso foram realizadas por meio do *Cuckoo Sandbox*. Embora o trabalho *UNVEIL* tenha introduzido uma abordagem para detectar *ransomware* em 2016, criando um ambiente artificial de usuário para capturar a interação maliciosa com arquivos ou a área de trabalho, ele também apresenta as mesmas limitações observadas no *ShieldFS*. Como ambos os estudos são de meados de 2016 e 2017, os binários de *ransomware* utilizados nas avaliações são antigos. Hoje, muitos desses binários maliciosos e suas técnicas são facilmente detectados por soluções *antimalware* modernas. Assim como o *ShieldFS*, o *UNVEIL* pode não ser tão eficaz contra *ransomwares* mais recentes que empregam técnicas de evasão avançadas, destacando a necessidade de atualizações e reavaliações para enfrentar as ameaças emergentes no cenário atual.

No trabalho proposto, o foco se concentra nos *ransomwares* mais utilizados atualmente pelos criminosos, de acordo com a Trend Micro[4,5], ou seja, nas ameaças mais recentes. Estudos anteriores empregaram uma variedade de *ransomwares* que são facilmente detectáveis por plataformas de *antimalware* e antivírus atuais, comprometendo a eficácia da análise. Muitas amostras utilizadas em pesquisas anteriores são antigas, enquanto novas versões de *ransomware* apresentam métodos mais sofisticados e representam um desafio maior. Em alguns trabalhos que propõem um conjunto de dados relacionado a *ransomwares*, o critério para a seleção dos *ransomwares* estudados não fica claro, o que compromete a confiabilidade do estudo.

A Tabela 1 mostra os tipos de *ransomware* utilizados neste estudo. Cada uma das famílias listadas foi coletada na plataforma *VirusShare*[6], um repositório amplamente utilizado para o compartilhamento e análise de amostras de *malware*. Embora o número de famílias de *ransomware* analisadas seja relativamente reduzido, a pesquisa foca em amostras recentes e avançadas, alinhadas com as ameaças contemporâneas enfrentadas por indivíduos e organizações. Essa abordagem permite uma investigação aprofundada e representa com maior precisão o atual cenário de ameaças, possibilitando que os dados coletados sejam utilizados para construir modelos de detecção que capturem as características mais relevantes dos *ransomwares* prevalentes. Dessa forma, a seleção criteriosa das amostras proporciona uma base sólida para análise, conforme demonstrado na Tabela 1.

Na Tabela 2 temos as diferenças entre os conjunto de dados utilizados nesses trabalhos

[4] Disponível em: <https://documents.trendmicro.com/assets/pdf/Conti-and-LockBit-Make-Waves-with-High-Profile-Attacks-pdf>

[5] Disponível em: <https://documents.trendmicro.com/assets/rpt/ransomware-in-q4-2022.pdf>

[6] virusshare.com

Tabela 1 – Amostras de ransomware: hashs e VirusShare.

Nome	sha256 Hash	Data do binário	Local
Conti	d826f4cb8240f894e43fea3c84b14fd85be9758d7ad4eafa113ad7d45c30bc26	2022-12	VirusShare
WannaCry	ed01ebfbc9eb5bbea545af4d01bf5f1071661840480439c6e5babe8e080e41aa	2017-05	VirusShare
Ryuk	dd4b13c694ca9d78d2a804149ac1919ee954564871e08c7f89c855d82c6c909a	2019-10	VirusShare
Cryptolocker	aa5272e0d783efba7b2903f1d2e64ce9a75eb7b2a073188b9dc7198bf6be4651	2017-03	VirusShare
Lockbit	fef1f9664fde9b23754c691b15a05fdc35a51a0ceb8a18fb9a5a0166e6377c69	2023-02	VirusShare
Sodinokibi	349bdb12a75fbfc2803f988862764ba6058b371728930f8dcb248f105ce607f7	2023-02	VirusShare

em comparação com o conjunto de dados empregado no trabalho atual.

Tabela 2 – Comparação dos conjuntos de dados em cada trabalho.

Pesquisa	Focado em Ransomware	Recência	Disponibilidade	Detalhes dos Binários	Análise de Chamadas de Sistema	Cuckoo Sandbox
(HERRERA-SILVA; HERNáNDEZ-ÁLVAREZ, 2023)	✓	×	✓	✓	✓	✓
(SGANDURRA et al., 2016)	✓	×	✓	✓	✓	✓
(RIECK et al., 2011)	×	×	✓	×	✓	×
(CARRIER et al., 2022)	×	×	✓	×	×	×
(HIRANO; HODOTA; KOBAYASHI, 2022)	✓	×	×	✓	×	×
(SCAIFE et al., 2016)	✓	×	×	×	✓	✓
(TANG et al., 2020)	✓	×	×	×	✓	×
(BERRUETA et al., 2020)	✓	×	✓	✓	×	×
(CONTINELLA et al., 2016)	✓	×	✓	×	✓	×
(KHARAZ et al., 2016)	✓	×	×	×	✓	✓
RansomSet	✓	✓	✓	✓	✓	✓

A Tabela 2 apresenta uma comparação detalhada das principais características dos conjuntos de dados utilizados nos estudos mencionados, incluindo o conjunto de dados do presente trabalho. Essa tabela foi criada para destacar as diferenças e semelhanças entre as abordagens adotadas pelos diferentes estudos na análise e detecção de *ransomwares*. As colunas da tabela foram organizadas de forma a incluir informações cruciais, como o tipo de análise, os tipos de *malware* estudados, o número de amostras utilizadas e as características específicas de cada conjunto de dados.

- **Pesquisa**: Nome ou referência do estudo sobre *ransomware* que propõe um conjunto de dados específico para análise.
- **Focado em Ransomware**: Indica se estudo é focado em *ransomwares* ou em outras formas de *malware*.
- **Recência**: Avalia o quão atuais são os binários utilizados na pesquisa.
- **Disponibilidade**: Indica se o conjunto de dados proposto por cada trabalho está disponível publicamente.
- **Detalhes dos Binários**: Fornece informações sobre o binário analisado, incluindo a data de coleta, o local da coleta, o nome do binário e a *hash* associada.
- **Análise de Chamadas de Sistema**: Indica se o estudo realiza a análise das chamadas de sistema nos binários, avaliando o comportamento do *ransomware* durante a execução.

- **Cuckoo Sandbox**: Especifica se o estudo utiliza o *Cuckoo Sandbox* como método de coleta de atributos para o conjunto de dados.

O RansomSet destaca-se em relação a outros trabalhos por seu foco nos binários de *ransomware* mais utilizados atualmente por cibercriminosos e pela ênfase na análise de chamadas de sistema, uma característica essencial para capturar o comportamento malicioso em tempo real. Essa abordagem diferencia-se significativamente de estudos anteriores, como os de Silva e Álvarez (HERRERA-SILVA; HERNáNDEZ-ÁLVAREZ, 2023), *EldeRan* (SGANDURRA et al., 2016), *CryptoLock* (SCAIFE et al., 2016) e *Unveil* (KHARAZ et al., 2016), que apresentam limitações evidentes.

Essas comparações evidenciam a diversidade de abordagens na pesquisa de detecção de *ransomware*, em que cada estudo ajusta seu conjunto de dados e metodologia conforme o foco da análise e os recursos disponíveis. A escolha do conjunto de dados e das ferramentas de análise não só influencia os resultados obtidos, mas também define a capacidade de comparação entre diferentes técnicas de detecção. O presente trabalho, por exemplo, se concentra nos *ransomwares* mais recentes e sofisticados, diferentemente de estudos anteriores que utilizaram amostras antigas e mais facilmente detectáveis por plataformas de *antimalware* e antivírus atuais, comprometendo a eficácia da análise. A importância de se trabalhar com métodos e amostras atualizadas reside na complexidade crescente e na evolução constante das técnicas utilizadas pelos *ransomwares* modernos.

Embora o número de *ransomwares* utilizados na pesquisa seja reduzido, a inclusão de amostras mais recentes e avançadas, disponibilizadas na plataforma *VirusShare*, permite uma investigação detalhada e alinhada com as ameaças contemporâneas, resultando em uma análise mais direcionada e relevante. A padronização e a disponibilização pública de conjuntos de dados, como proposto em alguns trabalhos, são essenciais para avançar na eficácia das soluções contra *ransomwares*. Isso possibilita que as comunidades de pesquisa avaliem e comparem seus métodos de maneira mais consistente e transparente, permitindo uma resposta mais eficaz às ameaças emergentes.

3 Ransomset

Neste capítulo, são apresentados e discutidos detalhadamente o processo de construção do conjunto de dados, o fluxo completo de coleta de informações que permitiu sua elaboração e o processo de instalação e configuração do ambiente. Primeiramente, descreve-se o fluxo de coleta de atributos realizado por meio do *Cuckoo Sandbox*, utilizado para monitorar e registrar o comportamento dos binários de *ransomware* em ambiente controlado. Esse processo inclui a captura de características dinâmicas das amostras, como chamadas de sistema, acessos a arquivos e interações de rede, essenciais para a detecção de padrões específicos de comportamento malicioso.

Em seguida, são abordadas as etapas de seleção, classificação e extração das amostras coletadas, detalhando o critério de escolha das amostras e como foram organizadas para compor o conjunto de dados. Esse processo de seleção é essencial para garantir que as amostras sejam representativas dos diferentes tipos de *ransomware* estudados e inclui uma análise cuidadosa dos atributos que melhor descrevem o comportamento malicioso. Por fim, discute-se a importância da classificação e filtragem dessas amostras, assegurando que o conjunto de dados final possua uma composição equilibrada de amostras benignas e maliciosas, permitindo uma avaliação confiável e robusta nos experimentos de detecção de *ransomware*.

3.1 Visão Geral

Esta seção apresenta uma visão abrangente do conjunto de dados denominado RansomSet, desenvolvido com o objetivo de fornecer uma base robusta para a análise comportamental de *ransomwares* e amostras benignas. O RansomSet foi projetado para representar os comportamentos de 6 variantes de *ransomwares*, incluindo *WannaCry*, *Ryuk*, *CryptoLocker*, *LockBit*, *Sodinokibi* e *Conti*, além de comportamentos típicos de *softwares* benigno. A criação desse conjunto de dados teve como finalidade contemplar diferentes classes de *ransomware* com foco na caracterização de suas atividades por meio de chamadas de sistema, com o intuito de aprimorar a precisão na classificação e seleção de atributos em variados cenários de análise.

Para a coleta e análise de dados comportamentais, utilizou-se a ferramenta *Cuckoo Sandbox*, que permite monitorar o comportamento dos binários de *ransomware* e das amostras benignas em um ambiente virtualizado controlado. Nessa configuração, os comportamentos observados foram registrados em arquivos no formato *JSON*, fornecendo uma base de dados rica em informações e adequada para a aplicação de técnicas de Análise Exploratória de Dados (EDA). O uso do *Cuckoo Sandbox* possibilitou a captura detalhada de interações com o sistema, como chamadas de API, acessos a arquivos e atividades de rede, aspectos essenciais para a identificação de padrões maliciosos específicos e desvios em relação ao comportamento benigno.

A estrutura do RansomSet permite incluir múltiplas classes de *ransomware* e amostras benignas, organizadas de modo a facilitar a análise comparativa entre comportamentos maliciosos e normais. Cada classe representa um tipo específico de atividade, proporcionando uma visão detalhada das características de cada tipo de comportamento. Esse conjunto de dados foi estruturado para auxiliar na compreensão das propriedades únicas dos *ransomwares* e suas distinções em relação ao comportamento benigno, oferecendo uma base sólida para o desenvolvimento de algoritmos de detecção e classificação.

No presente estudo, o conjunto de dados *RansomSet* foi desenvolvido com base em 30 execuções de cada binário capturado, abrangendo um conjunto de 26 binários benignos e 6 das famílias de *ransomware* mais amplamente utilizadas por cibercriminosos, totalizando 960 execuções. Esse volume de execuções permite uma captura abrangente do comportamento dos binários, proporcionando uma base de dados robusta e representativa para a análise e o estudo de detecção de *ransomware*. A inclusão de um número significativo de execuções e atributos no conjunto de dados é essencial, pois aumenta a quantidade e a diversidade de amostras disponíveis, ampliando a abrangência das análises e fornecendo uma base sólida para a criação de modelos de detecção de ameaças.

Dessa forma, o RansomSet constitui uma contribuição significativa para a pesquisa em segurança cibernética, oferecendo uma abordagem estruturada para o estudo de ameaças emergentes. A organização cuidadosa do conjunto de dados possibilita não apenas a detecção de ameaças de *ransomware*, mas também o aprimoramento das capacidades de resposta e defesa contra essas ameaças digitais.

3.2 Ambiente

Esta seção apresenta a configuração do ambiente de análise utilizado para monitorar e capturar o comportamento de diferentes variantes de *ransomware* e amostras benignas. Para garantir a segurança e precisão nas análises, foi criado um ambiente virtualizado que possibilita a execução controlada de cada amostra, minimizando riscos de contaminação ao sistema *host* e permitindo o isolamento adequado de cada execução.

A ferramenta principal utilizada para essa análise é o *Cuckoo Sandbox*, um software de código aberto amplamente utilizado na análise automatizada de *malwares*. O *Cuckoo Sandbox* foi escolhido devido à sua capacidade de monitorar interações detalhadas dos binários com o sistema, capturando eventos como chamadas de API, acessos a arquivos e conexões de rede, que são essenciais para a identificação de padrões e desvios comportamentais. A instalação foi realizada em uma máquina com sistema operacional *Ubuntu* 18.04, que oferece um ambiente estável e compatível com as dependências necessárias para o funcionamento do *Cuckoo*.

Para o ambiente virtual, foi utilizado o *VirtualBox* para a criação de uma *Virtual Machine* (VM) com *Windows* 7 de 64 bits, onde os binários foram executados. Essa configuração

foi escolhida devido à compatibilidade do *Cuckoo Sandbox* com sistemas operacionais mais antigos, como o *Windows* 7, garantindo maior confiabilidade na análise. O modelo de instalação estável do *Cuckoo* recomenda o uso desse sistema operacional, pois ele apresenta melhor desempenho na captura de dados em comparação com sistemas mais recentes, que podem interferir no processo devido a mudanças em sua arquitetura de segurança.

A utilização do *Windows* 7 também permitiu a observação detalhada das interações dos binários, sem comprometer o sistema principal. A comunicação entre o *Cuckoo Sandbox* e a VM foi estabelecida por meio de uma rede privada (*Host-Only Network*), possibilitando o controle total das operações de rede realizadas durante as execuções. Essa abordagem garantiu a coleta de dados consistente e detalhada, essencial para o estudo.

Cada execução foi monitorada por um agente `agent.py`, que capturou informações detalhadas e as enviou ao servidor do *Cuckoo*. Os dados coletados foram armazenados em arquivos no formato *JSON*, permitindo um registro completo das interações de cada amostra com o sistema. A Figura 2 apresenta a o fluxo básico que representa como o ambiente foi montado para o trabalho. Na Seção 3.3 são apresentados mais detalhes de todo o fluxo pelo qual o binário é processado.

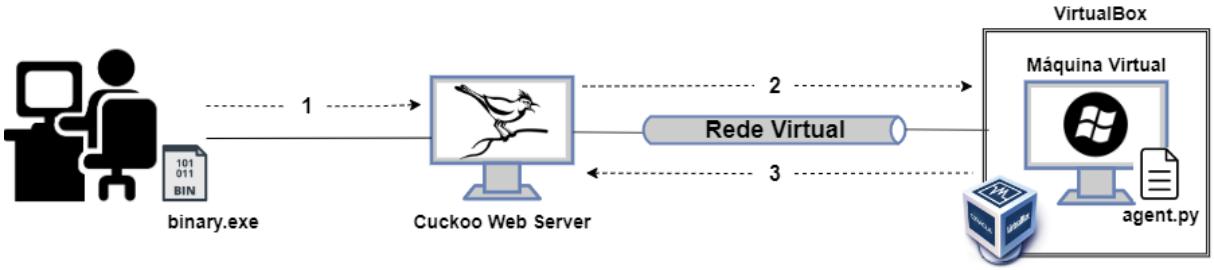

Figura 2 – Fluxo de básico do software Cuckoo Sandbox. Extraído de (SOUZA et al., 2023).

Dessa forma, o ambiente configurado permitiu a captura precisa dos comportamentos observados nas amostras, oferecendo uma base sólida para o desenvolvimento e aprimoramento de métodos de detecção e análise de *ransomware*. A estrutura organizada dos dados viabiliza futuras análises e contribui para o avanço das capacidades de defesa contra ameaças cibernéticas.

3.3 Coleta de Dados

O estudo propõe a criação de um conjunto de dados multiclasse expandido, abrangendo não apenas novas classes de *ransomware*, mas também novos binários benignos. Esse conjunto de dados visa fornecer uma análise e avaliação mais abrangentes, considerando as informações obtidas durante as execuções e a coleta de dados de cada binário utilizado no estudo. Enriquecer esses conjuntos de dados permite uma visão mais ampla e detalhada da análise de *ransomware*, da pesquisa científica e da detecção precoce em sistemas operacionais.

O processo de coleta de dados foi planejado para garantir a qualidade das amostras benignas e maliciosas. Foi utilizado o *Cuckoo Sandbox*, uma renomada ferramenta de análise automatizada de *malware*, para examinar os binários. Cada binário foi executado trinta vezes, com duração máxima de dez minutos, proporcionando uma análise detalhada de seu comportamento e interações. O ambiente controlado do *Cuckoo* permitiu uma visão precisa das ações dos binários, essenciais para a pesquisa. A Figura 3 descreve o cenário experimental, destacando o upload de binários na ferramenta e a configuração de máquinas virtuais via *VirtualBox*. O *Cuckoo Sandbox* utiliza um agente *Python* para monitorar e registrar todas as ações e modificações dos binários durante a execução (SOUZA et al., 2023). O ambiente de análise foi implementado em um Dell Inspiron 15 3000 com *Ubuntu* 18.04.

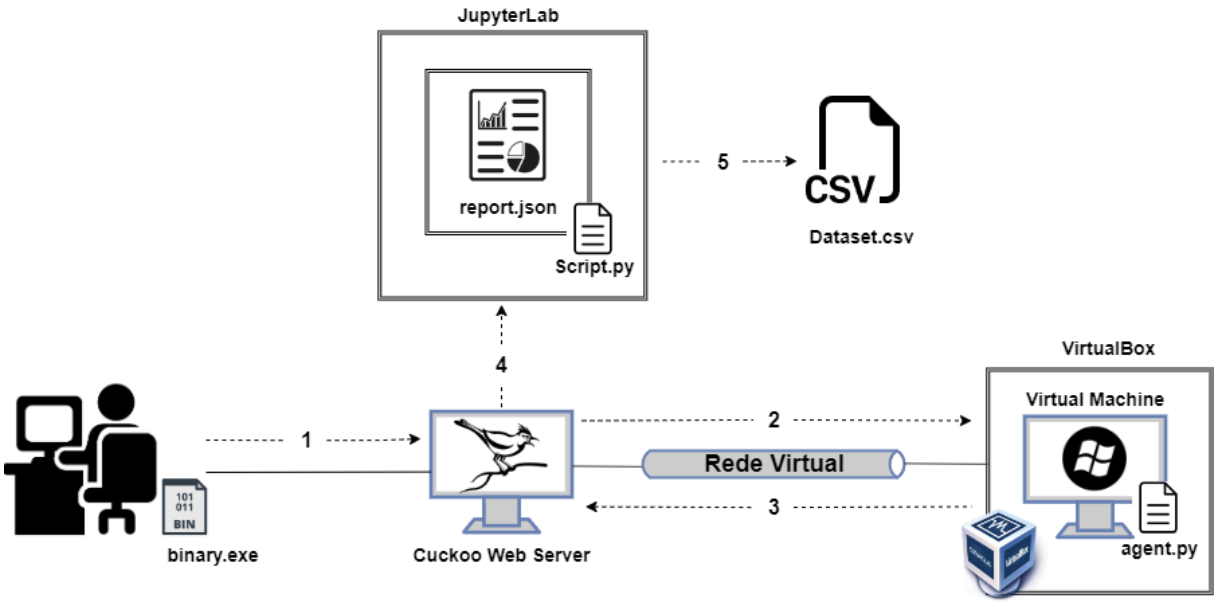

Figura 3 – Fluxo de trabalho de extração de dados. Extraído de (SOUZA et al., 2023).

Os passos ilustrados no cenário da Figura 3 demonstram a metodologia empregada na execução dos experimentos realizados. No Passo 1, o usuário insere o binário a ser analisado, definindo as especificações de tempo de execução e permissão de acesso à internet para outras ações durante a execução. Em seguida, no Passo 2, a aplicação utiliza a *Virtual Network* para enviar e inicializar a execução do binário na máquina virtual configurada para que o *Cuckoo Sandbox* realize sua análise. Durante a execução do binário, um programa chamado `agent.py` é executado dentro da máquina virtual, capturando todas as ações do binário.

Após a conclusão da execução do binário, no Passo 3, o agente `agent.py` captura todas as informações relevantes que ocorreram durante a execução e as envia para a aplicação por meio da *Virtual Network*. O *Cuckoo Sandbox* coleta todas essas informações e as armazena em um arquivo no formato *JavaScript Object Notation* (JSON), que é enviado no Passo 4 para a plataforma *JupyterLab*. Nessa plataforma, foi implementado um código que realiza a coleta e

filtragem das informações relevantes para o ranqueamento dos atributos, conforme descrito neste estudo.

Para processar o relatório resultante da análise do *Cuckoo*, foi implementado um *script* na linguagem *Python* chamado `script.py`, disponível no repositório mencionado. Esse *script* é responsável pela criação de um arquivo de *Comma-separated values* (*CSV*) que contém todos os atributos filtrados, conforme representado no Passo 5 da imagem.

O conjunto de dados *RansomSet* foi desenvolvido a partir de múltiplas execuções de cada amostra coletada durante a pesquisa, incluindo binários benignos e 6 das famílias de *ransomware* mais utilizadas por cibercriminosos: *WannaCry*, *Ryuk*, *CryptoLocker*, *LockBit*, *Sodinokibi* e *Conti*. Essa abordagem proporciona uma maior diversidade de dados e amplia a objetividade nas análises de comportamento dos binários. Ao incluir várias execuções para cada amostra, o conjunto de dados se torna mais representativo, permitindo uma análise detalhada e abrangente das características associadas a cada tipo de *ransomware*.

Por fim, utilizou-se o algoritmo *Information Gain* (*IG*) para medir a representatividade de cada um dos atributos gerados, descartando os menos representativos. Após a análise, iniciou-se o processo de rotulagem das amostras. As amostras foram classificadas como benignas, representando a classe normal, ou maliciosas, identificadas pelo nome de cada binário específico. Esse procedimento foi fundamental para organizar e categorizar corretamente as amostras coletadas, facilitando uma análise mais estruturada e criteriosa.

3.3.1 Seleção de amostras benignas e maliciosas

A coleta de amostras benignas e maliciosas é uma etapa fundamental na pesquisa de segurança cibernética, pois fornece os dados necessários para a análise e compreensão das ameaças em questão. Amostras maliciosas foram obtidas da plataforma *VirusShare*, escolhida por sua extensa coleção de *malware*. As amostras de *ransomware* selecionadas para o estudo totalizam 6 binários, incluindo: *Sodinokibi*, *Conti*, *WannaCry*, *Ryuk*, *Cryptolocker* e *Lockbit*.

As amostras benignas foram coletadas por meio da plataforma *Software Informer*, selecionada para garantir uma variedade de *software* comumente utilizado. O *software* benigno coletado totaliza 23 executáveis, incluindo: *Audacity*, *AVG*, *Bible Verse*, *BlueStacks*, *BrazosTweaker*, *CCleaner*, *EasyBits*, *FastStone Capture*, *Kaspersky*, *Legacy*, *McAfee*, *MPC-HC*, *.NET Framework*, *Notepad++*, *PDF Reader*, *PhotoInstrument*, *QuickTime*, *Streets of Rage*, *Sublime Text*, *Symtb*, *Thunderbird*, *Tor Browser* e *UC Browser*.

Para coletar amostras maliciosas, foi realizado um registro na plataforma *VirusShare* e cada *ransomware* estudado neste artigo foi pesquisado na plataforma. Para as amostras benignas, foram coletados diversos instaladores de *softwares*. Essa abordagem permitiu obter uma variedade de *softwares* comumente instalados, fornecendo uma ampla gama de cenários para as análises.

Amostras de *ransomware* foram selecionadas dos repositórios de *malware VirusShare*[1], priorizando as versões mais recentes de cada amostra, enquanto binários benignos foram coletados por meio da plataforma *Software Informer*[2]. As amostras benignas seguem um padrão de aplicativos comuns amplamente utilizados por um grande número de usuários. Priorizou-se a seleção de binários benignos representativos, cujo comportamento seja o mais semelhante possível aos binários maliciosos, visando melhorar o processo de aprendizado. Tanto os binários maliciosos quanto os benignos seguiram o mesmo processo no *Cuckoo Sandbox*, sendo executados trinta vezes, com um tempo de execução definido de dez minutos para cada binário.

A lista completa, juntamente com os resultados apresentados na Seção 3.3.3, os binários, conjuntos de dados e *scripts* utilizados nos experimentos, tanto benignos quanto maliciosos, está disponível em um repositório hospedado na plataforma[3].

3.3.2 Coleta de atributos no Cuckoo Sandbox

Quando uma amostra de *ransomware* é executada dentro do ambiente *Cuckoo Sandbox*, diversas informações são registradas em um arquivo *JSON*, fornecendo uma visão geral do comportamento do *ransomware* durante sua execução (HERRERA-SILVA; HERNáNDEZ-ÁLVAREZ, 2023).

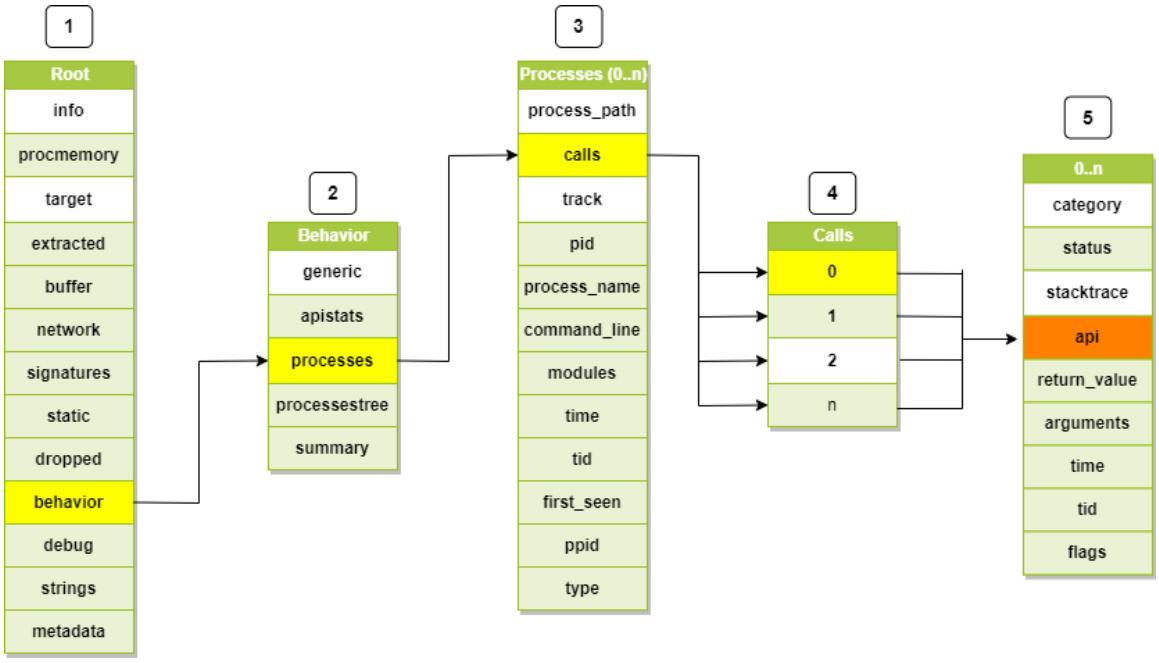

Figura 4 – Estrutura do arquivo *JSON* gerado pelo *Cuckoo Sandbox*.

A estrutura hierárquica do arquivo *JSON*, ilustrada na Figura 4, é organizada em seções específicas para diferentes conjuntos de dados. Cada seção é claramente rotulada, facilitando a

[1] Disponível em: <https://virusshare.com/>
[2] Disponível em: <https://software.informer.com/software/>
[3] Disponível em: <https://github.com/gabrielolivs/RansomSet>

identificação e interpretação de informações relacionadas à análise de *malware*. Neste estudo, chamadas de sistema foram utilizadas como atributos para analisar o comportamento do *ransomware* (ILIĆ et al., 2022)(HERRERA-SILVA; HERNáNDEZ-ÁLVAREZ, 2023). Na seção raiz, no bloco 1, a seção de comportamento foi selecionada para capturar os dados. Dentro desta seção, os processos no bloco 2 armazenam as execuções realizadas por cada binário durante sua execução no ambiente *Cuckoo Sandbox*. Além disso, há uma subseção chamada *calls* no bloco 3, que contém todas as chamadas de sistema feitas pelo binário durante sua execução.

Cada chamada de sistema na seção *calls* é identificada por um inteiro de 0 a N no bloco 5, onde N representa o número total de chamadas de sistema realizadas pelo binário. Esse valor pode ser significativamente alto, ultrapassando 1000 ou até 10000 chamadas. As chamadas de sistema, também conhecidas como *calls*, são instruções enviadas por um programa ao sistema operacional para solicitar recursos específicos, como acessar arquivos, estabelecer conexões de rede ou manipular (CHEN et al., 2017). Dentro de cada chamada de sistema, os nomes das APIs invocadas durante a operação do *malware* são registrados. Essas *calls* são capturadas pelo *Cuckoo Sandbox* a partir de registros de baixo nível, incluindo interações como criação de processos, acessos a arquivos, conexões de rede e interações com o registro do Windows.

3.3.3 Extração de atributos

Para processar o relatório de análise do *Cuckoo*, foi desenvolvido um *script* em *Python* chamado `script.py`, disponível no repositório *GitHub* associado a este artigo. Esse *script* é responsável pela criação de um arquivo *CSV* contendo todos os atributos filtrados, conforme ilustrado na etapa 5 da Figura 4.

Durante a implementação do *script*, duas funções principais foram definidas: `CapturaColuna` e `jsonarq`. A função `CapturaColuna`, relacionada ao Algoritmo 1, processa arquivos *JSON*, extrai informações comportamentais de binários maliciosos e identifica diferenças nas chamadas de *API* feitas por esses binários. Essa função tem como objetivo coletar uma lista de chamadas de *API* distintas utilizadas por binários maliciosos para análises subsequentes.

A função `jsonarq`, demonstrada no Algoritmo 2, cria ou atualiza um arquivo *CSV* com base nos dados coletados durante a análise comportamental dos binários. Essa função abre ou cria um arquivo *CSV* no diretório especificado e registra informações sobre os binários processados. Durante o processamento, extrai informações como *APIs* e classes binárias (consideradas benignas se não estiverem na lista predefinida de *ransomware*), contabiliza o número de vezes que cada chamada de *API* aparece na lista de colunas e registra essas informações no arquivo *CSV*, incluindo classe, resultados e o número de chamadas de *API* para cada binário analisado.

A descrição completa do *script* e suas funções está disponível no repositório *GitHub*[4] associado a este artigo, onde cada função é explicada em detalhes.

[4] Disponível em: <https://github.com/gabrielolivs/RansomSet>

Algoritmo 1 Captura de Colunas de atributos

```
Entrada: Lista de arquivos dir_list
Saída: Lista de colunas de atributos colunas
colunas ← []                                          ▷ Inicializa a lista de colunas
for arquivo in dir_list do
    data ← Ler JSON do arquivo arquivo
    for i in data['behavior']['processes'] do
        for j in i['calls'] do
            api ← j['api']
            if api ∉ colunas then
                Adicionar api a colunas
            end if
        end for
    end forjson.decoder.JSONDecodeError
    Imprimir "Erro ao decodificar JSON no arquivo " + arquivo
end for
return colunas                                        ▷ Retorna as colunas de atributos
```

Algoritmo 2 Processamento de arquivos *JSON* e gravação em *CSV*

```
Entrada: Diretório de arquivos dir_arq, colunas colunas
Saída: Arquivo CSV com dados processados
C ← Criar ou abrir arquivo CSV                        ▷ Abre o arquivo para escrita
Escrever linha de cabeçalho em C com ['timestamp', 'classe', 'score_binary'] + colunas
for arquivo in dir_arq do
    Abrir arquivo em modo leitura
    data ← Ler JSON do arquivo
    file_name ← data['target']['file']['name']
    if file_name ∈ {WannaCry, Ryuk, LockBit, Conti, Sodinikibi, CryptoLocker} then
        classe ← file_name
    else
        classe ← Normal
    end if
    score ← data['info']['score']
    for process in data['behavior']['processes'] do
        timestamp ← process['time']
        calls ← [call['api'] for call in process['calls']]
        numeros ← [calls.count(coluna) for coluna in colunas]
        Escrever linha em C com [timestamp, classe, score] + numeros
    end for
end for
```

A criação de conjuntos de dados é essencial para tarefas de detecção de *malware*, fornecendo conjuntos estruturados e detalhados que podem ser utilizados para treinar e avaliar modelos de classificação. Neste trabalho, a proposta de um conjunto de dados multiclasse é particularmente relevante, organizando informações de diferentes tipos de *malware* em um único conjunto. Essa abordagem multiclasse enriquece o conjunto de dados com exemplos e informações detalhadas sobre o comportamento de cada tipo de *malware*, aumentando a precisão e a eficiência dos algoritmos de detecção e classificação. Ao integrar várias características e tarefas em um único conjunto de dados, esta proposta estabelece uma base sólida para o desenvolvimento de algoritmos avançados para identificar e prevenir ameaças à segurança cibernética.

Na Tabela 3, são listados os 25 atributos classificados pelo algoritmo de ranqueamento

IG. Cada atributo está relacionado ao seu respectivo nome, acompanhado de uma breve descrição sobre sua função ao ser executado. A coluna de "Peso *IG*" demonstra o valor de *IG* para cada atributo (chamada de sistema), sendo que os atributos com maior peso têm impacto mais significativo na detecção, contribuindo para aumentar a precisão do modelo de classificação. A coluna denominada "Índice" organiza os atributos conforme sua posição no arquivo *CSV*, indicando a relevância de cada chamada de sistema na análise.

Tabela 3 – Lista de chamadas de sistema.

Índice	**Atributo**	**Descrição**	**Peso IG**	**IG *Rank***
1	score_binary	Função de pontuação binária	1,327	1
6	NtClose	Fecha um objeto	1,239	2
17	NtOpenKey	Abre uma chave de registro	1,056	3
18	NtQueryValueKey	Lê o valor de uma chave de registro	1,007	4
3	NtAllocateVirtualMemory	Aloca memória virtual	0,965	5
20	LdrGetProcedureAddress	Obtém o endereço de uma função de uma DLL	0.755	6
57	LdrGetDllHandle	Obtém o handle de uma DLL	0,671	7
54	GetSystemTimeAsFileTime	Obtém a hora do sistema	0,621	8
67	NtEnumerateValueKey	Enumera os valores de uma chave de registro	0,564	9
19	LdrLoadDll	Carrega uma DLL	0,562	10
135	SetUnhandledExceptionFilter	Define um filtro para exceções não tratadas	0,475	11
4	NtCreateFile	Cria ou abre um arquivo	0,418	12
27	RegOpenKeyExW	Abre uma chave de registro	0,414	13
58	SetErrorMode	Define o modo de erro do sistema	0,390	14
29	RegCloseKey	Fecha uma chave de registro	0,375	15
23	NtMapViewOfSection	Mapeia uma seção na memória	0,371	16
8	FindFirstFileExW	Procura o primeiro arquivo em um diretório	0,367	17
28	RegQueryValueExW	Lê o valor de uma chave de registro	0,360	18
7	NtFreeVirtualMemory	Libera memória virtual alocada	0,342	19
5	NtReadFile	Lê dados de um arquivo	0,340	20
66	NtEnumerateKey	Enumera chaves de registro	0,336	21
10	GetFileAttributesW	Obtém os atributos de um arquivo	0,321	22
46	NtUnmapViewOfSection	Desmapeia uma seção da memória	0,318	23
38	LdrUnloadDll	Descarrega uma DLL	0,308	24
22	NtCreateSection	Cria uma seção de memória	0,308	25

3.4 Descrição

O conjunto de dados gerado durante o desenvolvimento da pesquisa possui uma estrutura organizada para facilitar a compreensão e utilização das informações. A Tabela 3 apresenta uma lista detalhada dos atributos, ordenados com base no algoritmo *IG*. O algoritmo *IG* é uma técnica de seleção de atributos usada em aprendizado de máquina, baseada na teoria da informação, que quantifica a redução da incerteza ou entropia ao dividir os dados com base em um atributo específico (AZHAGUSUNDARI; THANAMANI et al., 2013). Ao avaliar o *IG* de diferentes atributos, identifica-se quais características têm maior impacto ou contribuição para a tarefa de classificação, auxiliando na criação de modelos mais precisos e eficientes. Portanto, a tabela fornece uma visão abrangente dos atributos mais influentes de acordo com o critério de ganho de informação (AZHAGUSUNDARI; THANAMANI et al., 2013)(QUINLAN, 1986). A tabela

completa contendo todos os atributos estudados neste artigo está disponível no repositório do *GitHub*. A Figura 5 demonstra a estrutura do arquivo: cada linha representa uma observação de cada *ransomware*; uma quantidade de linhas pode representar uma única execução, ou seja, a execução de um *ransomware* específico, enquanto as colunas representam as características coletadas durante essa execução. Entre as colunas, estão incluídos campos como pontuações, classes de *ransomwares* (quando aplicável), detalhes sobre as chamadas de sistema realizadas pelos binários, entre outros aspectos relevantes para a identificação e classificação das ameaças.

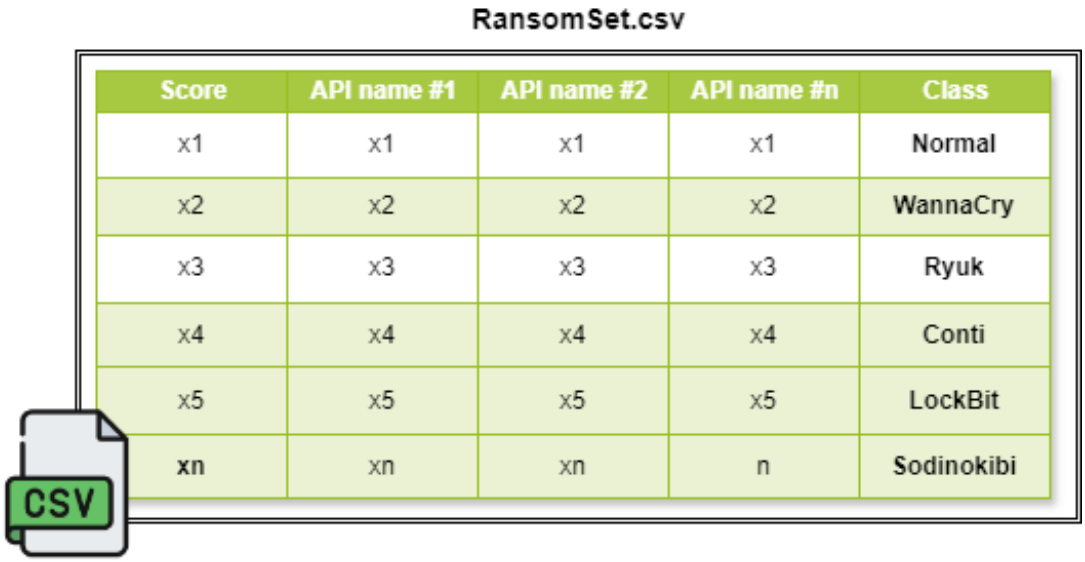

Figura 5 – Estrutura do conjunto de dados RansomSet.

O conjunto de dados atual é composto por um total de 19.235 linhas referentes a chamadas de sistema de cada execução do binário e 240 colunas de informações relevantes de cada atributo coletado. Esse conjunto de dados é fundamental para a pesquisa e o desenvolvimento de estratégias de detecção de ameaças. A estruturação cuidadosa e abrangente do conjunto de dados permite uma análise aprofundada do comportamento do *ransomware*, facilita a identificação de padrões e características únicas e prepara os dados para a aplicação de técnicas de aprendizagem automática e outras abordagens de análise. Essa organização bem definida é essencial para investigadores, cientistas de dados e profissionais de segurança da informação, fornecendo recursos inestimáveis na luta contra ameaças cibernéticas.

3.4.1 Sumário Estatístico

O gráfico apresentado na Tabela 4, que exibe o número de classes geradas a partir do conjunto de dados de análise de *ransomware*, mostra a distribuição das classes. A análise revelou diversas categorias ou classes de ameaças, representadas por diferentes tipos de *ransomware*. A classe "normal"no conjunto de dados refere-se a binários que não realizam qualquer tipo de ação maliciosa. Esses binários representam o comportamento esperado e legítimo de sistemas

e aplicações, servindo como referência para distinguir atividades benignas de atividades potencialmente prejudiciais. A inclusão dessa classe é fundamental para a análise, pois permite identificar desvios e padrões que caracterizam os ataques, auxiliando no treinamento e avaliação de modelos de detecção de ameaças cibernéticas. A mais comum dessas categorias é a classe *"Conti"*, com um total de 14.010 observações. Em seguida, estão a classe "Normal" com 2.463 observações, *"WannaCry"* com 1.480 observações, *"Ryuk"* com 802 observações, "LockBit" com 300 observações e, por fim, a classe *"Sodinokibi"* com 179 observações. Esse gráfico de contagem de classes fornece uma visão panorâmica da distribuição dos tipos de *ransomware* no conjunto de dados, permitindo visualizar de forma rápida e eficaz a frequência de cada categoria de ameaça cibernética. Essa análise é importante para compreender a representação de cada tipo de *ransomware* no conjunto de dados, fornecendo informações valiosas para o desenvolvimento de estratégias de detecção e defesa contra essas ameaças digitais.

Tabela 4 – Distribuição de classes no conjunto de dados.

	Análise dos dados	
Nome	**Contagem**	**Porcentagem**
Conti	14.010	72,87%
Normal	2.463	12,79%
WannaCry	1.480	7,69%
Ryuk	802	4,18%
CryptoLocker	360	1,87%
LockBit	300	1,56%
Sodinokibi	179	0,93%

A ofuscação do *ransomware Conti* revela uma lista complexa de estratégias projetadas para dificultar a detecção e a análise por pesquisadores de segurança. Uma das táticas mais notáveis é a abordagem proprietária à criptografia de *strings*, na qual quase todas as *strings* de texto usadas pelo *ransomware* são inseridas em um algoritmo de criptografia específico. São aplicados 277 algoritmos diferentes, um para cada *thread*, dos quais cerca de 230 são armazenados em sub-rotinas especiais, o que aumenta consideravelmente o tamanho do código do programa. Essa complexidade é ilustrada por diagramas específicos. O aspecto mais notável dessa ofuscação é a ocultação de muitas das chamadas de *API* do *Windows* necessárias para o funcionamento do *ransomware*. Muitos malwares são rastreados por meio dessas chamadas de *API* enquanto estão em execução. No entanto, o *Conti* adota uma abordagem ativa e evasiva. Esse sistema avançado de ofuscação fortalece a resiliência do *Conti* contra métodos de detecção tradicionais e destaca a necessidade de abordagens inovadoras para proteção contra *ransomwares* cada vez mais sofisticados (VMWARE, 2020)(ALZAHRANI; XIAO; SUN, 2022).

3.4.2 Matrizes de correlação

A análise do conjunto de dados é essencial para compreender e extrair informações significativas que possam orientar a tomada de decisões de forma rápida e facilitada. Durante a análise, foram utilizadas várias ferramentas e métricas, cada uma oferecendo um tipo específico de informação sobre o conjunto de dados gerado.

A contagem de atributos, ilustrada na Tabela **??**, fornece uma visão detalhada da distribuição das classes em cada categoria listada no conjunto de dados. Esse aspecto é crucial para conjuntos de dados multiclasses, pois permite compreender a frequência de ocorrência de cada uma das classes abordadas na pesquisa, facilitando a visualização de como as diferentes categorias estão representadas.

A matriz de correlação é uma ferramenta estatística para avaliar a relação linear entre duas ou mais variáveis em um conjunto de dados. Representada por uma tabela simétrica, a matriz de correlação fornece informações sobre a direção e a força da relação entre as variáveis, auxiliando pesquisadores e analistas na identificação de padrões e tendências nos dados.

O coeficiente de correlação de *Pearson* é o método mais comumente utilizado para calcular os valores na matriz de correlação. Esse coeficiente varia entre -1 e 1, onde -1 indica uma correlação negativa perfeita, 1 indica uma correlação positiva perfeita, e 0 indica ausência de correlação linear entre as variáveis. Valores próximos de -1 ou 1 indicam uma relação linear forte entre as variáveis, enquanto valores próximos de 0 sugerem uma relação fraca ou inexistente. Além de avaliar a relação entre pares de variáveis, a matriz de correlação é útil para detectar multicolinearidade em modelos de regressão, identificando variáveis independentes que estão altamente correlacionadas entre si. A multicolinearidade pode distorcer os resultados de modelos estatísticos e tornar as estimativas dos coeficientes menos confiáveis (STEIGER, 1980).

A matriz ilustrada na Figura 6 examina a relação entre a execução de binários sem ações maliciosas explícitas e as funções que eles desempenham durante a análise no *Cuckoo Sandbox*. Os atributos considerados incluem comportamentos legítimos, como tentativas de comunicação com servidores de comando e controle, estabelecimento de conexões, leitura e escrita de dados. Além disso, a matriz avalia correlações com variáveis relacionadas às chamadas de *API* do sistema operacional, como a presença de código associado à detecção de atividades potencialmente maliciosas.

A análise da matriz de correlação dos binários benignos permite identificar padrões e obter insights valiosos sobre seus comportamentos e sua relação com atributos específicos monitorados no *Cuckoo Sandbox*. Valores de correlação próximos de 1 entre determinadas características podem indicar uma forte associação positiva, sugerindo que tais comportamentos, embora legítimos, possuem elementos semelhantes aos que podem ser explorados por *ransomwares*. Por outro lado, valores próximos de 0, como entre a presença de vulnerabilidades conhecidas e a eficácia das medidas de segurança, podem indicar ausência de correlação, revelando que fatores

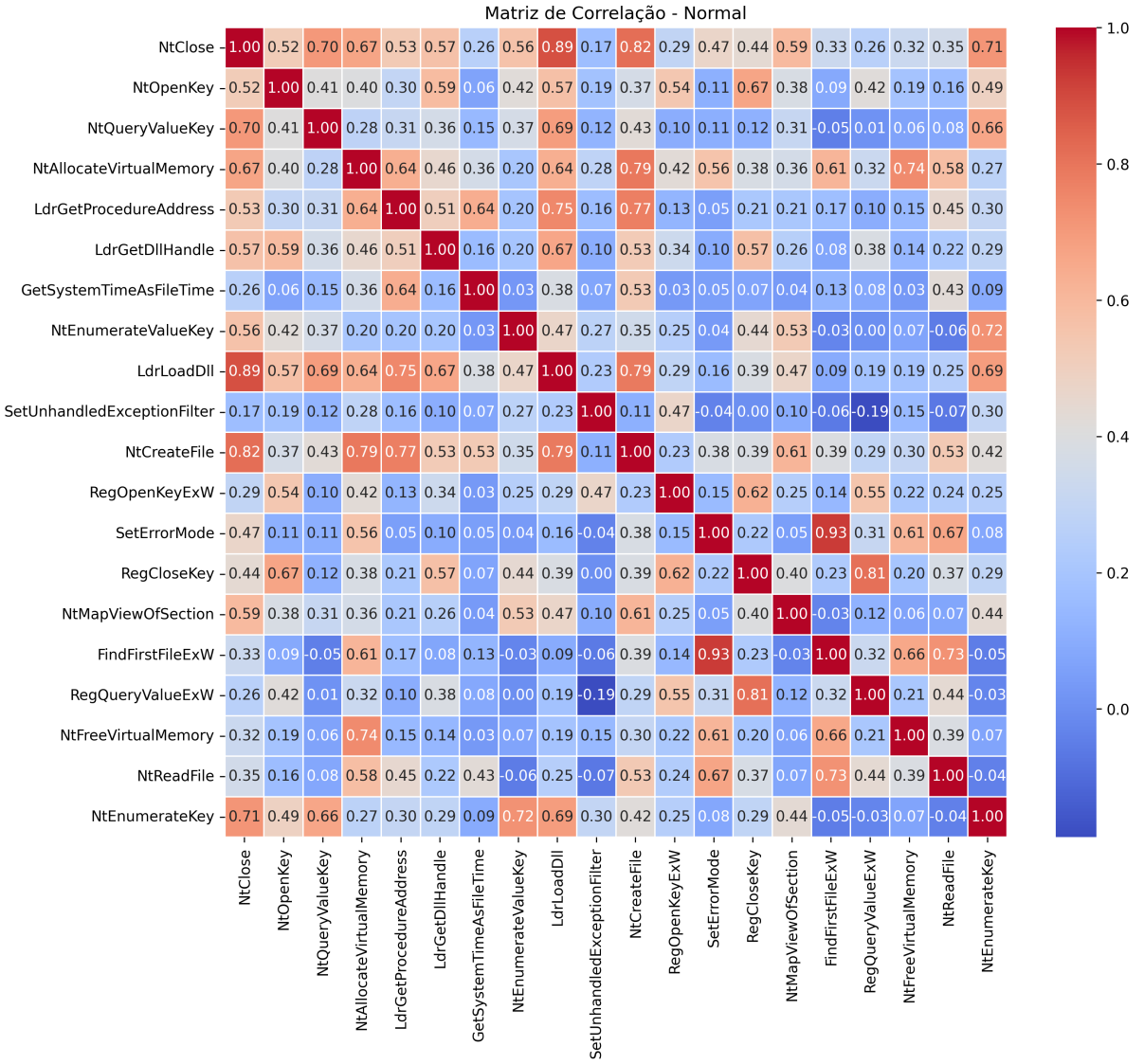

Figura 6 – Matriz de correlação da classe normal, relações entre chamadas de sistema.

adicionais não contemplados na análise podem exercer influência significativa na propagação de ameaças.

A análise da matriz de correlação da classe normal, apresentada na Figura 6, revela padrões de associação entre diferentes chamadas de sistema. Observa-se uma correlação forte entre *NtClose* e *NtOpenKey*, com valor de correlação de 0,52, indicando que essas funções frequentemente aparecem em conjunto nas operações analisadas. De maneira similar, funções relacionadas ao gerenciamento de memória, como *NtAllocateVirtualMemory* e *NtFreeVirtualMemory*, exibem correlação significativa, sugerindo um padrão de uso conjunto em processos normais do sistema.

Adicionalmente, a função *RegOpenKeyExW*, que interage com chaves de registro, apresenta correlações notáveis com várias outras funções de manipulação de registro, como *RegClo-*

seKey e *RegQueryValueExW*. Este comportamento reflete a natureza sequencial de operações de acesso ao registro em processos benignos, onde essas chamadas são realizadas para leitura e fechamento de chaves.

Outro ponto de destaque é a ausência de correlação forte entre funções de acesso a arquivos e chamadas relacionadas ao gerenciamento de memória. Esse padrão sugere que processos benignos mantêm uma separação entre operações de manipulação de arquivos e operações de alocação de memória. Em resumo, a matriz de correlação fornece uma visão detalhada dos padrões operacionais dos processos normais do sistema, permitindo a distinção de comportamentos entre classes maliciosas e não maliciosas com base nos atributos identificados.

A matriz ilustrada na Figura 7 investiga a relação entre a disseminação do *ransomware Conti* e as chamadas de sistema executadas durante seu processo de ação para enganar *anti-malware*s ou antivírus. Essas chamadas de sistema são projetadas para parecerem legítimas, o que dificulta a detecção e o bloqueio pelo software de segurança.

As chamadas fraudulentas de sistema são um atributo especial do *ransomware* Conti, permitindo que ele passe despercebido e comprometa de forma efetiva o sistema da vítima. Ao analisar a correlação entre essas chamadas de sistema e a disseminação do *Conti*, a matriz busca identificar padrões e tendências que possam auxiliar na compreensão do funcionamento do *ransomware* e no desenvolvimento de estratégias de detecção e prevenção mais eficazes (CSA, 2022). Um valor próximo de 1 entre certas chamadas de sistema específicas do *Conti* e a taxa de infecção pode indicar uma correlação positiva forte, sugerindo que esses métodos de evasão são altamente eficazes na disseminação do *ransomware*. Por outro lado, um valor próximo de -0,75 entre a presença de vulnerabilidades conhecidas nos sistemas afetados e a eficácia das medidas de segurança implementadas pode indicar uma correlação negativa forte, evidenciando que a presença de vulnerabilidades conhecidas está associada a uma menor eficácia das medidas de segurança contra o *Conti*.

A análise da matriz de correlação da classe *Conti*, ilustrada na Figura 7, evidencia padrões de correlação significativos entre diversas chamadas de sistema, refletindo as táticas de ofuscação e evasão empregadas por esse ransomware. Observa-se uma correlação extremamente alta entre *NtClose* e *NtOpenKey* (0,99), sugerindo uma frequência de uso conjunto que pode estar associada a manipulações repetidas de objetos e chaves de registro, o que é comum em operações maliciosas para manter persistência no sistema.

Funções de gerenciamento de memória, como *NtAllocateVirtualMemory* e *NtMapViewOfSection*, também apresentam alta correlação com outras chamadas de sistema, indicando uma possível manipulação estratégica da memória para dificultar a detecção. Esse comportamento é característico de *ransomwares* avançados que utilizam técnicas para alocar e proteger memória de forma a esconder sua presença.

Além disso, *LdrGetDllHandle* e *LdrLoadDll* exibem correlações elevadas, sugerindo um

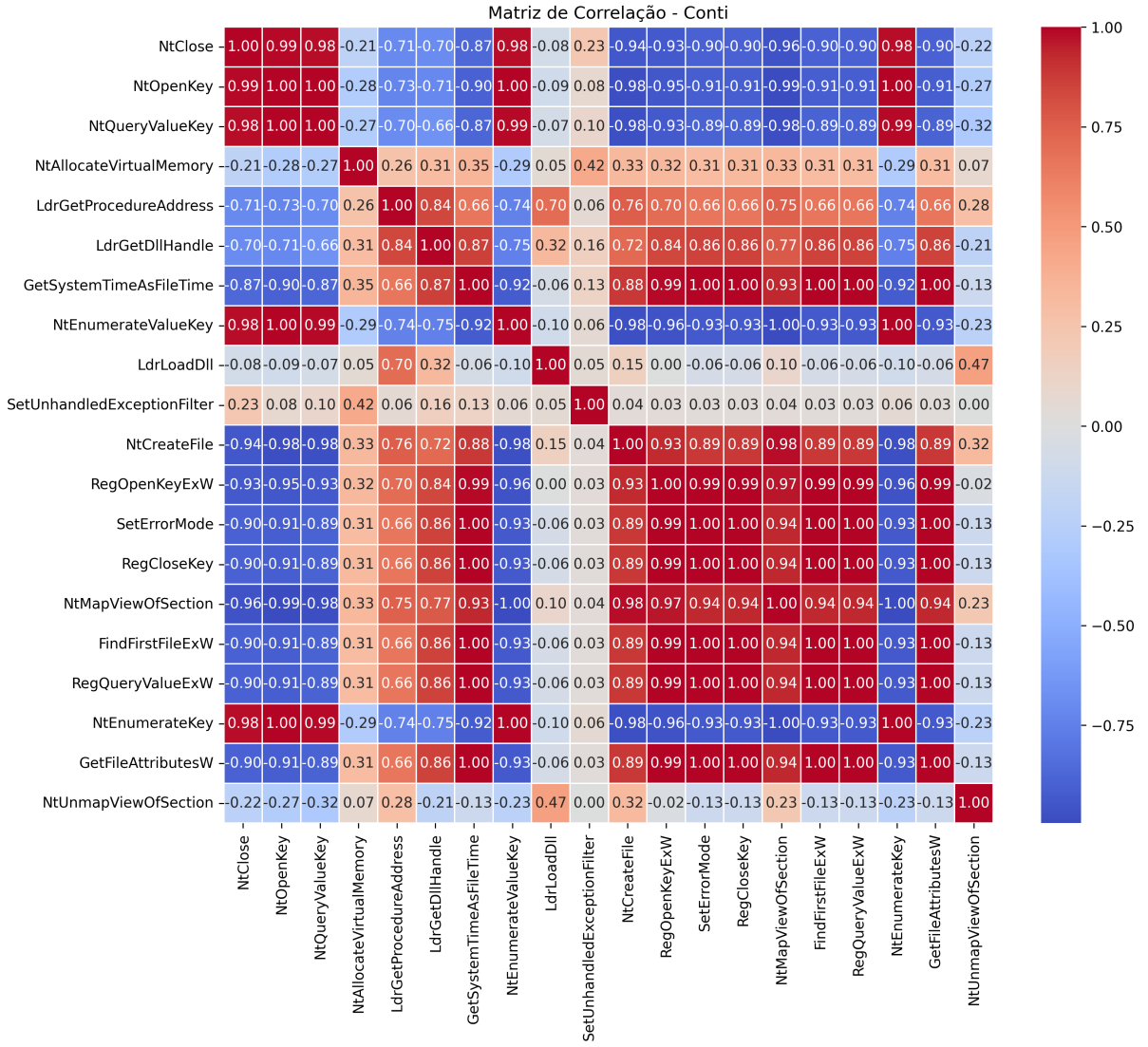

Figura 7 – Matriz de correlação da classe *Conti*, relações entre chamadas de sistema.

padrão de carregamento e descarregamento de bibliotecas dinâmicas durante a execução. Esse processo é uma estratégia conhecida para dificultar a análise, pois permite que o *ransomware* carregue e descarregue funcionalidades conforme necessário, evitando detecção contínua.

Esses padrões de correlação, distintos dos processos benignos, fornecem uma visão abrangente das estratégias de evasão e manipulação de sistema utilizadas pelo *Conti*. Tais informações são para aprimorar modelos de detecção que consigam identificar comportamentos anômalos associados a *ransomware*, promovendo uma resposta mais eficaz contra essas ameaças cibernéticas.

4 Experimentos

Este capítulo apresenta os experimentos realizados para validar e avaliar a eficiência do conjunto de dados desenvolvido, denominado *RansomSet*, no contexto de detecção de *ransomwares*. Os experimentos têm como objetivo identificar e analisar atributos relevantes para distinguir entre atividades maliciosas, especificamente de *ransomwares*, e atividades benignas. Para esse fim, foram configurados ambientes de teste controlados e foi aplicado um processo rigoroso de coleta e pré-processamento dos dados, seguido da seleção de atributos de impacto na classificação.

Inicialmente, descreve-se a estrutura do ambiente de experimentação, projetado para garantir a reprodutibilidade dos testes e o isolamento das execuções dos *ransomwares*. Em seguida, aborda-se o processo de coleta de dados, que utilizou a ferramenta *Cuckoo Sandbox* para registrar atividades comportamentais dos binários de *ransomware* e binários benignos. Esses dados brutos foram organizados e preparados para análise, assegurando a qualidade e consistência dos registros.

Para refinar a análise, aplicou-se a técnica de *IG*, permitindo identificar e ranquear os atributos mais relevantes. Essa etapa contribuiu para a criação de um modelo de classificação mais preciso, voltado à distinção eficaz entre atividades maliciosas e normais. A conclusão deste capítulo enfatiza os resultados obtidos com os experimentos e os benefícios da seleção estratégica de atributos para o aprimoramento de mecanismos de segurança cibernética e resposta a ameaças.

4.1 Coleta e Pré-processamento dos Dados

Para este estudo, foram coletados dados comportamentais de amostras de *ransomware* e binários benignos em um ambiente de análise controlado. O ambiente de captura utilizou o *Cuckoo Sandbox*, uma ferramenta amplamente utilizada em pesquisas de cibersegurança para a análise dinâmica de *malwares* em ambientes virtuais isolados (OKTAVIANTO; MUHARDIANTO, 2013). Esta ferramenta permite a execução e monitoramento detalhado dos binários, registrando suas interações com o sistema operacional e seus comportamentos.

4.1.1 Configuração do Ambiente de Coleta

O ambiente de análise foi configurado para executar cada binário em uma máquina virtual isolada, onde as atividades de *malware* não comprometem o sistema de análise. Para assegurar a representatividade dos dados coletados, foram realizadas múltiplas execuções de cada amostra de *ransomware* e de cada binário benigno. Cada execução teve duração média de dez mi-

nutos, permitindo a captura de atividades maliciosas e normais ao longo do tempo. Essa prática de múltiplas execuções visa aumentar a quantidade e a diversidade dos registros comportamentais, ampliando a representatividade do conjunto de dados final (CANALI et al., 2012).

4.1.2 Coleta de Dados com o *Cuckoo Sandbox*

Durante a execução dos binários, o *Cuckoo Sandbox* registrou atividades do sistema, como chamadas de *API*, interações com o sistema de arquivos e comunicação em rede. Esses registros foram exportados em formato *JSON*, facilitando a estruturação e a organização dos dados coletados. A escolha das informações registradas foi realizada com base na relevância desses atributos para a análise comportamental, permitindo que os dados coletados fossem analisados posteriormente por algoritmos de aprendizado de máquina e técnicas de análise exploratória (IJAZ; DURAD; ISMAIL, 2019).

4.1.3 Pré-processamento dos Dados

O pré-processamento dos dados coletados foi essencial para garantir a consistência e qualidade das amostras. Primeiramente, os arquivos *JSON* foram convertidos para o formato *CSV*, o que permitiu uma análise mais eficiente. Em seguida, realizou-se a filtragem dos atributos para remover informações redundantes e irrelevantes. O Atributos considerados essenciais para a detecção e classificação de *ransomwares*, como chamadas de *API* relacionadas ao sistema e acesso ao registro, foram preservados, enquanto atributos com baixo valor informativo foram descartados.

Após a filtragem, foi realizada a rotulagem das amostras, identificando os binários como benignos ou maliciosos, de acordo com as características do comportamento registrado. Este processo de rotulagem envolveu a classificação das amostras de *ransomware* de acordo com sua família e especificidade, o que contribui para uma análise mais detalhada e específica de cada tipo de ameaça (CANALI et al., 2012).

4.2 Seleção de atributos e avaliação dos resultados

Conforme mencionado anteriormente, utiliza-se o algoritmo *IG*, fundamentado na redução de entropia, para avaliar a importância dos atributos em relação às classes de amostras examinadas. O cálculo do *IG* permite determinar a relevância de cada atributo para a classificação. Quanto maior o valor do *IG*, mais significativa é a diminuição da entropia, indicando que o atributo contribui de maneira relevante para a diferenciação entre as classes (QUINLAN, 1986).

A seleção de atributos no estudo sobre *ransomware*, utilizando o *IG*, revela *insights*. Ao analisar a Figura 8, observam-se atributos como *score_binary* e *NtClose*, com pesos respectivos de 1,3267 e 1,2389, que são cruciais, indicando atividades suspeitas. Da mesma forma,

operações de sistema como *NtOpenKey*, *NtQueryValueKey* e *NtAllocateVirtualMemory*, com pesos de 1,0563, 1,0068 e 0,9647 respectivamente, destacam-se na identificação de atividades mal-intencionadas relacionadas à manipulação do registro e gestão de memória. *LdrGetProcedureAddress*, com um peso de 0,7554, também se mostra vital na detecção de comportamentos típicos de *ransomware*, evidenciando a importância de operações de busca de endereços de procedimentos. Este conjunto de atributos, com altos valores de *IG*, contribui para aumentar a precisão do modelo na distinção entre comportamentos benignos e maliciosos, destacando a relevância da seleção estratégica de atributos no combate ao *ransomware* (CANALI et al., 2012).

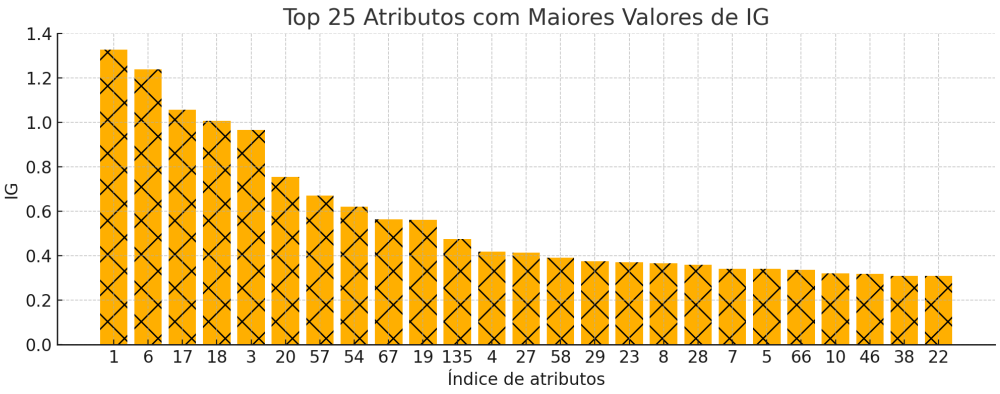

Figura 8 – Gráfico dos principais atributos classificados pelo algoritmo IG.

4.3 Cenários de utilização do RansomSet

Esta seção apresenta exemplos práticos e contextos nos quais o RansomSet pode ser aplicado para contribuir com a segurança cibernética. Esta seção explora situações diversas, desde o desenvolvimento de modelos de detecção para ferramentas antivírus até o uso em ambientes corporativos para avaliar políticas de segurança. O objetivo é demonstrar o potencial do *RansomSet* como um conjunto de dados versátil e útil para a análise de comportamento de *ransomware*, destacando sua aplicabilidade em áreas essenciais para a prevenção e mitigação de ataques.

4.3.1 Cenário 1: Desenvolvimento de Modelos de Detecção para Antivírus

O *RansomSet* oferece uma ampla gama de atributos comportamentais específicos de *ransomware*, o que possibilita o desenvolvimento e o aprimoramento de modelos de detecção para ferramentas antivírus e *antimalware*. Ao utilizar técnicas de aprendizado de máquina, como redes neurais, é possível treinar modelos que aprendam a identificar padrões de comportamento suspeitos, presentes apenas em amostras maliciosas. Este cenário atende a empresas de segurança cibernética que desejam implementar modelos robustos para identificar novos ataques de *ransomware* com base em dados reais e comportamentais, aprimorando, assim, a taxa de detecção e reduzindo a ocorrência de falsos positivos.

4.3.2 Cenário 2: Análise de Comportamento para Pesquisa Acadêmica

No contexto acadêmico e de pesquisa científica, o RansomSet pode ser utilizado para estudar e entender as características únicas e os padrões de comportamento das principais famílias de *ransomware*. Esse conjunto de dados permite que pesquisadores investiguem com precisão como diferentes *ransomwares* interagem com o sistema, acessam arquivos, executam chamadas de sistema e realizam operações de rede. Este cenário é adequado para laboratórios de pesquisa e universidades que desenvolvem projetos voltados à análise comportamental de *ransomwares*, onde o objetivo é aprofundar o entendimento sobre as táticas e técnicas empregadas por criminosos, permitindo a descoberta de novas abordagens para a detecção e mitigação de ameaças.

4.3.3 Cenário 3: Teste e Avaliação de Políticas de Segurança Corporativa

Em ambientes corporativos, o RansomSet é uma dados valiosos para testar e avaliar políticas de segurança de rede e práticas de defesa cibernética. Utilizando o conjunto de dados em simulações de ataque, as equipes de segurança da informação podem verificar a eficácia das suas medidas de proteção e identificar possíveis vulnerabilidades que poderiam ser exploradas por *ransomware*. Ao realizar essas simulações com base em dados reais e representativos, os profissionais conseguem ajustar as políticas de controle de acesso, segmentação de rede e monitoramento de atividades, de forma a criar uma camada de defesa mais eficaz. Esse cenário beneficia especialmente grandes empresas e organizações que precisam garantir a integridade de seus dados, protegendo-se de ataques cada vez mais sofisticados.

4.4 Método de inclusão de classes no RansomSet

Esta seção descreve o processo de expansão do RansomSet com novas classes de *ransomware* e amostras benignas, com o objetivo de tornar o conjunto de dados mais abrangente e atualizado, acompanhando a evolução das ameaças cibernéticas. A metodologia de inclusão de novas classes no RansomSet visa garantir a continuidade e relevância do conjunto de dados, de modo que ele possa atender a análises e demandas crescentes na área de segurança digital.

4.4.1 Novas classes

Para adicionar uma nova classe de *ransomware* ao RansomSet, recomenda-se seguir uma abordagem sistemática para garantir a integridade dos dados e a confiabilidade das análises. Inicialmente, é necessário obter amostras seguras e autenticadas da nova variante de *ransomware* a ser incluída. A coleta dessas amostras pode ser realizada a partir de repositórios confiáveis de *malware*, como *MalwareBazaar* ou *VirusShare*, garantindo a procedência das amostras e minimizando o risco de falsos positivos nas análises.

Após a obtenção das amostras, cada binário deve ser executado em um ambiente isolado, como o *Cuckoo Sandbox*, que registra as atividades comportamentais do *ransomware* de forma controlada. Durante a execução, é essencial capturar dados como chamadas de sistema. Esse processo deve ser repetido várias vezes, com diferentes configurações, para gerar uma quantidade suficiente de observações que representem adequadamente o comportamento da nova classe que será incluída no conjunto de dados.

Os dados coletados são convertidos em um formato estruturado, como *JSON* ou *CSV*, facilitando a integração com o RansomSet. Cada nova classe é então submetida ao algoritmo de ranqueamento de *IG*, que prioriza os atributos mais relevantes para a análise. O algoritmo aplicado se baseia no código da pesquisa desenvolvida por (QUINCOZES, 2021), conforme referência. Após essa etapa, o algoritmo descrito na Seção 3.3.3 é utilizado para consolidar o novo conjunto de dados, agregando-o de forma coerente ao RansomSet.

5 Conclusão

Este estudo apresentou o *RansomSet*, um conjunto de dados abrangente que coletou dados comportamentais de 6 variantes diferentes de *ransomware*, incluindo *WannaCry*, *Ryuk*, *CryptoLocker*, *LockBit*, *Sodinokibi* e *Conti*, juntamente com exemplos de comportamento usual. Utilizou-se o *Cuckoo Sandbox* para analisar o comportamento desses *ransomwares* e criar arquivos *JSON* como base para filtrar, coletar e analisar dados. Demonstrou-se o potencial dos *RansomSets* ao explorar e compreender as propriedades dos dados que os compõem, utilizando técnicas de análise exploratória de dados. A estrutura do conjunto de dados permitiu incluir múltiplas classes de *ransomware* e binários benignos em estudo. Cada classe representou um tipo específico de comportamento, permitindo analisar e comparar o comportamento do *ransomware* com o comportamento benigno do sistema. O principal objetivo foi entender o comportamento do *ransomware* em análise e identificar padrões e desvios em relação ao comportamento benigno do sistema.

Em trabalhos futuros, o objetivo será aprimorar as capacidades de classificação funcional dos *RansomSets*, visando aumentar a precisão da análise de dados e a detecção eficaz de ameaças. Como trabalhos futuros, pretende-se empregar técnicas de Inteligência Artificial Explicável (XAI) a fim de estudar de maneira mais profunda o impacto dos atributos do RansomSet na detecção de diferentes classes de ataques. Ademais, pretende-se ampliar a base de amostras contidas no conjunto de dados ao abordar maior diversidade de amostras benignas para obter-se maior generalização dos modelos de aprendizado de máquina empregados no processamento do RansomSet. Esses princípios melhoram a capacidade de detectar e responder a ameaças de *ransomware*, contribuindo para a segurança cibernética geral.

5.1 Produções bibliográficas

Como parte dos resultados deste trabalho, foi publicado o artigo intitulado *"Um Estudo Acerca da Seleção de Atributos para a Detecção dos Ransomwares WannaCry, Ryuk e CryptoLocker"*[1], em que são apresentados os métodos de seleção de atributos aplicados na detecção de *ransomwares* específicos. Esta publicação está alinhada com os objetivos da pesquisa e representa uma contribuição significativa para a área.

[1] <https://doi.org/10.5753/semish.2023.229616>

Referências

ABBASI, M. S.; AL-SAHAF, H.; MANSOORI, M.; WELCH, I. Behavior-based ransomware classification: A particle swarm optimization wrapper-based approach for feature selection. **Applied Soft Computing**, Elsevier, v. 121, p. 108744, 2022. Citado na página 10.

ADVISOR, C. **Accenture confirma invasão e vazamento de dados**. 2021. Acesso em: 12 nov. 2024. Disponível em: <https://www.cisoadvisor.com.br/accenture-confirma-invasao-e-vazamento-de-dados/>. Citado na página 14.

AKBANOV, M.; VASSILAKIS, V. G.; LOGOTHETIS, M. D. Wannacry ransomware: Analysis of infection, persistence, recovery prevention and propagation mechanisms. **Journal of Telecommunications and Information Technology**, Instytut Łączności-Państwowy Instytut Badawczy, n. 1, p. 113–124, 2019. Citado na página 13.

ALZAHRANI, S.; XIAO, Y.; SUN, W. An analysis of conti ransomware leaked source codes. **IEEE Access**, IEEE, v. 10, p. 100178–100193, 2022. Citado na página 32.

AZHAGUSUNDARI, B.; THANAMANI, A. S. et al. Feature selection based on information gain. **International Journal of Innovative Technology and Exploring Engineering (IJITEE)**, Citeseer, v. 2, n. 2, p. 18–21, 2013. Citado na página 30.

BEAMAN, C.; BARKWORTH, A.; AKANDE, T. D.; HAKAK, S.; KHAN, M. K. Ransomware: Recent advances, analysis, challenges and future research directions. **Computers & Security**, v. 111, p. 102490, 2021. ISSN 0167-4048. Citado na página 10.

BERRUETA, E.; MORATO, D.; MAGAÑA, E.; IZAL, M. Open repository for the evaluation of ransomware detection tools. **IEEE Access**, IEEE, v. 8, p. 65658–65669, 2020. Citado 2 vezes nas páginas 18 e 20.

BREWER, R. Ransomware attacks: detection, prevention and cure. **Network security**, Elsevier, v. 2016, n. 9, p. 5–9, 2016. Citado na página 10.

CANALI, D.; LANZI, A.; BALZAROTTI, D.; KRUEGEL, C.; CHRISTODORESCU, M.; KIRDA, E. A quantitative study of accuracy in system call-based malware detection. In: **Proceedings of the 2012 International Symposium on Software Testing and Analysis**. [S.l.: s.n.], 2012. p. 122–132. Citado 2 vezes nas páginas 38 e 39.

CARRIER, T.; VICTOR, P.; TEKEOGLU, A.; LASHKARI, A. H. Detecting obfuscated malware using memory feature engineering. In: **Icissp**. [S.l.: s.n.], 2022. p. 177–188. Citado 2 vezes nas páginas 16 e 20.

CHEN, Z.-G.; KANG, H.-S.; YIN, S.-N.; KIM, S.-R. Automatic ransomware detection and analysis based on dynamic api calls flow graph. In: **Proceedings of the international conference on research in adaptive and convergent systems**. [S.l.: s.n.], 2017. p. 196–201. Citado na página 28.

CONTINELLA, A.; GUAGNELLI, A.; ZINGARO, G.; PASQUALE, G. D.; BARENGHI, A.; ZANERO, S.; MAGGI, F. Shieldfs: a self-healing, ransomware-aware filesystem. In: **Proceedings of the 32nd annual conference on computer security applications**. [S.l.: s.n.], 2016. p. 336–347. Citado 2 vezes nas páginas 18 e 20.

CSA. **Conti Ransomware**. 2022. <https://media.defense.gov/2021/Sep/22/2002859507/-1/-1/0/CSA_Conti_Ransomware_20220309.PDF>. Accessed: 2024-04-25. Citado na página 35.

CURRAN, K. Cyber security and the remote workforce. **Computer Fraud & Security**, MA Business London, v. 2020, n. 6, p. 11–12, 2020. Citado na página 10.

DAMODARAN, A.; TROIA, F. D.; VISAGGIO, C. A.; AUSTIN, T. H.; STAMP, M. A comparison of static, dynamic, and hybrid analysis for malware detection. **Journal of Computer Virology and Hacking Techniques**, Springer, v. 13, n. 1, p. 1–12, 2017. Citado 3 vezes nas páginas 10, 14 e 15.

GAGULIC, D.; ZUMTAUGWALD, L.; SAHU, S.; PLETKA, R. Ransomware detection with machine learning in storage systems. 2023. Citado na página 14.

HERRERA-SILVA, J. A.; HERNáNDEZ-ÁLVAREZ, M. Dynamic feature dataset for ransomware detection using machine learning algorithms. **Sensors**, v. 23, n. 3, 2023. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/23/3/1053>. Citado 5 vezes nas páginas 15, 20, 21, 27 e 28.

HIRANO, M.; HODOTA, R.; KOBAYASHI, R. Ransap: An open dataset of ransomware storage access patterns for training machine learning models. **Forensic Science International: Digital Investigation**, v. 40, p. 301314, 2022. ISSN 2666-2817. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666281721002390>. Citado 2 vezes nas páginas 17 e 20.

HUMAYUN, M.; JHANJHI, N.; ALSAYAT, A.; PONNUSAMY, V. Internet of things and ransomware: Evolution, mitigation and prevention. **Egyptian Informatics Journal**, Elsevier, v. 22, n. 1, p. 105–117, 2021. Citado na página 12.

HYUNA, E. L. T. L. Tech Report, **KARA ransomware trends report**. 2024. Acessado em: 3 out. 2024. Disponível em: <https://www.skshieldus.com/download/files/download.do?o_fname=KARA%20ëđœìĎňìŻĺìŰť%20ëŘŹíŰěëşťêşăìĎœ%202024.05(ENG).pdf&r_fname=20240529154106821.pdf>. Citado 2 vezes nas páginas 10 e 14.

IJAZ, M.; DURAD, M. H.; ISMAIL, M. Static and dynamic malware analysis using machine learning. In: IEEE. **2019 16th International bhurban conference on applied sciences and technology (IBCAST)**. [S.l.], 2019. p. 687–691. Citado na página 38.

ILIĆ, S. Ž.; GNJATOVIĆ, M. J.; POPOVIĆ, B. M.; MAČEK, N. D. A pilot comparative analysis of the cuckoo and drakvuf sandboxes: An end-user perspective. **Vojnotehnički glasnik**, v. 70, n. 2, p. 372–392, 2022. Citado na página 28.

KHARAZ, A.; ARSHAD, S.; MULLINER, C.; ROBERTSON, W.; KIRDA, E. {UNVEIL}: A {Large-Scale}, automated approach to detecting ransomware. In: **25th USENIX security symposium (USENIX Security 16)**. [S.l.: s.n.], 2016. p. 757–772. Citado 3 vezes nas páginas 18, 20 e 21.

MICRO, T. **2020 Report on Threats Affecting ICS Endpoints**. 2020. 1-22 p. Disponível em: https://documents.trendmicro.com/assets/white_papers/wp-2020-report-on-threats-affecting-critical-industrial-endpoints.pdf. Acesso em: Junho/2023. Citado 2 vezes nas páginas 10 e 12.

____. **Defending the Expanding Attack Surface**. 2022. 1-50 p. Disponível em: https://documents.trendmicro.com/assets/rpt/rpt-defending-the-expanding-attack-surface-trend-micro-2022-midyear-cybersecurity-report.pdf. Acesso em: Junho/2023. Citado 2 vezes nas páginas 10 e 12.

____. **LockBit, BlackCat, and Royal Dominate the Ransomware Scene in Q4 2022**. 2022. Disponível em: <https://documents.trendmicro.com/assets/rpt/rpt-defending-the-expanding-attack-surface-trend-micro-2022-midyear-cybersecurity-report.pdf>. Citado na página 12.

____. **Ransomware in Q1 2024**. 2024. Disponível em: <https://www.trendmicro.com/vinfo/us/security/news/ransomware-by-the-numbers/phobos-emerges-as-a-formidable-threat-in-q1-2024-lockbit-stays-in-the-top-spot>. Citado 2 vezes nas páginas 12 e 14.

OKTAVIANTO, D.; MUHARDIANTO, I. **Cuckoo malware analysis**. [S.l.]: Packt Publishing Ltd, 2013. Citado na página 37.

QUINCOZES, S. E. **Ereno IDS**. 2021. Disponível em: <https://github.com/sequincozes/ereno-ids>. Citado na página 41.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, p. 81–106, 1986. Citado 2 vezes nas páginas 30 e 38.

RAZAULLA, S.; FACHKHA, C.; MARKARIAN, C.; GAWANMEH, A.; MANSOOR, W.; FUNG, B. C.; ASSI, C. The age of ransomware: A survey on the evolution, taxonomy, and research directions. **IEEE Access**, IEEE, 2023. Citado na página 10.

RICHARDSON, R.; NORTH, M. M. Ransomware: Evolution, mitigation and prevention. **International Management Review**, v. 13, n. 1, p. 10, 2017. Citado na página 12.

RIECK, K.; TRINIUS, P.; WILLEMS, C.; HOLZ, T. Automatic analysis of malware behavior using machine learning. **Journal of computer security**, IOS Press, v. 19, n. 4, p. 639–668, 2011. Citado 2 vezes nas páginas 16 e 20.

SCAIFE, N.; CARTER, H.; TRAYNOR, P.; BUTLER, K. R. Cryptolock (and drop it): stopping ransomware attacks on user data. In: IEEE. **2016 IEEE 36th international conference on distributed computing systems (ICDCS)**. [S.l.], 2016. p. 303–312. Citado 3 vezes nas páginas 17, 20 e 21.

SGANDURRA, D.; MUÑOZ-GONZÁLEZ, L.; MOHSEN, R.; LUPU, E. C. Automated dynamic analysis of ransomware: Benefits, limitations and use for detection. **arXiv preprint arXiv:1609.03020**, 2016. Citado 3 vezes nas páginas 16, 20 e 21.

SHARMA, N.; SHANKER, R. Analysis of ransomware attack and their countermeasures: A review. In: IEEE. **2022 International Conference on Electronics and Renewable Systems (ICEARS)**. [S.l.], 2022. p. 1877–1883. Citado na página 13.

SIHWAIL, R.; OMAR, K.; ARIFFIN, K. Z. A survey on malware analysis techniques: Static, dynamic, hybrid and memory analysis. **Int. J. Adv. Sci. Eng. Inf. Technol**, v. 8, n. 4-2, p. 1662–1671, 2018. Citado na página 14.

SOUZA, G. O.; QUINCOZES, S. E.; KAZIENKO, J. F.; QUINCOZES, V. E.; FARIA, N. N. R. Um estudo acerca da seleçao de features para a detecçao dos ransomwares wannacry, ryuk e cryptolocker. In: SBC. **L Seminário Integrado de Software e Hardware (SEMISH)**. [S.l.], 2023. p. 36–47. Citado 4 vezes nas páginas 5, 10, 24 e 25.

STEIGER, J. H. Tests for comparing elements of a correlation matrix. **Psychological bulletin**, American Psychological Association, v. 87, n. 2, p. 245, 1980. Citado na página 33.

TANG, F.; MA, B.; LI, J.; ZHANG, F.; SU, J.; MA, J. Ransomspector: An introspection-based approach to detect crypto ransomware. **Comput. Secur.**, v. 97, p. 101997, 2020. Disponível em: <https://api.semanticscholar.org/CorpusID:221666937>. Citado 2 vezes nas páginas 17 e 20.

UMAR, R.; RIADI, I.; KUSUMA, R. S. Mitigating sodinokibi ransomware attack on cloud network using software-defined networking (sdn). **International Journal of Safety and Security Engineering**, v. 11, n. 3, p. 239–246, 2021. Citado na página 13.

UPPAL, D.; MEHRA, V.; VERMA, V. Basic survey on malware analysis, tools and techniques. **Int. Jrnl on Computational Sciences & Applications (IJCSA)**, v. 4, n. 1, p. 103, 2014. Citado 2 vezes nas páginas 14 e 15.

VMWARE. **TAU Threat Discovery: Conti Ransomware**. 2020. <https://blogs.vmware.com/security/2020/07/tau-threat-discovery-conti-ransomware.html>. Accessed: 2024-02-19. Citado na página 32.