DISSERTAÇÃO DE MESTRADO PROFISSIONAL

METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO

ALESSANDRA DE MELO E SILVA

Brasília, 18 de dezembro de 2020

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

INTEGRATIVE METHODOLOGY TO PRODUCE CYBER THREAT INTELLIGENCE USING OPEN SOURCE PLATFORMS

METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO

ALESSANDRA DE MELO E SILVA

ORIENTADOR: DR. JOÃO JOSÉ COSTA GONDIM
COORIENTADOR: DR. ROBSON DE OLIVEIRA ALBUQUERQUE

DISSERTAÇÃO DE MESTRADO PROFISSIONAL
EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPEE.MP.008

BRASÍLIA/DF: DEZEMBRO - 2020

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO

ALESSANDRA DE MELO E SILVA

Dissertação de Mestrado Profissional submetida ao Departamento de Engenharia Elétrica como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica

Banca Examinadora

Prof. João José Costa Gondim, Dr., CIC/UnB _____
Orientador

Profa. Edna Dias Canedo, Dra., FT/UnB _____
Examinador Interno

Prof. André Ricardo Abed Grégio, Dr., DI/UFPR _____
Examinador Externo

Prof. Georges Daniel Amvame Nze, Dr., FT/UnB _____
Suplente

FICHA CATALOGRÁFICA

SILVA, ALESSANDRA DE MELO
METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO [Distrito Federal] 2020.
xvi, 60 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2020).
Dissertação de Mestrado Profissional - Universidade de Brasília, Faculdade de Tecnologia.
Departamento de Engenharia Elétrica

1. Segurança Cibernética	2. Inteligência de Ameaças
3. Plataformas de CTI	4. Plataformas *open source*
I. ENE/FT/UnB	II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

SILVA, A.M. (2020). *METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO*. Dissertação de Mestrado Profissional, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 60 p.

CESSÃO DE DIREITOS

AUTOR: Alessandra de Melo e Silva
TÍTULO: METODOLOGIA INTEGRATIVA PARA PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS UTILIZANDO PLATAFORMAS DE CÓDIGO ABERTO.
GRAU: Mestre em Engenharia Elétrica ANO: 2020

Alessandra de Melo e Silva
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

AGRADECIMENTOS

Primeiramente, agradeço em conjunto ao meu orientador, Prof. João José Costa Gondim, e Co-orientador, Prof. Robson de Oliveira Albuquerque, por todos os ensinamentos, conselhos e oportunidades oferecidos para realização desse trabalho, além de toda paciência, dedicação e prontidão para me ajudar durante esse processo.

Agradeço aos meus pais, Constantino e Neide, e irmão, Caio, pelo suporte incondicional durante toda a minha vida e principalmente por toda a força para que eu não desistisse do meu mestrado.

Agradeço a minha melhor amiga, Camila Fernandes, por toda a companhia e incentivos diários para a conclusão deste trabalho.

Por fim, agradeço a toda minha família e amigos pelo companheirismo e por se fazerem presentes durante essa jornada.

RESUMO

Título: Metodologia Integrativa para Produção de Inteligência de Ameaças Cibernéticas Utilizando Plataformas de Código Aberto
Autor: Alessandra de Melo e Silva
Orientador: Dr. João José Costa Gondim
Programa de Pós-Graduação Profissional em Engenharia Elétrica – Área de Concentração em Segurança Cibernética
Brasília, 18 de dezembro de 2020.

Neste trabalho foi proposta uma metodologia integrativa de plataformas de CTI de código aberto visando a produção de inteligência de ameaça de qualidade. Primeiramente, foi desenvolvida uma metodologia de avaliação para analisar padrões e plataformas de CTI e definir soluções com potencial de consolidação no mercado. Essa metodologia de avaliação baseou-se em uma estratégia de seleção de soluções de CTI populares de código aberto e no estabelecimento de critérios de avaliação para analisar e comparar essas soluções. Os resultados dessa avaliação comparativa mostraram a existência de boas soluções de CTI de código aberto e possibilitaram a definição de uma metodologia integrativa para a produção de inteligência de ameaça, baseada na complementaridade das plataformas MISP e OpenCTI. Alguns cenários de teste foram simulados e analisados com base em uma proposta definida que utiliza o método 5W3H para avaliar a completude da inteligência produzida e, consequentemente, entender sua qualidade e eficácia no processo de tomada de decisão contra incidentes. A partir dos resultados produzidos foi possível verificar que a metodologia proposta é satisfatória quando aplicada à conjuntos de dados de ameaça de tamanho controlado e contextualizados. Além disso, foi possível identificar algumas desvantagens em sua aplicação que podem proporcionar o desenvolvimento de trabalhos futuros.

Palavras-chave: Segurança Cibernética, Inteligência de Ameaças, Plataformas de CTI, Plataformas *open source*.

ABSTRACT

Title: Integrative Methodology to Produce Cyber Threat Intelligence Using Open Source Platforms
Author: Alessandra de Melo e Silva
Supervisor: Dr. João José Costa Gondim
Professional Post-Graduate Program in Electrical Engineering – Cybersecurity Concentration Area
Brasília, December 18th, 2020.

In this work, an integrative methodology of open source CTI platforms was proposed, aiming to produce quality threat intelligence. First, an evaluation methodology was developed to analyze CTI standards and platforms and define solutions with potential for consolidation in the market. This evaluation methodology was based on a strategy for selecting popular open source CTI solutions and establishing evaluation criteria to analyze and compare these solutions. The results of this comparative evaluation showed the existence of great open source CTI solutions and have led to the definition of an integrative methodology for the production of threat intelligence, based on the complementarity of the platforms MISP and OpenCTI. Some test scenarios were simulated and analyzed based on a defined proposal that uses the 5W3H method to assess the completeness of the intelligence produced and, consequently, understand its quality and effectiveness in the decision making process. From the results produced, it was possible to verify that the proposed methodology is satisfactory when applied to contextualized and controlled size threat data sets. In addition, it was possible to identify some disadvantages in its application that may provide the development of future works.

Keywords: Cyber Security, Threat Intelligence, CTI Platforms, Open Source Platforms.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ACRÔNIMOS

APT	*Advanced Persistent Threats*
ARF	*Abuse Reporting Format*
CAIF	*Common Announcement Interchange Format*
CAPEC	*Common Attack Pattern Enumeration and Classification*
CEE	*Common Event Expression*
CIDF	*Common Intrusion Detection Framework*
CIF	*Collective Intelligence Framework*
CRITs	*Collaborative Research into Threats*
CTI	*Cyber Threat Intelligence* (Inteligência de ameaças cibernéticas)
CybOX	*Cyber Observable eXpression*
GNU	*General Public License*
HTTPS	*Hyper Text Transfer Protocol Secure*
IDMEF	*Intrusion Detection Message Exchange Format*
IDS	*Intrusion Detection System*
IETF	*Internet Engineering Task Force*
IOC	*Indicator Of Compromise* (Indicador de comprometimento)
IODEF	*Incident Object Description Exchange Format*
IP	*Internet Protocol Version 4*
MAEC	*Malware Attribute Enumeration and Characterization*
MISP	*Malware Information Sharing Platform*
OpenIOC	*Open Indicator Of Compromise*
OpenTPX	*Open Threat Partner Exchange*
RID	*Real-time Inter-network Defense*
SIEM	*Security Information and Event Management*
STIX	*Structured Threat Information eXpression*
TAXII	*Trusted Automated Exchange of Intelligence Information*
TI	*Threat Intelligence* (Inteligência de ameaças)
TTP	*Tactics, Techniques, and Procedures*
URL	*Uniform Resource Locator*

1 INTRODUÇÃO

Nos últimos anos, com o relevante aumento do poder computacional e das tecnologias de comunicação, um novo mercado de dispositivos de rede variados e diferentes sistemas tecnológicos emergiu rapidamente e esses dispositivos estão trazendo uma ampla gama de vulnerabilidades exploráveis [10]. Consequentemente, o número de ataques cibernéticos e seus custos também aumentaram [11]. Além disso, esses novos ataques cibernéticos são mais complexos e direcionados [12], gerando cenários de ataque mais sofisticados e aprimorados. Esses fatos indicam que o espectro da segurança cibernética está mudando fundamentalmente e se tornando cada vez mais desafiador.

O progressivo crescimento dos ataques cibernéticos atuais surge de uma cascata de novos sistemas sofisticados que estão sendo desenvolvidos por invasores e especialistas em segurança, e quanto mais complexo um sistema fica, mais inseguro ele se torna [13]. Outro motivo para a evolução dos ataques é o fato de esses estarem sendo melhor planejados e aplicados de forma mais específica [14], o que os torna mais complexos. A maioria deles são desenvolvidos para não serem detectados pelas defesas de primeiro nível, podendo persistir no sistema [15]. Além disso, essas novas ameaças estão em constante processo de modificação e melhoria, tornando sua detecção e defesa mais complicada [14]. Os avanços e modificações no ecossistema de ataque cibernético têm estimulado mudanças no modelo de defesa tradicional e a busca por métodos mais eficientes e proativos [10] [15].

Considerando o cenário apresentado, a ideia de *Cyber Threat Intelligence* (CTI - Inteligência de Ameaça Cibernética) tem se popularizado rapidamente e muitas vezes é apresentada como uma nova solução para aplicar segurança efetiva em diferentes organizações [2]. Qualquer tipo de informação que possa ser usada para identificar, caracterizar ou auxiliar na resposta a ameaças cibernéticas é comumente referida como informação de ameaça cibernética e a análise deste tipo de informação pode produzir inteligência para informar o usuário sobre ameaças ao seu sistema [16].

Nesse contexto, novos sistemas automatizados com a capacidade de consumir uma grande quantidade de dados, fornecer recursos de defesa sofisticados e responder a incidentes em tempo real estão sendo desenvolvidos e comumente referidos como plataformas de *Threat Intelligence* (TI - Inteligência de Ameaça) [17] [18]. Essas plataformas devem incluir processos automáticos de transformação de dados e produção de inteligência para garantir um modelo de defesa mais eficiente, proativo e oportuno [19]. Além disso, devido à heterogeneidade dos dados inseridos no contexto de CTI, esforços consideráveis têm sido feitos no sentido de padronizar os dados [17] e torná-los compatíveis entre os diferentes sistemas [20]. A interoperabilidade dos dados é importante para facilitar a coleta e análise automática dos dados e o compartilhamento de inteligência contra ameaças cibernéticas [21].

No entanto, a diversidade de iniciativas em relação ao tema ocasionou o crescimento de um mercado heterogêneo de plataformas CTI e diversos sistemas e ferramentas, com diferentes objetivos, capacidades e níveis de desempenho foram disponibilizados de mercado de *Threat Intelligence Platforms* (TIPs - Plataformas de Inteligência de Ameaça) [18]. Embora algumas dessas sejam ótimas soluções, ainda é um desafio encontrar uma solução completa para uma defesa baseada em inteligência de ameaças, uma vez

que as plataformas têm focos divergentes e conseqüentemente correspondem ou focam em apenas algumas etapas do processo de produção de inteligência de ameaças [22].

Além disso, ainda existem obstáculos a serem ultrapassados para uma consolidação da abordagem de CTI. Dentro dessas limitações, está a massiva quantidade de dados disponíveis para coleta [23]. Muitas vezes, a quantidade de dados é valorizada em relação à qualidade, implicando em uma menor eficiência para tomada de decisões [13]. Para que a informação seja capaz de auxiliar no processo de resposta a ameaças e incidentes e, consequentemente, seja considerada informação eficiente e de qualidade, é necessário que ela atenda requisitos como ser relevante, precisa, oportuna, coerente e clara [24], algo que não ocorre na maioria das soluções existentes atualmente. Além disso, outra limitação existente é a heterogeneidade dos dados envolvidos no processo [21] [25]. Por serem derivados de diferentes fontes, os dados são disponibilizados em diferentes formatos e não constituem um conjunto de dados estruturados para análise. Assim, para uma utilização eficaz das informações sobre ameaças cibernéticas, são necessários mecanismos capazes de consumir, analisar, avaliar e classificar as informações [20], para assim produzir inteligência de ameaças cibernéticas de qualidade.

Dessa forma, este trabalho desenvolverá uma metodologia que utilize a interoperabilidade entre diferentes plataformas de TI visando fornecer uma solução mais completa para produzir e gerenciar dados de inteligência de ameaças. Para isso, serão estudadas plataformas com características complementares, de modo que seus pontos fortes sejam ressaltados e otimizados. Além disso, o trabalho propõe um método de análise da qualidade da inteligência produzida, como meio de comprovar a eficiência da metodologia adotada.

1.1 MOTIVAÇÃO

Mesmo com sua rápida popularização nos últimos anos e consequente surgimento de estudos e iniciativas para a utilização de inteligência de ameaça cibernética como mecanismo de defesa cibernética, pesquisas comprovam diversas problemáticas que ainda estão sendo trabalhadas para o desenvolvimento de soluções eficientes. Dentre essas problemáticas, pode-se dizer que o quesito de qualidade dos dados disponibilizados por plataformas de TI é um dos mais abordados.

Além disso, apesar de já existirem boas soluções para aplicação de TI, a maioria delas não constitui uma abordagem completa quando considerado o fluxo de processos de CTI em sua integridade. Além disso, algumas delas ainda possuem pontos falhos em relação a performance e cenários de aplicação, impactando na qualidade final das informações produzidas. Por tais motivos, há indicações que existem obstáculos a serem superados em relação ao tema.

Por ser um tema promissor no contexto de segurança cibernética, a elaboração de uma pesquisa detalhada sobre o assunto e o desenvolvimento de um método para a otimização da qualidade dos dados de inteligência de ameaça produzidos a partir de plataformas já existentes no mercado, podem trazer grandes vantagens e abrir novos caminhos para a elaboração de soluções eficientes de defesa contra ameaças cibernéticas.

1.2 OBJETIVOS

O presente trabalho visa produzir dados de inteligência de ameaça capazes de disponibilizar inteligência de ameaça cibernética de qualidade, que atenda aos requisitos de ser relevante, oportuna e clara, para auxiliar Na consolidação de um mecanismo eficiente de defesa. Nesse sentido, esta pesquisa almeja:

1. Realizar um estudo do estado da arte de soluções de CTI com foco em identificar e selecionar as principais ferramentas disponíveis atualmente.
2. Analisar as ferramentas de CTI selecionadas, comparando e classificando-as com base em uma metodologia de avaliação proposta.
3. Implementar uma solução de inteligência de ameaça baseada na complementariedade de plataformas já existentes, seguindo os resultados da avaliação anteriormente realizada e visando destacar as principais características de cada uma delas.
4. Produzir inteligência de ameaça cibernética de qualidade utilizando a solução implementada.
5. Analisar e avaliar os dados produzidos de modo a garantir que sejam relevantes, oportunos e claros, sendo assim capazes de auxiliar efetivamente no processo de tomada de decisão contra ameaças cibernéticas.

1.3 CONTRIBUIÇÕES DO TRABALHO

O presente trabalho busca trazer as seguintes contribuições:

1. Apresentação do estado da arte de padrões e plataformas de TI, principalmente daquelas de código aberto;
2. Apresentação de uma estratégia de seleção de potenciais padrões e plataformas de TI e de uma metodologia de avaliação dessas soluções.
3. Avaliação dos principais padrões e plataformas de TI de código aberto existentes e disponibilização de um descritivo de suas funcionalidades.
4. Apresentação de uma arquitetura de integração de plataformas de TI com capacidade de produção de inteligência de ameaça cibernética.
5. Implementação funcional da arquitetura proposta com disponibilização de análises acerca dos resultados obtidos e da qualidade da informação produzida.

Vale ressaltar que o desenvolvimento deste trabalho e o alcance das contribuições descritas nos itens 1 a 3, levaram a publicação de um artigo em periódico indexado:

- SILVA, A. de Melo e; GONDIM, J. J. C.; ALBUQUERQUE, R. de O.; VILLALBA, L. J. G. (2020, Junho). A methodology to evaluate standards and platforms within cyber threat intelligence. Future Internet 2020, 12 (6), 108.

1.4 ESTRUTURA DO TRABALHO

Este trabalho foi ordenado em cinco capítulos, sendo este primeiro o de Introdução. Para facilitar o entendimento desta pesquisa os demais capítulos estão organizados como descrito a seguir:

O Capítulo 2 traz uma revisão do referencial teórico necessário para a compreensão da metodologia proposta no trabalho e a descrição das tecnologias aplicadas. Além disso, são apresentados trabalhos correlatos e relevantes ao tema.

O Capítulo 3 oferece uma revisão geral do mercado atual de ferramentas de CTI. Além disso, propõe uma metodologia para selecionar as princpais soluções existentes e avalia-as de maneira comparativa.

O Capítulo 4 apresenta a metodologia proposta para a implementação da solução proposta, otimização da qualidade dos dados de inteligência de ameaça produzidos e análise da qualidade desses dados.

O Capítulo 5 apresenta resultados experimentais obtidos a partir da aplicação prática, com conjuntos de dados previamente plaejados, do método proposto.

O Capítulo 6 conclui este trabalho. Nele, sintetizam-se os resultados encontrados e são sinalizados caminhos futuros que podem ser seguidos para a sequência deste trabalho.

2 REFERENCIAL TEÓRICO

Este capítulo contém a revisão teórica dos principais conceitos acerca do tema inteligência de ameaça cibernética (CTI), além de uma descrição sobre o estado da arte das soluções relacionadas ao tema. Para facilitar o entendimento, o presente capítulo está organizado em dois tópicos maiores. Na Seção 2.1, são tratados os conceitos referentes ao surgimento de novas ameaças cibernéticas e a inteligência de ameaça. Na Seção 2.2, é realizada uma revisão sobre o estado da arte e são apresentados trabalhos relacionados relevantes ao tema em questão.

2.1 CONCEITOS RELACIONADOS

Nesta seção serão apresentados os conceitos básicos para o embasamento e entendimento sobre o tema inteligência de ameaça cibernética. Estes conceitos vão desde os desafios enfrentados pelos modelos tradicionais de segurança na defesa de novas ameaças até os conceitos de inteligência de ameaça como nova abordagem de solução para essa problemática.

2.1.1 Novo cenário de ameaças

A grande evolução da computação nos últimos anos decorre, em grande parte, do surgimento de uma diversidade de dispositivos com capacidade de interação com a internet [26]. No entanto, a heterogeneidade e a interconexão desses dispositivos levaram a um aumento significativo no número de ataques de segurança [27] e o cenário de ameaças cibernéticas está se expandindo em proporções alarmantes. O aumento no número de ataques cresce em paralelo com a complexidade dos cenários de ataque e com o surgimento de ameaças cada vez mais sofisticadas.

Atualmente, o comportamento do adversário está mais focado em atingir alvos específicos e utiliza ataques que visam persistir no hospedeiro de modo a causar danos continuos [28]. Além disso, a maioria desses ataques ocorrem de forma transparente ao usuário, dificultando a sua detecção.

Algumas dessas novas ameaças são denominadas *Advanced Persistent Threats* (APTs). Elas executam um modelo de ataque sofisticado, caracterizado por estabelecer uma hospedagem persistente no alvo e permanecer indetectável por um longo período de tempo [14]. Outra característica dessa ameaças é o polimorfismo, ou seja, a capacidade de constante modificação, o que torna sua detecção uma tarefa bastante complexa. Além disso, outro tipo de ameaça amplamente explorada são as vulnerabilidades *zero-day*, que exploram vulnerabilidades não publicadas de um *software* e, por isso, acabam não sendo detectadas por longos períodos de tempo [13].

2.1.2 Modelo Tradicional de Segurança Cibernética

O método de defesa utilizado tradicionalmente pela maioria das organizações é baseado na coleta de informações sobre indicadores de segurança e ataques já disseminados, monitoramento de ativos com base nas informações reunidas e resposta a incidentes [29]. O fluxo de processos para uma arquitetura de segurança clássica, ainda amplamente aplicada, pode ser vista na Figura 2.1.

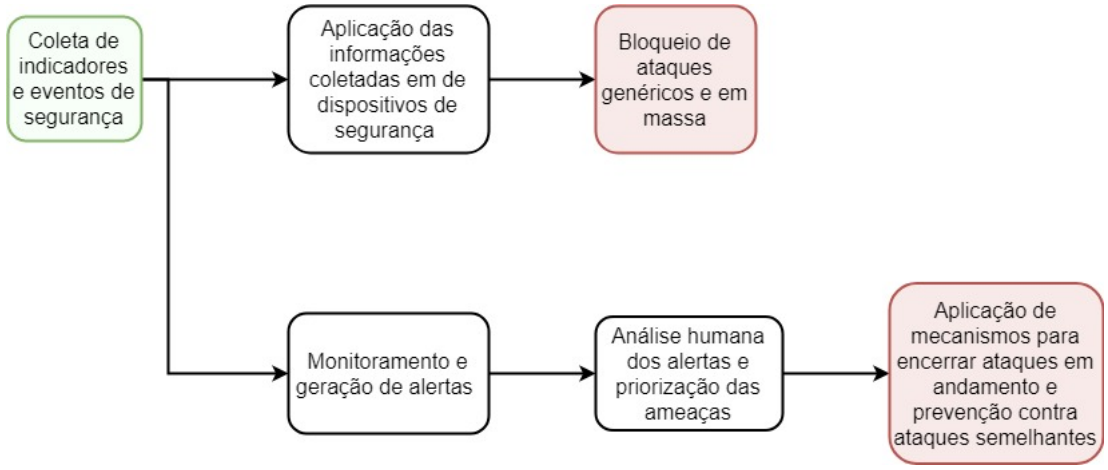


Figura 2.1: Fluxo dos processos de arquiteturas clássicas de segurança

Em linhas gerais, o fluxo apresentado descreve processos de ação reativos. Primeiramente, após a ocorrência de um incidente ou divulgação de uma vulnerabilidade, os dados sobre o incidente ou vulnerabilidade são coletados e estudados por especialistas de segurança. Em seguida, indicadores claros, como IPs e URLs maliciosos, são utilizados para aplicar filtros e regras em dispositivos de segurança. Essa ação proporciona apenas a possibilidade de defesa contra ataques generalizados, pois não envolve a variável contexto. Outros tipos de indicadores são incluídos em ferramentas de monitoramento que gerem alertas para possíveis ameaças. Esses alertas devem ser analisados, classificados e priorizados por equipes especializadas em segurança da informação. Essa vertente de ação permite a defesa contra ataques em andamento e gera informações que auxiliam na prevenção de ataques semelhantes, porém, devido ao grande número de alertas que devem ser analisados, esse processo se torna demorado e ineficiente.

No cenário atual de ataques cibernéticos, esse modelo de arquitetura focado em definir indicadores de comprometimento, do inglês *Indicators of Compromisse* (IOCs), adotar monitoramento humano e resposta direta a incidentes não é considerado muito produtivo. Primeiramente, em um nível tático, como todo tipo de ameaça é monitorada, o número de alertas reportados as equipes de segurança são imensos e muitas vezes geram falsos positivos, o que torna difícil a filtragem das ameaças mais relevantes. Em relação ao contexto operacional, o tempo demandado para receber alertas, filtrar as ameaças monitoradas, analisá-las de forma aprofundada e construir mecanismos de defesa é demasiadamente grande, o que oferece maior probabilidade de um ataque bem sucedido [29]. Além disso, considerando a vertente estratégica, na maioria dos casos as equipe encarregadas pela priorização das ameaças encontradas não possuem a quantidade de informações necessária para realizar essa ação e tomar decisões precisas e eficientes.

Deve-se considerar também que as táticas, técnicas e procedimentos (TTPs) aplicados por atacantes

atualmente se tornaram menos previsíveis, mais persistentes e complexos, melhor organizados e muitas vezes bem financiados [2]. Desse modo, a aplicação de métodos simples de monitoramento e a utilização de dispositivos de segurança comuns, se tornaram procedimentos incapazes de bloquear determinados tipos de ameaças.

2.1.3 Inteligência

Dentro da literatura, o termo inteligência possui as mais diversas definições. Isso pode ser explicado devido ao fato que a inteligência é um conceito fortemente relacionado à variável contexto. Em outras palavras, dependendo do escopo no qual é definido (acadêmico, psicológico, governamental, militar), o termo inteligência pode englobar diferentes conceitos. Uma definição generalizada do termo foi apresentada em [30] e pode ser aplicada de maneira propícia ao tema deste trabalho. Ela descreve a inteligência como o processo de transformar tópicos do estágio de total desconhecimento, para o de simples conhecimento de existência, até alcançar o estágio de completo entendimento do tópico. Uma maneira de exemplificar o fluxo pelo qual passa o ruído geral de dados até sua transformação em inteligência, pode ser vista na Figura 2.2. Os dados gerais e aleatórios são primeiramente filtrados em conjuntos menores de dados relevantes de ameaças, que em seguida são processados e transformados em informação. Essa informação, quando analisada e melhor contextualizada, se converte em conhecimento. Por fim, a capacidade de aplicar conhecimento de maneira correta e eficiente constitui a ideia de inteligência [2].

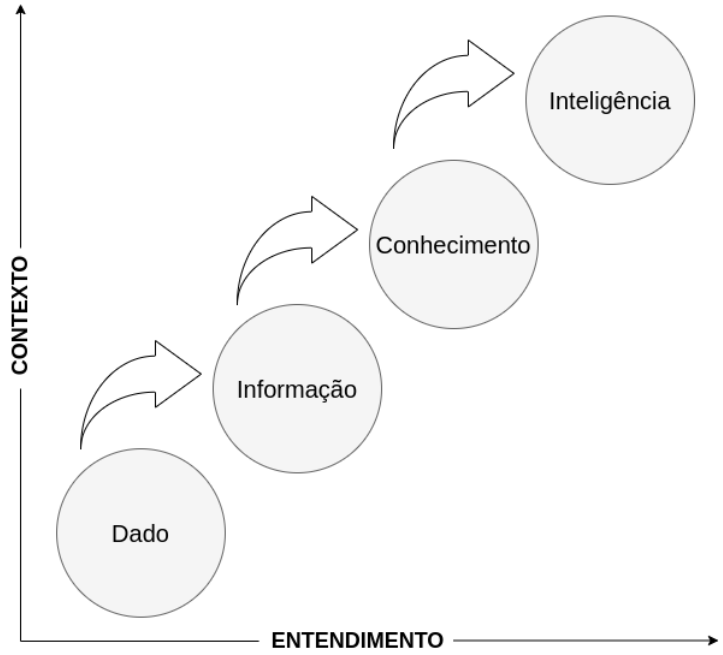


Figura 2.2: Evolução do dado até a formação de inteligência [1]

Tendo em vista os diferentes escopos nos quais o termo inteligência está inserido e considerando que devido aos diferentes interesses e mecanismos de ação aplicados em cada área, a inteligência pode possuir uma definição diferente. Desse modo, é válido dividir o conceito em três vertentes: inteligência humana,

inteligência artificial e inteligência de ação, como proposto em [31]. Dentro desta última, que foi definida como o resultado do processo que combina informações visando responder um problema específico, são estabelecidos quatro domínios: inteligência militar, inteligência de negócio, inteligência governamental e, por último, a inteligência de ameaça.

2.1.3.1 Inteligência de ameaça

O conceito de inteligência de ameaças, do inglês *Threat Intelligence* (TI), está inserido no conceito de inteligência de ação. Para que a inteligência seja considerada acionável, é necessário que ela atenda aos requisitos de ser oportuna, precisa e relevante [29]. Isso porque ela deve ser capaz de auxiliar no desenvolvimento de mecanismos proativos e eficientes para tomada de decisões. Seguindo essa perspectiva, o fluxo de produção de inteligência acionável deve incluir, além dos processos mencionados em 2.1.3, as etapas de implantação e disseminação da inteligência.

Desse modo, no contexto deste trabalho, o fluxo para produção de inteligência de ameaças é composto de cinco etapas principais, ilustradas na Figura 2.3:

1. **Coleta**: essa etapa se refere a extração e junção de dados, que são apenas fatos ou indicadores.
2. **Processamento**: trabalha na formatação e combinação dos dados, objetivando responder perguntas específicas para gerar informação.
3. **Análise**: avalia os dados e informações de forma conjunta para auxiliar na descoberta de padrões e na produção de inteligência acionável.
4. **Implantação**: após a produção da inteligência é possível implantá-la para garantir a tomada de decisão contra ameaças de forma proativa.
5. **Disseminação**: expande o conhecimento compartilhando com partes interessadas.

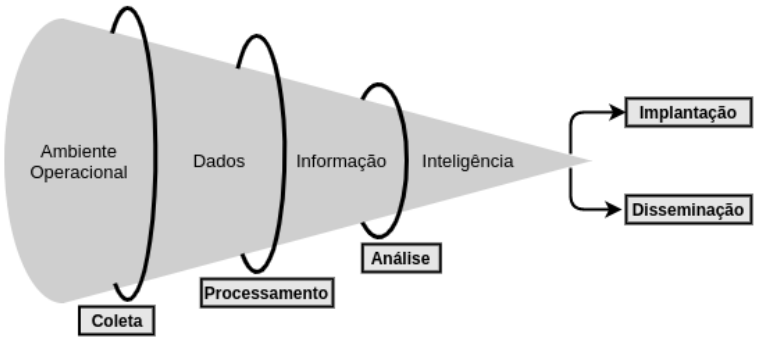


Figura 2.3: Fluxo para produção de inteligência de ameaça [2]

2.1.4 Inteligência de ameaças cibernéticas

Dentro do espectro de inteligência de ameaça, está o conceito de inteligência de ameaça cibernética, do inglês *Cyber Threat Intelligence* (CTI). Essa é uma abordagem relativamente nova, que se tornou altamente abrangente e passou a ser utilizada para definir diferentes tipos de serviços oferecidos. Pode ser considerada uma inteligência acionável gerada com base em evidências de mecanismos, indicadores, implicações e contextos relativos a ameaças ou incidentes no domínio cibernético. Fornece conhecimento sobre adversários e métodos que podem auxiliar no processo de tomada de decisão de resposta a ameaças [29].

Para que CTI seja aplicada corretamente e tenha resultados efetivos, é necessário estabelecer um fluxo de processos para sua produção [10]. Primeiramente, é importante entender as necessidades dos usuários da inteligência que está sendo desenvolvida e o contexto no qual ela está inserida [32], para que seus requisitos sejam definidos adequadamente.

Uma vez definidos os requisitos para a obtenção de inteligência de ameaças cibernéticas, inicia-se a etapa de coleta de dados. Sabe-se que dados e informações sem tratamento e contexto não são considerados inteligência, mas são os materiais básicos para sua produção. Então, existem mecanismos que consomem essas informações e realizam o processamento e a análise para gerar informações estruturadas e encontrar padrões. As informações tratadas podem ser integradas a outros mecanismos de defesa e então utilizadas para executar e desenvolver métodos de defesa e mitigação de ameaças [29]. Finalmente, como muitas organizações não possuem a capacidade de entender o cenário de ameaças cibernéticas de forma holística, o estágio de compartilhamento e disseminação de informações sobre ameaças entre as organizações é de extrema importância [28].

Por ser uma abordagem complexa, viu-se necessário o desenvolvimento de tecnologias e ferramentas que pudessem auxiliar e desempenhar os processos descritos na Figura 2.3.

2.1.4.1 Padrões de CTI

Um aspecto crucial de todo o processo de inteligência de ameaças é o formato dos dados compartilhados. Em primeiro lugar, para um processamento adequado e automatizado dos dados coletados, é importante que eles sejam formatados em um modelo estruturado e delineados em uma linguagem comum [22]. Além disso, o estabelecimento de padrões fornece uma definição prévia sobre o tipo de informação a ser compartilhada e a densidade dessa informação [33]. Como resultado, diversas iniciativas surgiram com o objetivo de padronizar as informações coletadas, consumidas e disseminadas no ecossistema do CTI [34].

2.1.4.2 Plataformas de CTI

O estabelecimento de um novo cenário de ameaças estimulou a mudança dos modelos tradicionais de defesa. Novos sistemas com ação proativa e recursos de resposta em tempo real a incidentes estão sendo desenvolvidos e comumente chamados de *Threat Intelligence Platforms* (TIPs) [34]. Eles são sistemas de software especializados que implementam os processos de coleta, processamento, análise, produção,

implantação e integração de inteligência de ameaças internas e externas. O principal objetivo deste tipo de plataformas é servir de assistente aos decisores relacionados com a resposta a incidentes [35].

2.2 TRABALHOS CORRELATOS

Embora muitos trabalhos tenham sido conduzidos acerca do tema inteligência de ameaça cibernética nos últimos anos, a maioria deles não possui foco em analisar e avaliar as soluções de CTI utilizadas atualmente. Por esse motivo, ainda existem muitos sistemas chamados de plataformas de CTI, porém incapazes de gerar informação de qualidade ou inteligência acionável. Além disso, considerando que este tipo de tecnologia evolui rapidamente, alguns trabalhos e resultados obtidos na área se tornam desatualizados. Assim, para obter um quadro detalhado dos trabalhos e oportunidades de pesquisa disponíveis na área, foi realizada uma revisão da literatura.

Vários esforços foram empenhados na tentativa de compreender e apresentar o panorama geral do tema inteligência de ameaça cibernética. Um survey [36] fornece uma ampla descrição do tópico CTI e menciona brevemente algumas plataformas e padrões utilizadas nesse contexto. Em [35] a pesquisa é mais focada em plataformas de TI e apresenta uma visão geral do cenário atual das plataformas disponíveis no mercado, incluindo software livre e plataformas comerciais. Outro trabalho [37] descreve algumas plataformas de CTI de código aberto selecionadas, porém nenhuma avaliação é feita. Um survey mais recente [38] discorre sobre oportunidades de pesquisa em relação de padrões para compartilhamento de informação de ameaças cibernéticas e menciona, sem realizar nenhum tipo de análise, algumas das linguagens mais populares para descrição e compartilhamento de CTI.

No contexto de avaliação das soluções existentes, incluindo padrões e plataformas de TI, algumas pesquisas foram conduzidas. Um dos primeiros trabalhos relevantes referentes a esse tópico [39], introduz 8 diferentes formatos de compartilhamento de dados se ameaça, sendo esses: *Common Intrusion Detection Framework* (CIDF), IODEF, *Common Announcement Interchange Format* (CAIF), *Intrusion Detection Message Exchange Format* (IDMEF), *Abuse Reporting Format* (ARF), *Common Event Expression* (CEE), *Extended Abuse Reporting Format* (X-ARF) and *Syslog*. Esses formatos são analisados com base em uma metodologia de avaliação proposta que consiste em 10 diferentes critérios como interoperabilidade, confidencialidade e aplicação prática. A avaliação desempenhada aplica uma metodologia significativa e produz bons resultados, porém alguns formatos importantes, que são atualmente relevantes, não foram abordados, mostrando que alguns resultados estão desatualizados. Em [24], o panorama completo do CTI é abordado e alguns padrões de CTI são apresentados. Além disso, é feita uma avaliação satisfatória de algumas ferramentas de inteligência de ameaças de código aberto com o objetivo de compará-las com a ferramenta proposta no trabalho. Em [40], é realizada uma classificação e análise de 23 plataformas de inteligência de ameaça com base no modelo de licença, padrões suportados, tipo de plataforma e tipo de informação compartilhada. O resultado da análise apresenta alguns fatos interessantes sobre o panorama do CTI, como a descoberta de que a maioria das plataformas de inteligência de ameaças são de código fechado, a descrição do STIX como o padrão de-facto para descrever a inteligência de ameaças e a descoberta de que a maioria das plataformas prioriza o compartilhamento à análise das informações. No entanto, os resultados são consolidados em oito tópicos, o que não permite um conhecimento aprofundado das funcionalidades e

funcionamento das plataformas.

Na mesma perspectiva, um trabalho recente [17], fornece uma análise comparativa de fontes, formatos e linguagens de CTI. Diversas fontes de CTI são apresentadas e examinadas e, com base nos resultados da avaliação em conjunto com uma pesquisa bibliográfica, alguns padrões de CTI foram selecionados para análise posterior. Muitos critérios e recursos foram considerados na comparação, fornecendo uma descrição excelente e detalhada dos recursos de alguns padrões de CTI relevantes. Outro trabalho relevante ao tema [18] apresenta um *framework* capaz de analisar e comparar plataformas de compartilhamento de inteligência de ameaça. Com base em uma revisão sistemática da literatura, 40 publicações diferentes, que continham características ou requisitos de Threat Intelligence Sharing Platforms (TISPs), foram estudadas. Com isso, 62 critérios essenciais de avaliação foram determinados e divididos em seis categorias principais que foram utilizadas pelo *framework* proposto para avaliar as plataformas. O trabalho menciona que o método proposto foi aplicado a 10 TISPs diferentes, mas apenas três delas tiveram os resultados descritos.

Os estudos focados em descrever e avaliar o cenário de soluções de CTI apresentam alguns padrões e plataformas satisfatórias, no desempenho de sua função, porém ainda existem limitações a serem enfrentadas para implementar mecanismos de CTI como modelo de defesa. Desse modo, alguns trabalhos foram realizados visando propor métodos para solução desses obstáculos.

Uma das principais limitações acerca do assunto é a qualidade dos dados produzidos e compartilhados. Na grande maioria das soluções disponibilizadas, a quantidade de dados compartilhados acaba sendo priorizada em relação a qualidade dos mesmos [13] e o uso de técnicas de *machine learning* se tornou uma abordagem bastante utilizada para mitigar essa limitação. Em [41] são apresentadas boas perspectivas para o uso de inteligência artificial no contexto de segurança cibernética. A pesquisa discorre sobre a enorme quantidade de dados de ameaças disponíveis e coloca os poderosos recursos de automação e análise de dados disponíveis com o uso de técnicas de aprendizado de máquina como uma solução para lidar com o volume de dados. Em [42], utilizando algoritmos de aprendizado de máquina, um *framework* foi desenvolvido para coletar e analisar dados e atribuir incidentes de ameaças a seus atores. Outro trabalho [43], propôs uma plataforma de TI com uma arquitetura baseada em sistemas de última geração, como Malware Information Sharing Platform (MISP) e Collaborative Research into Threats (CRITs). As plataformas aplicam algoritmos de aprendizado de máquina para analisar e classificar o conteúdo de emails e se defender ativamente contra a engenharia social.

Seguindo a mesma abordagem de otimização da qualidade dos dados de TI, alguns trabalhos desenvolveram métodos de enriquecimento de indicadores de comprometimento. Em [44], é proposta uma plataforma capaz de classificar automaticamente dados de ameaças em diferentes categoria de ameaça e correlacionar eventos de mesma categoria. Em [45], para aumentar a estimativa de impacto de indicadores de ameaça, Kazato et al. utiliza um método baseado em rede convolucional de grafos. O método proporcionou uma melhoria na precisão das estimativas de ameaça e impacto de indicadores e reduziu o tempo alocado para análises manuais de indicadores. Na mesma linha, em [21] é proposto um novo método para extrair automaticamente indicadores e aplicar marcas de domínio de mídia social. O método inclui uma rede neural convolucional para reconhecer domínios de CTI e classificar corretamente os dados de ameaças nesses domínios.

Em síntese, os principais trabalhos relacionados aos objetivos deste trabalho foram sumarizados na

Tabela 2.1. Tendo em vista o exposto, a maioria das iniciativas e trabalhos desenvolvidos na área, com o objetivo de comparar ou avaliar as soluções de CTI existentes, se concentra na comparação de um grande número de plataformas ou padrões e não fornece uma análise crítica de cada uma delas. Além disso, poucos trabalhos apresentam métodos de avaliação relevantes e efetivos. Em paralelo, ainda existem diversos desafios a serem enfrentados no sentido de otimizar a qualidade dos dados produzidos por soluções de CTI. Desse modo, é notória uma oportunidade de pesquisa relacionada a análise de plataformas de CTI, com foco em encontrar potenciais soluções para a otimização da qualidade dos dados produzidos.

Tabela 2.1: Síntese dos principais trabalhos correlatos

Foco	**Abordagem**	**Método**	**Referência**
Estudo, análise e comparação de formatos e protocolos utilizados no contexto de eventos de segurança	Divisão entre formatos e protocolos e proposição de critérios de avaliação para cada categoria	Revisão sistemática da literatura	[39]
Estudo do panorama geral do tema CTI e proposta de uma nova ferramenta de TI	Estudo e avaliação de ferramentas de TI de código aberto visando comparar funcionalidades com uma nova ferramenta proposta no trabalho	Análise de documentação oficial e literatura acadêmica	[24]
Análise e comparação do mercado de plataformas de compartilhamento de inteligência de ameaça	Comparar 22 plataformas de TI, incluindo código aberto e fechado, considerando diferentes quesitos de avaliação	Revisão sistemática da literatura	[40]
Explorar as capacidades de formatos e linguagens de TI disponíveis	Selecionar padrões proeminentes e avaliar a capacidade de transmissão e correlação desses padrões quando aplicados a diferentes fontes de dados	Análise de documentação e utilização prática das ferramentas	[17]
Apresentar um framework para análise e comparação de plataformas de compartilhamento de inteligência de ameaça	Realizar uma revisão da literatura para obter critérios de avaliação relevantes e combina-los em uma ferramenta de avaliação	Demonstração prática da ferramenta desenvolvida em três TIPs	[18]
Proposta de plataforma para produção de inteligência de ameaça de qualidade	Utilização da plataforma MISP em conjunto com módulos de enriquecimento, análise e classificação	Desenvolvimento de protótipo funcional	[44]
Desenvolvimento de sistema para produção de inteligência de ameaça	Processar linguagem natural proveniente de diferentes fontes e converter em inteligência acionável	Aplicação prática e análise dos resultados obtidos	[43]
Estimar o grau de impacto de IoCs	Utilização de redes neurais para grafos visando melhorar a acurácia na análise de impacto de um indicador	Resultados experimentais e comparação com outros métodos	[45]
Proposta de framework para extração de IoCs de documentos não estruturados	Produção de IoCs em linguagem de alto nível utilizando técnicas de Machine Learning	Aplicação prática e análise dos resultados obtidos	[42]

3 FERRAMENTAS DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

Com a popularização do tema inteligência de ameaças cibernéticas, hoje o conceito de CTI carece de uma definição consistente [40] [23] e um mercado heterogêneo de plataformas de CTI emergiu. Diversos sistemas e ferramentas, que cumprem objetivos diferentes, são implementados como soluções completas de inteligência de ameaça [18]. Além disso, funcionalidades, níveis de desempenho e casos de uso aplicáveis sofrem grande diferenciação entre as plataformas existentes.

Nesse contexto, notou-se uma oportunidade de pesquisa envolvendo a análise das ferramentas disponíveis no mercado a fim de descrever detalhadamente suas características e analisar a qualidade das informações produzidas por elas. Desse modo, um estudo comparativo foi conduzido em [22], com o objetivo principal de desenvolver uma metodologia de avaliação para as plataformas de CTI existentes no mercado.

Para alcançar resultados relevantes, o estudo realizado em [22] foi fragmentado em três etapas principais. Primeiro, foi feita uma revisão dos padrões e plataformas existentes no contexto de CTI visando identificar soluções com potencial de consolidação. Em seguida, foi elaborada uma estratégia de seleção para definir os padrões e plataformas mais populares atualmente. Por fim, as soluções selecionadas foram analisadas, de forma prática, e avaliadas com base em critérios de avaliação holísticos.

3.1 METODOLOGIA DE AVALIAÇÃO DE PADRÕES E PLATAFORMAS DE CTI

Esta seção descreve a abordagem proposta para avaliar os padrões e plataformas de CTI. Primeiro, um método é desenvolvido para restringir e definir quais padrões e plataformas serão analisados no escopo deste estudo. Em seguida, os critérios de avaliação para as soluções selecionadas são introduzidos e explicados.

3.1.1 Estratégia de Seleção

O cenário de inteligência de ameaças atual é bastante extenso e inclui uma ampla gama de sistemas, plataformas, ferramentas, padrões e formatos. Para realizar uma avaliação relevante e precisa, é necessário definir uma estratégia de inclusão e exclusão para selecionar alguns deles. Assim, no contexto desse estudo, serão considerados padrões e plataformas de código aberto e populares na comunidade.

Com o objetivo de obter um panorama completo do panorama do CTI, foi realizada uma revisão da literatura acerca do tema. Para auxiliar nessa tarefa, alguns mecanismos de pesquisa na Internet foram utilizados, incluindo: Google Scholar, Biblioteca Digital IEEE, Springer e ScienceDirect. As seguintes *strings* de busca foram combinadas para encontrar resultados relevantes: (*Threat Intelligence* OR *Cyber Threat Intelligence*) AND (*platform* OR *tools* OR *standards*). O processo de busca resultou em um grande número de trabalhos e alguns deles foram filtrados de acordo com sua relevância ao tema e ao número de

citações.

Mesmo após a aplicação de um primeiro processo de filtragem, uma quantidade considerável de iniciativas de padrões e plataformas de CTI foram coletados. Assim, para reduzir o número de resultados encontrados, com base em breves leituras dos *sites* e documentação oficial, padrões ou plataformas sem a capacidade de abordar dois ou mais estágios do fluxo do processo de produção de inteligência de ameaças, apresentado na 2.3, não foram considerados para o processo de seleção.

Em seguida, visando selecionar os padrões e plataformas mais relevantes no campo de CTI para serem avaliados, os resultados encontrados com o processo de pesquisa foram descritos em termos de popularidade e modelo de licença:

1. **Popularidade:** No contexto deste trabalho, a popularidade foi estimada de acordo com o número de vezes que o padrão ou plataforma foi mencionado em trabalhos e fontes confiáveis, combinado com estatísticas coletadas sobre a porcentagem de utilização entre as organizações.
2. **Modelo de licença:** O tipo de licença foi analisado visando limitar a obra aos padrões ou plataformas disponíveis para todas as comunidades interessadas, incluindo apenas soluções e iniciativas gratuitas ou de código aberto.

Por fim, com base nos critérios supracitados, um conjunto de resultados de tamanho razoável, composto pelos padrões e plataformas mais populares e com modelo de licença livre ou permissiva, foi selecionado para avaliação.

3.1.2 Critérios de Avaliação

Para definir os critérios de avaliação dos padrões e plataformas de CTI, dois aspectos principais foram levados em consideração. Primeiro, foi analisada a aplicabilidade em diferentes casos de uso, por meio da definição de um modelo de dados holístico. Depois, alguns critérios gerais de avaliação foram inferidos do fluxo de produção de inteligência.

3.1.2.1 Modelo de Dados

Em termos de arquitetura, para definir a aplicabilidade em diferentes casos de uso, quatro entidades principais foram utilizadas para representar o cenário de inteligência de ameaças cibernéticas de uma maneira genérica. Nesta abordagem, as entidades derivam indiretamente do método 5W3H (*what, who, why, when, where, how, how much and how long*), que visa responder às questões apresentadas na Tabela 3.1. Este é um método genérico, aplicado em diferentes áreas, com o objetivo principal de esclarecer um tópico em sua completude [8].

Método 5W3H

Esse método foi escolhido como parâmetro de para representação do modelo de dados por ser um mecanismo efetivo para obter a caracterização completa de uma ameaça, dado que as questões levantadas por ele auxilia na geração de conhecimento e tomada de decisão. Além disso, ao correlacionar dados e indicado-

Tabela 3.1: Descrição do método 5W3H [8][9]

Questionamento	Descrição
What	Diretamente descreve o tópico em questão
Where	Específica referências de localização sobre o tópico
When	Específica informações temporais relevante ao tópico como data e hora
Who	Associa o tópico a uma entidade capaz de executá-lo
Why	Descreve possíveis motivações para a ocorrência do tópico
How	Descreve as principais características e mecanismos do tópico
How much	Refere-se aos custos e impactos gerados pelo tópico
How long	Descrição da eficácia do tópico em termos de tempo

res com informações geradas a partir das questões levantadas pelo método 5W3H, a tarefa de implementar mecanismos eficazes para detectar e mitigar ameaças se torna mais simples.

Primeiro, *what* é usado para definir diretamente o tópico que está sendo analisado. No contexto de inteligência de ameaças cibernéticas, normalmente pode ser resumido pelos termos ameaças e incidentes, que abrangem desde evidências e prováveis ataques até ocorrências maliciosas reais. Para uma definição adequada, esses termos devem vir acompanhados de outras informações como o tipo de ameaça ou incidente e o contexto em que está inserido. Depois de definido, o tópico pode ser caracterizado usando *where*, *when* e *how*. *Where* pode se referir à localização geográfica na qual teve inicio, além de partes do caminho que percorreu até chegar ao alvo. *When* fornece um intervalo de tempo, especificando a data e hora da ocorrência. *How* é usado para descrever a forma como a ameaça ou incidente ocorreu e as táticas e técnicas aplicadas. É importante dizer que a granularidade das informações que descrevem essas entidades é variável, dependendo do caso de uso.

Outro ponto essencial é associar a ameaça ou incidente ao seu ator, que pode ser descrito por *who* e *why*. *Who* pode ser uma organização ou indivíduo responsável pela ameaça ou incidente. *Why* é importante para caracterizar melhor o ator da ameaça, entendendo as motivações por trás do evento.

Algumas características detalhadas sobre a ameaça ou incidente podem ser descobertas utilizando *how long* e *how much*. *How long* indica a durabilidade efetiva da ameaça ou incidente se nenhuma ação for tomada. *How much* é utilizado para medir a intensidade do ataque e analisar sua capacidade de dano e custo de defesa. A informação recolhida com as declarações *how long* e *how much*, juntamente com todas as características descritas com a declaração *how*, também podem ser utilizadas para analisar e medir a capacidade de ação de um adversário.

Desse modo, ao utilizar a correlação entre todas as informações levantadas a partir da aplicação do método 5W3H, é muito provável que inteligência acionável tenha sido produzida.

Tendo em vista o exposto, as quatro entidades principais utilizadas para delinear uma representação holística do cenários de inteligência de ameaças cibernéticas são: ameaça, incidente, ator da ameaça e defesa. Um diagrama, apresentado na Figura 3.1, ilustra o contexto em que essas entidades estão inseridas e as relações entre elas.

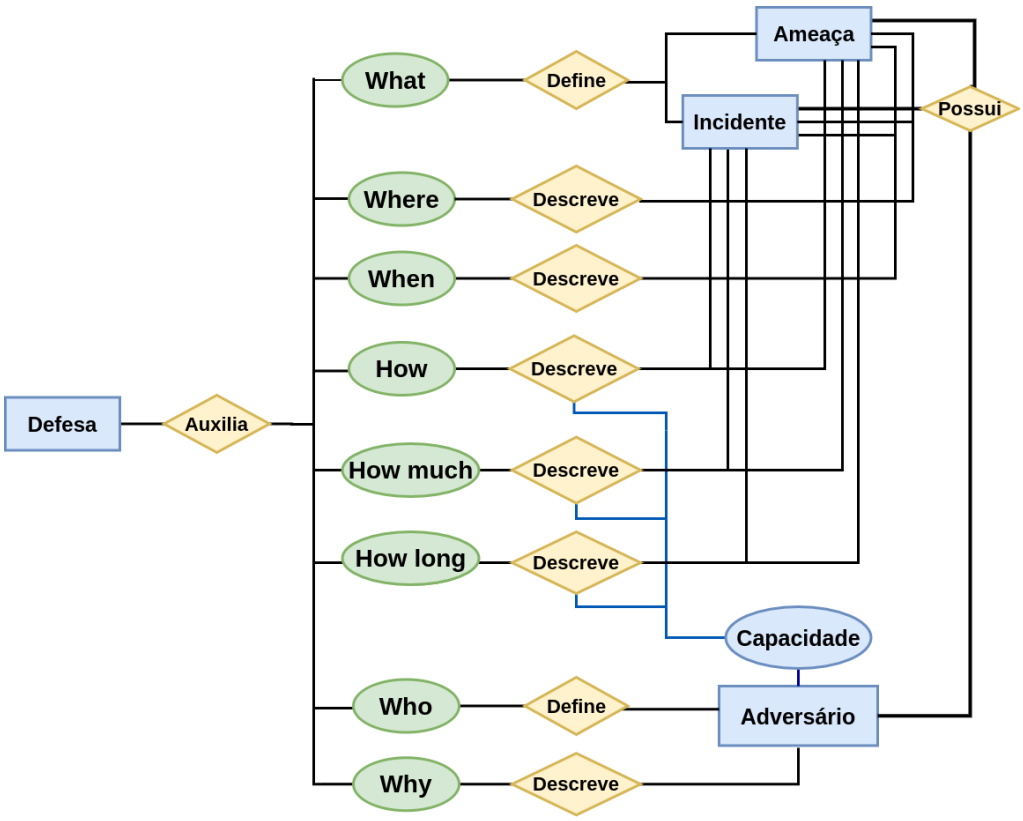


Figura 3.1: Diagrama de relacionamento entre entidades e o método 5W3H

3.1.2.2 Fluxo de Produção de Inteligência

Visando avaliar critérios gerais, recursos e funcionalidades essenciais para atingir um fluxo completo do processo de produção de inteligência foram delineados, incluindo alguns critérios propostos em [39] and [33]. Considerando o fluxo de produção de inteligência de ameaça apresentado na seção inteligencia-ameaca, para a etapa de coleta é importante fornecer os dados em um formato comum para facilitar o processo de coleta. Em seguida, para processar e normalizar os dados, um formato estruturado e legibilidade por máquinas são essenciais. Além disso, dados mais leves propiciam um sistema mais eficiente em processamento. A etapa de análise requer um modelo de dados não ambiguo para realizar correlações e classificar as informações, além de mecanismos de relacionamento para representar essas correlações. Com a informação analisada acessível, interoperabilidade entre formatos, sistemas e plataformas são necessárias para que a inteligência acionável possa ser implantada de forma correta e automática. Por fim, para disseminar informações e inteligência, em conjunto com alguns aspectos supracitados, é relevante a adoção de um mecanismo de transporte específico e um bom uso prático na comunidade.

3.1.2.3 Adicional

Ao tratar de plataformas de CTI, considera-se que a facilidade de uso e flexibilidade para implementação de novos recursos são aspectos relevantes, e por isso alguns critérios adicionais foram aplicados. Portanto, a quantidade e qualidade da documentação e as permissões declaradas em suas licenças também foram avaliadas.

Baseado no exposto, todos os critérios de avaliação de padrões e plataformas de CTI foram definidos. As Tabelas 3.2 e 3.3 resumem os critérios descritos nesta seção.

Tabela 3.2: Critérios de avaliação para padrões de CTI

Modelo de Dados	
Arquitetura Holística	Ameaça
	Incidente
	Adversário
	Defesa
Processo de Produção de Inteligência	
Coleta	Formatação comum
Processamento	Formato estruturado
	Tamanho
	Legibilidade para máquinas
Análise	Modelo não ambíguo
	Mecanismos de Relacionamento
Implantação	Interoperabilidade
Disseminação	Mecanismo de transporte
	Aplicação prática

Tabela 3.3: Critérios de avaliação para plataformas de CTI

Modelo de Dados	
Arquitetura Holística	Casos de uso
Método 5W3H	Capacidade de resposta
Processo de Produção de Inteligência	
Coleta	Fomato de importação
	Coleta automática
Processamento	Formato de exportação
	Visualização gráfica
Análise	Correlação
	Classificação
Implantação	Integração com outros sistemas
Disseminação	Mecanismo de compartilhamento
Adicionais	
Usabilidade	Documentação
	Modelo de Licença

3.2 RESULTADOS DA AVALIAÇÃO DOS PADRÕES DE CTI

Para iniciar o processos de avaliação, os padrões foram selecionados com base na estratégia proposta na Seção 3.1.1. Além disso, os resultados da avaliação foram baseados nos critérios de avaliação propostos na Seção 3.1.2. Estes resultados relativos à seleção e avaliação dos padrões são apresentados e explicados.

Após coletar os resultados mais relevantes derivados do processo de busca por padrões de CTI, algumas iniciativas interessantes foram encontradas. Em [24] e [46] alguns projetos que buscam padronizar dados

de inteligência de ameaças são mencionados, como STIX, TAXII, CybOX, OpenIOC, CAPEC, MAEC e ATT&CK, sendo o STIX considerado o padrão mais utilizado.

Em [33] outros padrões foram mencionados, como VERIS, STATL, ARF and X-ARF e alguns deles são avaliados. Outros trabalhos [40] é [37] apenas comentam sobre padrões considerados consolidados no mercado, sendo esses: OpenIOC, CybOX, STIX, TAXII e IODEF. Um survey [47] apresenta, em termos estatísticos, os padrões mais utilizados: STIX, OpenIOC, CybOX e IODEF. Em [8] é realizada uma comparação entre os padrões IODEF/RID e STIX/TAXII, considerados os mais populares. Por fim, alguns trabalhos recentes [21][20] descrevem padrões considerados proeminentes atualmente.

Como todos os padrões encontrados são liberados gratuitamente para uso da comunidade, a popularidade foi o principal critério para selecioná-los. Considerando os resultados obtidos com a pesquisa bibliográfica, complementada com a revisão dos sites oficiais e documentações da maioria dos padrões, os padrões foram classificados em termos de popularidade e os resultados estão apresentados na Tabela 3.4.

Tabela 3.4: Padrões de CTI descritos por popularidade e modelo de licença

Padrão	**Popularidade**	**Modelo de Licença**	**Referências**
STIX	++++	Licença permissiva garantida pela MITRE	[21][20][17][18][40][38] [24][33][28]
TAXII	++++	Licença permissiva garantida pela MITRE	[21][17][40][8][38][24] [33]
CybOX	+++	Licença permissiva garantida pela MITRE	[21][40][24][47]
IODEF	++	IETF TLP 5.0	[21][20][38][24][33]
RID	++	IETF TLP 5.0	[8][24][33]
OpenIOC	+++	Apache 2.0	[21][40][36][24]
VERIS	+	*Attribution-ShareAlike 4.0 International*	[38][33]
X-ARF	+	*Open Source* (GNU *General Public License*)	[33] [39]
Legenda: muito alto (++++) alto (+++) médio (++) baixo (+)			

3.2.1 Padrões Selecionados para Avaliação

Diante dos resultados apresentados, os padrões selecionados para posterior análise e avaliação são: STIX, TAXII, IODEF, RID, CybOX e OpenIOC. A seguir, é fornecida uma apresentação sucinta dos padrões selecionados.

3.2.1.1 Structured Threat Information eXpression

STIX é uma linguagem criada pela MITRE e desenvolvida para a especificação, captura, caracterização e comunicação de informações no contexto de inteligência de ameaças cibernéticas [28]. Ela provê mecanismos para representar essas informações de forma estruturada em diferentes cenários do ecossistema de ameaças cibernéticas.

Essa linguagem foi projetada tendo em vista princípios como interoperabilidade, extensibilidade, foco em automatização e legibilidade por máquinas. Em sua primeira versão, a arquitetura do STIX foi modelada em formato XML, com atributos e campos previamente definidos, e era composta por oito *cores* principais. A segunda versão da linguagem foi desenvolvida utilizando serialização em formato JSON, fato que proporcionou otimizações no *overhead* de processamento da informação [48]. A utilização do formato JSON na representação de objetos STIX também proporcionou um entendimento mais intuitivo da infor-

mação, como pode ser visto no exemplo apresentado nas Figuras 3.2 e 3.3. Analisando graficamente o diagrama de relacionamento entre os objetos STIX, disposto na Figura 3.3, nota-se que devido a clareza no modelo adotado para descrição da informação que produz o diagrama, a legibilidade essa informação possui uma baixa complexidade tanto para máquinas como para seres humanos.


```
{
    "type": "bundle",
    "id": "bundle--601cee35-6b16-4e68-a3e7-9ec7d755b4c3",
    "objects": [
        {
            "type": "threat-actor",
            "spec_version": "2.1",
            "id": "threat-actor--dfaa8d77-07e2-4e28-b2c8-92e9f7b04428",
            "created": "2014-11-19T23:39:03.893Z",
            "modified": "2014-11-19T23:39:03.893Z",
            "name": "Disco Team Threat Actor Group",
            "description": "This organized threat actor group operates to create profit from all types of crime.",
            "threat_actor_types": [
                "crime-syndicate"
            ],
            "aliases": [
                "Equipo del Discoteca"
            ],
            "roles": [
                "agent"
            ],
            "goals": [
                "Steal Credit Card Information"
            ],
            "sophistication": "expert",
            "resource_level": "organization",
            "primary_motivation": "personal-gain"
        },
        {
            "type": "identity",
            "spec_version": "2.1",
            "id": "identity--733c5838-34d9-4fbf-949c-62aba761184c",
            "created": "2016-08-23T18:05:49.307Z",
            "modified": "2016-08-23T18:05:49.307Z",
            "name": "Disco Team",
            "description": "Disco Team is the name of an organized threat actor crime-syndicate.",
            "identity_class": "organization",
            "contact_information": "disco-team@stealthemail.com"
        },
        {
            "type": "relationship",
            "spec_version": "2.1",
            "id": "relationship--a2e3efb5-351d-4d46-97a0-6897ee7c77a0",
            "created": "2020-02-29T18:01:28.577Z",
            "modified": "2020-02-29T18:01:28.577Z",
            "relationship_type": "attributed-to",
            "source_ref": "threat-actor--dfaa8d77-07e2-4e28-b2c8-92e9f7b04428",
            "target_ref": "identity--733c5838-34d9-4fbf-949c-62aba761184c"
        }
    ]
}
```


Figura 3.2: Exemplo de implementação com padrão STIX [3]

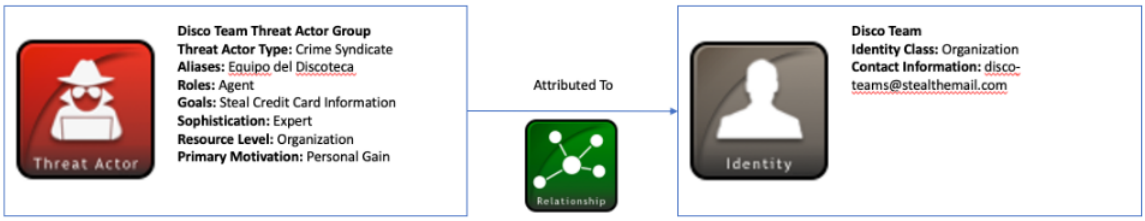


Figura 3.3: Diagrama de relacionamento STIX obtidos a partir do JSON disposto na Figura 3.2 [3]

Em sua segunda versão, a arquitetura estrutural do padrão STIX foi significativamente modificada, sendo composta atualmente por doze objetos principais que correspondem diretamente a conceitos embutidos no contexto de CTI [49]. Com sua arquitetura holística, a STIX é capaz apresentar a informação de forma padronizada, compreensiva e estruturada ao mesmo tempo que permite aplicação em diferentes casos de uso. Além disso, é diretamente integrável com outras linguagens do contexto de CTI, fato que expande ainda mais seu perímetro de aplicação [36].

CybOX é uma linguagem criada pela MITRE e desenvolvida para especificar, caracterizar e comunicar informações sobre observáveis cibernéticos, de maneira padronizada e com alta fidelidade. Era comumente utilizada em conjunto com a linguagem STIX para contextualização da informação e consequente concepção de inteligência de ameaça. CybOX foi integrado a segunda versão da linguagem STIX, e atualmente não é mais utilizada como uma linguagem independente. A especificação do STIX 2.0 define uma representação estruturada para objetos observáveis no domínio cibernético, denominada de *Cyber Observable Object* [50]. Essa padronização pode ser utilizada para descrever diversos tipos de dados, desde a caracterização de um host até informação de forense digital. A representação dos objetos é feita com serialização no formato JSON [51]. Assim como em sua primeira idealização, o CybOX é geralmente utilizado em conjunto com os objetos do core da linguagem STIX para adicionar contexto aos dados descritos por cada objeto.

3.2.1.2 Trusted Automated Exchange of Intelligence Information

TAXII é um protocolo da camada de aplicação, que define um conjunto de serviços de trocas de mensagens para detecção, prevenção, mitigação e compartilhamento de informações de CTI entre organizações [52]. Ele foi projetado pela MITRE especificamente para o transporte de informações formatadas na linguagem STIX, porém não é limitado a ela. Esse protocolo utiliza o protocolo HTTPS para transporte das mensagens e suporta diferentes modelos de compartilhamento, incluindo *hub-and-spoke*, P2P e *publish-subscribe*.

3.2.1.3 Incident Object Description Exchange Format

IODEF surgiu como uma iniciativa da *Internet Engineering Task Force* (IETF) para facilitar o compartilhamento de informações entre organizações e aumentar a probabilidade de mitigação de ameaças cibernéticas [53]. Em sua primeira versão, define um modelo de representação de dados para compartilhamento de informações focadas em incidentes cibernéticos. Trazendo uma abordagem mais holística, a segunda versão do IODEF trouxe uma significante evolução em sua parte estrutural, que passou a incluir estruturas para descrição de indicadores, atacantes e metodologias de resposta aos incidentes [54]. Ambas as versões utilizam o formato XML.

Como o modelo IODEF inclui uma grande variedade de classes e tipos de dados que são definidos em seu esquema XML, existem diversas construções disponíveis que podem ser utilizadas para descrever um incidente ou ameaça. Essa grande variabilidade de atributos pode ser um desafio na implementação do IODEF e no desenvolvimento de ferramentas interoperáveis [4]. Além disso, a legibilidade das informações em XML, conforme mostra o exemplo na Figura 3.4, se torna menos intuitiva para seres humanos.

```
<!-- ...XML code omitted... -->
 <IndicatorData>
   <Indicator>
     <IndicatorID name="csirt.example.com" version="1">
     G90823490
     </IndicatorID>
     <Description>C2 domains</Description>
     <IndicatorExpression operator="and">
       <IndicatorExpression operator="or">
         <Observable>
           <System category="source" spoofed="no">
             <Node>
               <Address category="ipv4-addr">
                 10.10.10.104
               </Address>
             </Node>
           </System>
         </Observable>
         <Observable>
           <System category="source" spoofed="no">
             <Node>
               <Address category="ipv4-addr">
                 10.10.10.106
               </Address>
             </Node>
           </System>
         </Observable>
       </IndicatorExpression>
       <Observable>
         <System category="target" spoofed="no">
           <Node>
             <Address category="ipv4-addr">
               10.1.1.1
             </Address>
           </Node>
         </System>
       </Observable>
     </IndicatorExpression>
   </Indicator>
 </IndicatorData>
<!-- ...XML code omitted... -->
```

Figura 3.4: Exemplo de implementação com padrão IODEF [4]

3.2.1.4 Real-time Inter-network Defense

RID é uma iniciativa desenvolvida pela IETF, com o objetivo de facilitar o compartilhamento de dados sobre incidentes de segurança, principalmente estruturadas sob o formato IODEF. Ele esboça uma rede de comunicação interna proativa, com capacidade de integração com mecanismos de detecção, identificação, mitigação e resposta a incidentes, visando constituir uma solução completa no tratamento de incidentes de segurança [55]. As mensagens RID são transportadas sob o protocolo HTTPS e, para fornecer mais segurança, o protocolo adiciona outra camada de segurança para gerenciar as sessões.

3.2.1.5 Open Indicator Of Compromise

OpenIOC é um *framework* que oferece um formato padronizado e estruturado, legível por máquinas, utilizado para registrar, definir e compartilhar informações englobadas no contexto de ataques e incidentes

cibernéticos. Esse formato é escrito em XML, com um design razoavelmente leve e pequeno, exemplificado na Figura 3.5. O *framework* é focado em realizar uma descrição técnica de indicadores, em um formato assimilável por diferentes dispositivos empregados no cenário da segurança da informação. Sua arquitetura é composta por mais de 500 tipos específicos de dados incorporados para representação de indicadores.

DOMAIN_852808bf99be59a2902e089e26d5976a.ioc

```
<ioc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns="http://schemas.mandiant.com/2010/ioc"
id="df83c379-d757-5242-971e-640c92016a0b" last-modified="2016-10-27T14:01:31.6275083Z">
  <short_description>ioc for ddns.net</short_description>
  <description>ZONE:Green</description>
  <authored_by>KasperskyThreatLookup</authored_by>
  <authored_date>2016-10-27T14:01:31.6275083Z</authored_date>
  <links />
  <definition>
    <Indicator operator="OR" id="b6fe04cb-dfbd-5205-a920-00f7a62308cc">
      <IndicatorItem condition="contains" id="852808bf-99be-59a2-902e-089e26d5976a">
        <Context document="Network" search="Network/URI" type="mir" />
        <Content type="string">DDNS.NET</Content>
      </IndicatorItem>
      <Indicator operator="AND" id="cfbcee05-35d8-5f37-8837-62fd716b7e1b">
        <Indicator operator="OR" id="dfac3ba6-e2ed-5b9f-b1d3-aa1ef3046bc2">
          <IndicatorItem condition="is" id="cfcf62ca-6194-5916-becb-024a6cd5db18">
            <Context document="DnsEntryItem" search="DnsEntryItem/RecordData/IPv4Address" type="mir"
/>
            <Content type="IP">213.128.81.34</Content>
          </IndicatorItem>
          <IndicatorItem condition="is" id="e2d0d2bb-cbd9-5bf7-ad07-7de1c4d9e366">
            <Context document="DnsEntryItem" search="DnsEntryItem/RecordData/IPv4Address" type="mir"
/>
            <Content type="IP">8.23.224.108</Content>
          </IndicatorItem>
        </Indicator>
      </Indicator>
    </Indicator>
</definition>
</ioc>
```

Figura 3.5: Exemplo de implementação com padrão OpenIOC [5]

3.2.2 Avaliação dos Padrões

Posteriormente à seleção e definição dos padrões, foi feita uma avaliação com base em literatura acadêmica, estudo de documentação oficial e demonstrações práticas. Levando em consideração os critérios supracitados, a avaliação dos padrões está resumida na Tabela 3.5.

Em relação aos resultados ilustrados na Tabela 3.5, duas considerações devem ser destacadas. Primeiro, como explicado antes na Seção 3.2.1.1, uma vez que CybOX e STIX eram comumente utilizados juntos e ambos os padrões são mantidos pela mesma organização, CybOX foi integrado à segunda versão do STIX e não é mais usado como uma linguagem independente. Assim, como CybOX agora faz parte da estrutura STIX, foi considerado mais plausível avaliá-los como um único padrão. Em segundo lugar, percebeu-se que, embora TAXII e RID sejam protocolos autônomos, eles são utilizados principalmente em combinação com STIX e IODEF, respectivamente. Isso decorre do fato de que TAXII e RID são protocolos projetados especificamente para facilitar o transporte de STIX e IODEF. Assim, optou-se por avaliar esses protocolos como pares (STIX/TAXII e IODEF/RID), considerando o fato de suas funções serem complementares.

Na perspectiva da arquitetura dos dados, STIX é a linguagem com a arquitetura mais holística e é aplicável em diferentes casos de uso. As quatro entidades consideradas essenciais para delinear uma contextualização holística do cenário de inteligência de ameaças cibernéticas pode ser totalmente representada e caracterizada com as classes que compõem o esquema STIX. IODEF e OpenIOC também têm uma arquitetura satisfatório, porém com alguns pontos de falha. Ambos os padrões não possuem os recursos necessários para uma definição adequada dos mecanismos de defesa ou curso de ações. Além disso, OpenIOC tem alguns deficiências no processo de caracterização de um adversário de forma mais específica.

Do ponto de vista do fluxo de produção de inteligência, o STIX tem a capacidade de atender à maioria dos requisitos propostos. O uso de serialização no formato JSON fornece um formato comum e estruturado, com baixo overhead e legibilidade por máquina. Os doze objetos que compõem a arquitetura STIX são bem descritos e documentados, fornecendo um modelo de dados não ambíguo com mecanismos de relacionamento acoplados. Quando utilizado em conjunto com o TAXII, oferece um mecanismo de transporte confiável. Finalmente, uma vez que tem uma aplicação prática significativa, a maioria das plataformas e ferramentas de CTI possuem métodos de integração com este padrão.

IODEF e OpenIOC são baseados no formato XML, então também fornecem um formato comum, estruturado e legível por máquinas. No entanto, IODEF pode apresentar alguns problemas devido aos campos de texto livre que compõem seu modelo de dados. Quanto aos mecanismos de relacionamento, OpenIOC fornece operadores lógicos (E/OU) para criar conexões entre indicadores, por outro lado, além de interconexões no modelo de dados, o IODEF não apresenta mecanismos específicos para relacionar informações. IODEF e OpenIOC são suportados em muitas plataformas e ferramentas e podem ser integrados com diferentes sistemas. IODEF pode ser usado em conjunto com o protocolo RID, proporcionando um sistema eficiente e seguro de transporte, enquanto o OpenIOC não tem foco na implementação de mecanismos de transporte.

Como resultado da avaliação, pode-se dizer que STIX é o padrão de-facto no contexto de inteligência de ameaças. Em primeiro lugar, STIX é o mais popular e compatível, sendo suportado por muitas plataformas e ferramentas, e utilizado pela maioria das organizações. Em segundo lugar, devido à sua arquitetura holística e sua capacidade de abordar uma diversidade de cenários no âmbito da inteligência de ameaças, pode ser considerado o padrão mais completo. Mesmo que os outros padrões ainda sejam suportados por algumas plataformas e tenham um desempenho satisfatório, os recursos oferecidos pela STIX têm se destacado.

3.3 RESULTADOS DA AVALIAÇÃO DAS PLATAFORMAS DE CTI

Os resultados relativos à seleção e avaliação das plataformas são apresentados e explicados.

A partir do processo de busca por plataformas de CTI, um número massivo de projetos foram identificados. Em [57] e [40] um número significativo de plataformas foram analisadas, totalizando 30 e 23, respectivamente. Em [38] um conjunto menor de plataformas foi mencionado e considerado consolidado na área. Em estudos mais específicos, [24] [37] e [58], apenas plataformas de código aberto e populares foram avaliadas. Outro trabalho [18] propôs um um *framework* para avaliar plataformas de CTI e descreveu

Tabela 3.5: Avaliação dos padrões de CTI

	STIXv2 [48][49] & TAXII [54]	**IODEFv2 [54] & RID [55]**	**OpenIOC [56]**
Arquitetura Holística			
Ameaça	++++	++++	++++
Incidente	++++	++++	+++
Adversário	++++	++++	++
Defesa	++++	++	+
Processo de Produção de Inteligência			
Formatação comum	++++	++++	++++
Formato estruturado	++++	++++	++++
Tamanho	+++	+++	+++
Legibilidade para máquinas	++++	+++	++++
Modelo não ambíguo	++++	+++	++++
Mecanismos de relacionamento	++++	++	+++
Interoperabilidade	++++	+++	+++
Mecanismo de transporte	++++	++++	+
Aplicação prática	++++	++	+++
Legenda: muito alto (++++) alto (+++) médio (++) baixo (+)			

os resultados para três plataformas. Algumas fontes confiáveis também mencionaram plataformas novas que possuem grande potencial de ascensão [59][60]. Um parte considerável das plataformas encontradas foram excluídas de acordo com o filtro de exclusão que considera a aderência ao processo de produção de inteligência. Desse modo, um total de 16 plataformas foram classificadas em termos de popularidade e os resultados estão apresentados na Tabela 3.6.

Tabela 3.6: Plataformas de CTI descritas por popularidade e modelo de licença

Plataforma	**Popularidade**	**Modelo de Licença**	**Referências**
Accenture CIP	+	Código fechado	[40][57]
Anomali STAXX	+++	Código fechado com versão gratuita	[40][38][57]
MISP	++++	*Open Source* (GNU *General Public License*)	[17][18][40][37][38][24][57]
CRITs	+++	*Open Source* (GNU *General Public License*)	[40][37][24]
OpenCTI	+++	*Open Source* (Apache License)	[21][59][60]
Facebook TE (beta)	++	*Open Source* (BSD License)	[40][38]
Falcon Intelligence	++	Código fechado	[40]
MANTIS	++	*Open Source* (GNU *General Public License*)	[40][37]
McAfee TIE	+	Código fechado	[40][57]
Microsoft Interflow	+	Código fechado	[40][57]
Soltra Edge	+++	Código fechado	[40][38][37][57]
ThreatQ	++	Código fechado	[18][40][38][57]
ThreatConnect	++	Código fechado	[40][38][57]
EcleticIQ	+	Código fechado	[40][38][57]
IBM X-Force	++	Código fechado	[40][38][57]
CIF	+++	*Open Source* (GNU *General Public License*)	[17][40][37][24]
Legenda: muito alto (++++) alto (+++) médio (++) baixo (+)			

Considerando que o escopo deste trabalho se restringe a plataformas de código aberto ou gratuitas, cerca de metade das plataformas foram excluídas. Em seguida, o critério de popularidade foi aplicado para selecionar as plataformas.

3.3.1 Plataformas Selecionados para Avaliação

Diante dos resultados apresentados, as plataformas selecionadas para posterior análise e avaliação são: **MISP, CIF, CRITs, OpenCTI e Anomali STAXX**. é fornecida uma apresentação sucinta das plataformas selecionadas.

3.3.1.1 Malware Information Sharing Platform

MISP é uma plataforma de código aberto de CTI que permite o compartilhamento, armazenamento e a correlação de indicadores de comprometimento (IOCs) [61]. A ferramenta disponibiliza um banco de dados de indicadores, incluindo informações técnicas e gerais sobre ameaças cibernéticas, que são armazenados em formato estruturado e com um modelo de dados flexível. Os dados armazenados são automaticamente correlacionados para descrever as relações entre eventos, atributos e indicadores.

3.3.1.2 Collaborative Research into Threats

Collaborative Research into Threats (CRITs) é uma ferramenta de código aberto que integra um repositório de malwares e ameaças cibernéticas com outro software capaz de oferecer mecanismos de análise e correlação dessas informações [62]. Essa iniciativa foi desenvolvida pela MITRE, com o objetivo principal de auxiliar a comunidade de segurança cibernética no processo de análise e compartilhamento de dados sobre ameaças [63].

3.3.1.3 OpenCTI

OpenCTI é uma plataforma de código aberto que possui como objetivo principal agregar, de maneira compreensiva, informações gerais e técnicas do contexto de CTI [6]. Ela auxilia organizações no processo de administração de seu conteúdo sobre ameaças cibernéticas, permitindo a estruturação, armazenamento, organização e visualização gráfica dessas informações. O modelo para estruturação dos dados na plataforma segue a padronização definida pela linguagem STIX v2.

3.3.1.4 Collective Intelligence Framework

CIF é um sistema focado em velocidade, desempenho e integração utilizado no gerenciamento de informações de ameaças [64]. Ele auxilia os usuários na formatação, normalização, processamento, armazenamento, compartilhamento e construção de conjuntos de dados de ameaças. O sistema extrai informações sobre ameaças cibernéticas de uma ampla gama de fontes e cria um agrupamento sequencial e cronológico sobre uma ameaça específica [65].

3.3.1.5 Anomali STAXX

É uma ferramenta que fornece compartilhamento bidirecional entre fontes de inteligência de ameaças que utilizam os padrões STIX e TAXII [66], facilitando o acesso a informações sobre ameaças cibernéticas. A plataforma fornece um painel intuitivo para apresentar dados obtidos de diferentes fontes.

3.3.2 Avaliação das Plataformas

Posteriormente à definição das plataformas, foi feita uma avaliação com base na literatura acadêmica, estudo de documentação oficial e demonstrações práticas. De acordo com os critérios supracitados, a avaliação das plataformas foi feita e está resumida na Tabela 3.7

Na perspectiva da arquitetura dos dados, MISP e OpenCTI são aquelas com a abordagem mais holística e aplicáveis em diversos cenários. Além disso, quando utilizadas de forma correta, ambas as plataformas têm a capacidade de abordar de forma eficiente o método 5W3H e fornecer suporte ao processo de tomada de decisão. As demais plataformas apresentam alguns pontos de falha quanto à representação de entidades incorporadas ao espectro das ameaças cibernéticas. Além disso, as plataformas não estão focadas em mecanismos de classificação e correlação. Como resultado, nem todos os aspectos do método 5W3H são suportados.

Do ponto de vista do fluxo de produção de inteligência, a avaliação forneceu algumas descobertas significativas. Primeiro, todas as plataformas suportam importação e exportação de dados em pelo menos um dos formatos mais comuns, como XML, CSV e JSON. Além disso, com exceção do CIF, as plataformas são compatíveis com o STIX, considerado o padrão consolidado no ecossistema de CTI, juntamente com outros padrões. Nesses quesitos, MISP pode ser considerada a plataforma mais flexível considerando a compatibilidade com diferentes formatos.

Quanto ao processo de coleta, todas as plataformas têm a capacidade de realizar a coleta automática de informações. Para as plataformas CIF e Anomali STAXX, esse recurso é integrado, enquanto para as outras plataformas alguma integração pode ser necessária. Outro ponto importante a analisar são os mecanismos de correlação e classificação, que são bem executados pela MISP e OpenCTI. As outras plataformas não estão focadas no estágio de análise do processo de CTI e fornecem apenas alguns mecanismos de filtragem ou agregação.

Em relação à visualização da informação, exceto CIF que é baseado na linha de comando, todas as plataformas utilizam dashboards para apresentar as informações. MISP, CRITs e Anomali STAXX fornecem um painel genérico para todas as informações na plataforma, enquanto o OpenCTI constrói painéis personalizados para as diferentes informações na plataforma. Ele também fornece gráficos de relacionamento intuitivos e completos com base no padrão STIXv2. MISP e o CRITs também oferecem serviços de visualização de relacionamento que devem ser acoplados por meio de módulos.

Considerando a integração entre plataformas, sistemas e ferramentas, MISP e OpenCTI são os mais adaptáveis. CIF possui alguns códigos de extensão para utilizar na integração com alguns Sistemas de Detecção de Intrusão (IDSs) e as demais plataformas não possuem mecanismos específicos para este tipo de integração. Sobre os critérios de compartilhamento, a Anomali STAXX pode se comunicar com qualquer

plataforma usando TAXII, enquanto MISP, CIF e CRITs se concentram em estabelecer um grupo confiável de instâncias.

Por fim, todas as plataformas possuem documentação disponível e, para MISP, OpenCTI e Anomali STAXX, a documentação muito extensa e elaborada. Para a plataforma CRITs foi encontrada uma quantidade satisfatória de informações e detalhes. Já no caso da plataforma CIF, houveram algumas dificuldades em encontrar e organizar a documentação que é limitada em detalhes e apresenta descrições bastante sucintas.

Como resultado da avaliação, pode-se afirmar que, atualmente, existem no mercado algumas plataformas de CTI satisfatórias. Cada um delas possui diferenciais que podem otimizar o processo de criação de inteligência contra ameaças. Portanto, é importante analisar o contexto em que serão aplicadas e os objetivos que devem ser alcançados para decidir qual plataforma utilizar.

3.4 SÍNTESE DOS RESULTADOS

Para obter uma boa capacidade de detecção e prevenção de ameaças cibernéticas, a maioria das organizações precisa contar com as plataformas de CTI de código aberto disponíveis. Da mesma forma, essas plataformas precisam de padrões consolidados para fornecer um serviço automatizado, compartilhável e confiável. Portanto, é essencial analisar recursos e modos de operação dessas duas vertentes do domínio de inteligência de ameaças.

Ao avaliar alguns padrões comuns de CTI, é notório que o STIX, combinado com os recursos do TAXII, pode ser considerado o padrão mais holístico e aplicável em diferentes casos de uso. Além disso, as estatísticas mostram que é o padrão mais utilizado entre as organizações [47]. Portanto, um passo importante é a consolidação definitiva deste padrão, para que o objetivo de estabelecer ampla integração e interoperabilidade entre as organizações possa ser alcançado.

Em relação à análise das plataformas de CTI, várias soluções interessantes foram encontradas. Algumas delas focam em fornecer velocidade e desempenho, outras trouxeram grandes esforços na visualização das informações, enquanto algumas implementam um pouco de cada recurso. Na verdade, diversos tipos de sistemas, com objetivos diferentes, são definidos como plataformas de inteligência de ameaça. Isso provavelmente deriva do fato de que atualmente não existe uma definição padronizada para o conceito ou processo de produção de inteligência de ameaças cibernéticas. Como CTI é um domínio muito extenso, seria relevante estabelecer escopos para melhor caracterizar as plataformas disponíveis, tornando mais fácil decidir quais plataformas são mais bem aplicadas em cada caso de uso.

Levando em consideração o fato de que as plataformas de CTI possuem objetivos distintos, pode-se dizer que atualmente não existe uma plataforma completa, com capacidade para atender todos os processos de CTI adotados neste estudo. Assim, uma possibilidade de ampliar e otimizar os resultados obtidos com a aplicação dos processos de CTI seria a integração entre diferentes plataformas de CTI, com objetivos complementares. Adotando essa perspectiva, é possível conciliar diferentes aspectos como desempenho e mecanismos visuais, alcançando um processo de CTI muito mais desenvolvido e que ofereça desde a coleta de dados até a transformação dos dados em inteligência acionável.

Pelos motivos expostos, ainda é necessário realizar estudos no sentido de caracterizar o conceito de CTI de forma mais específica. Não apenas uma definição deve ser estabelecida, mas também os processos que estão envolvidos neste domínio. Assim, a ampla gama de sistemas disponíveis, denominados de plataformas de CTI, poderão ser melhor aproveitados e aplicados e, novos sistemas que serão desenvolvidos, poderão ser melhor desenhados.

Tabela 3.7: Avaliação das plataformas de CTI

	MISP [61]	**OpenCTI [6]**	**CIF [64][65]**	**CRITs [62][63]**	**Anomali STAXX [66]**
Arquitetura Holística					
Casos de uso	++++	++++	+++	+++	+++
Aderência ao método 5W3H	++++	++++	+	++	+
Processo de Produção de Inteligência					
Fomato de importação	OpenIOC, STIX, CybOX, JSON, CSV, XML	STIX, CybOX, JSON, CSV, XML	XML, JSON, Zip	CSV, STIX, CybOX	STIX
Coleta automática	MISP feeds	Utiliza conectores com fontes ou outras plataformas	Sincronização automática com diferentes fontes	Possível integração com ferramentas de coleta	Sincronização automática com feeds configurados
Formato de exportação	MISP, OpenIOC, CSV, XML, JSON	CSV, STIX	CSV, JSON, HTML, XLS	CSV, STIX, CybOX	CSV, JSON
Visualização gráfica	Dashboard geral e intuitivo e gráficos de relacionamento	Dashboards diversos e grafos de relacionamento baseados em STIXv2	Interface de linha de comando com possibilidade de integração com ferramenta visual	Dashboard simples e extensão para gerar gráficos de relacionamento	Dashboard geral
Correlação	Automática para todos os dados na plataforma	Automática para todos os dados na plataforma	Não convém	Necessario módulo de extensão	Não convém
Classificação	Baseada no tipo de indicador	Baseada em objetos STIXv2	Baseada no tipo de indicador	Baseada em modelo de dados proposto	baseada no tipo de indicador, a partir de mecanismo de busca
Integração	IDS, SIEMs e outras plataformas de CTI	outras plataformas de T	IDSs (Snort, Splunk, Bro, Bind)	Não convém	Não convém
Mecanismo de compartilhamento	Grupo de instâncias confiáveis	Instância particular com compartilhamento entre usuários	Grupo de instâncias confiáveis com serviço centralizado	Grupo de instâncias confiáveis	Qualquer sistema com suporte TAXII
Adicionais					
Documentação	Extensiva e bem elaborada	Extensiva e bem elaborada	Poucos detalhes e descrições sucintas	Quantidade e descrição satisfatória	Extensiva e bem elaborada
Modelo de Licença	*Open Source* (GNU *General Public License*)	*Open Source* (Apache License)	*Open Source* (GNU *General Public License*)	*Open Source* (GNU *General Public License*)	Código fechado com versão gratuita
Legenda: muito alto (++++) alto (+++) médio (++) baixo (+)					

4 METODOLOGIA PROPOSTA

Uma vez que o conceito e a implementação de inteligência de ameaças cibernéticas é algo relativamente novo, ainda existem diversas oportunidades de pesquisa relacionadas ao assunto. Uma delas diz respeito a qualidade dos dados gerados pelas soluções de CTI adotados atualmente. Dado que não existe padronização quanto aos processos que devem ser seguidos para a geração de inteligência de ameaça, as informações disponibilizadas são muitas vezes superficiais e descontextualizadas, além de não possuírem uma formatação padrão.

Este capítulo apresenta a metodologia proposta para otimização da qualidade de dados de inteligência de ameaça cibernética. O método compreende a integração de duas plataformas de inteligência de ameaça de código aberto, disponíveis no mercado, com o objetivo de integrar seus pontos fortes de maneira complementar e assim, gerar dados de TI com maior qualidade e eficácia. Além disso, de modo a garantir a qualidade dos dados proposta, estes serão analisados com base no método 5W3H, explicado nas Seção 3.1.2.1, dado que esse método é capaz de mensurar a capacidade da informação produzida em caracterizar e detalhar uma ameaça.

4.1 METODOLOGIA DE INTEGRAÇÃO DAS PLATAFORMAS DE CTI

Atualmente, o cenário de inteligência de ameaça é bastante vasto e pouco padronizado. Isso se dá, principalmente, devido a popularização do tema e consequente surgimento de uma grande quantidade de sistemas, plataformas e formatos que se colocam como soluções de CTI, porém desempenham diferentes funções. Além disso, muitas dessas soluções não compreendem um processo completo de produção de inteligência de ameaças cibernéticas e acabam produzindo uma quantidade massiva de dados e informações de ameaça, muitas vezes ineficazes na prevenção de ameaças.

Sabendo disso, viu-se necessário definir e adotar um fluxo de processos para a produção de inteligência de ameaças cibernéticas, de modo a garantir que a informação final seja relevante ao contexto em que está inserida e eficaz na prevenção de ameaças. O fluxo adotado neste trabalho, discutido na Seção 2.1.3.1, é composto de cinco etapas: coleta, processamento, análise, implantação e disseminação. Dentre esses processos, o de coleta de dados pode ser considerado o mais consolidado atualmente, dado que existe uma diversidade de soluções capazes de realizá-lo com eficiência para as mais diversas fontes de dados. Os processos de implantação e disseminação também possuem problemáticas e desafios a serem superados, porém estão além do escopo deste trabalho. Por outro lado, as etapas de processamento e análise ainda estão sendo aperfeiçoadas e diferentes abordagens são utilizadas nas soluções existentes. Essas etapas são responsáveis pelo tratamento dos dados, contextualização e correlação da informação, além da produção de sua saída para o usuário final. Desse modo, como o objetivo deste trabalho é otimizar a qualidade das informações produzidas, o método proposto tem foco nas etapas de processamento e análise dos dados.

Além do fluxo padrão de processos descritos na Figura 2.3, para o contexto deste trabalho foi considerada também uma etapa de visualização da informação. Isso decorre da necessidade de analisar as

informações geradas para garantir a qualidade, demonstrar de maneira descritiva os dados gerados para o usuário final e representar a possibilidade de implantação e disseminação dessas informações.

Com base no exposto, elaborou-se uma arquitetura para a implementação de uma solução, apresentada na Figura 4.1, focada na etapas de processamento e análise dos dados para a produção de inteligência de qualidade.

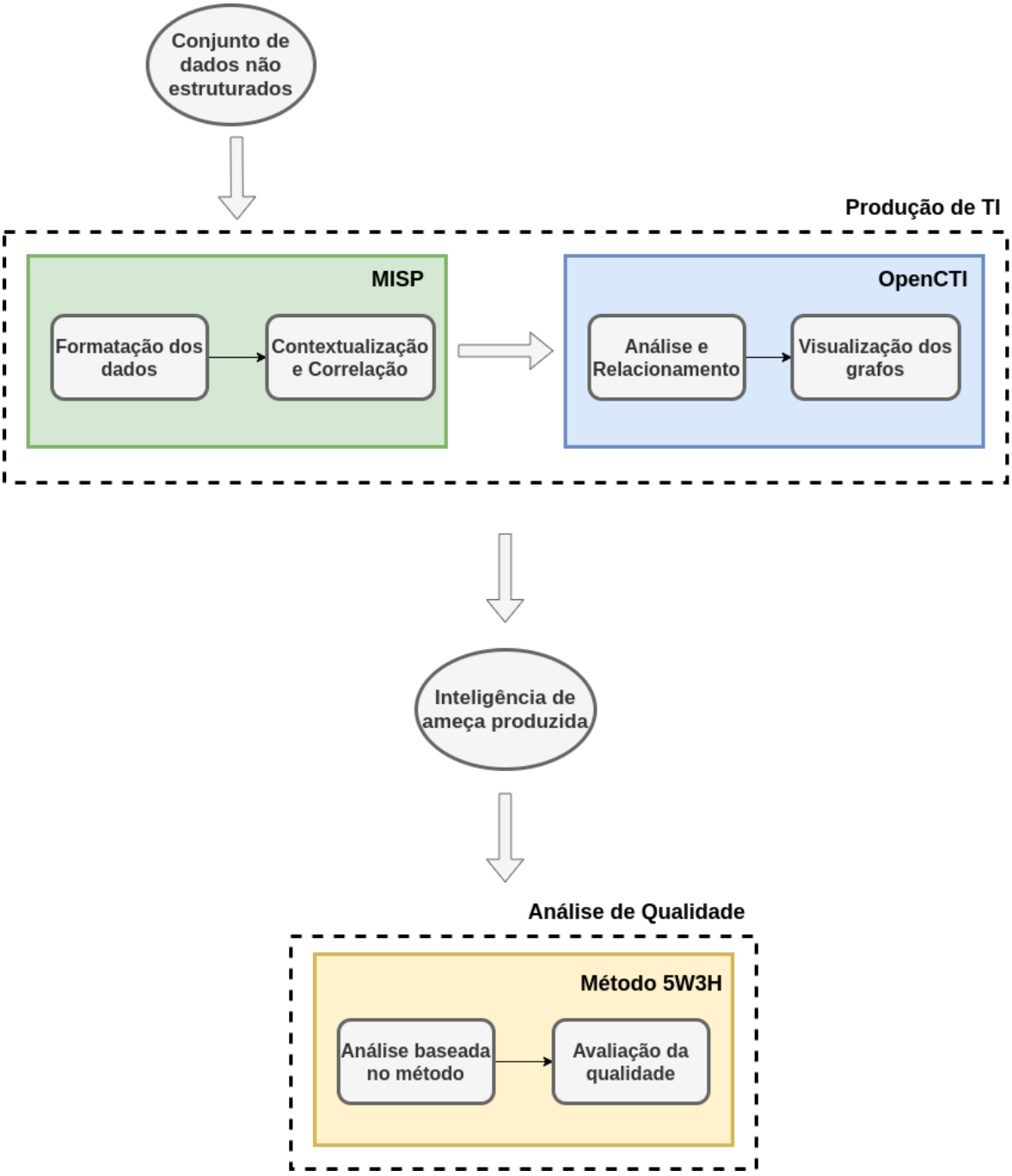


Figura 4.1: Aquitetura da solução proposta

A arquitetura consiste na integração entre duas plataformas de TI, MISP e OpenCTI, para a realização dos estágios de processamento e análise, respectivamente. Em seguida, a inteligência produzida a partir da complementaridade das funcionalidades dessas plataformas estará visualmente disponível por meio de

grafos criados pela plataformas OpenCTI. Por fim, a qualidade da inteligência produzida será avaliada com base no método proposto.

4.1.1 Configurações MISP

Como proposto na arquitetura esboçada para este trabalho, apresentada na Figura 4.1, com base nos resultados obtidos em [22] e apresentados no Capítulo 3, a ferramenta MISP foi selecionada para realizar as etapas de formatação e contextualização dos dados.

Para atingir os resultados esperados, algumas configurações foram fundamentais durante a instalação e primeiras utilizações da plataforma. Para que indicadores não estruturados fossem extraídos e formatados, foi necessária a instalação e ativação de módulos de enriquecimento de dados disponíveis para a plataforma. No contexto dos testes simulados neste trabalho, os módulos utilizados foram: *pdf-enrich* e *xlsx-enrich*. Em seguida, para especificar os dados que deveriam ser exportados para a plataforma OpenCTI, foi criada uma nova *tag*. As *tags* são de suma importância na plataforma MISP, pois é com base nelas que a plataforma consegue relacionar os eventos de forma mais assertiva e criar contextos específicos para os eventos.

4.1.2 Configurações OpenCTI

Como proposto na arquitetura esboçada para este trabalho, apresentada na Figura 4.1, com base nos resultados obtidos em [22] e apresentados no Capítulo 3, a ferramenta OpenCTI foi selecionada para realizar as etapas de análise e relacionamento das informações, além de prover a visualização das relações.

Para que os eventos de interesse fossem exportados da plataforma MISP para a plataforma OpenCTI, foi necessário configurar e ativar dois conectores disponíveis para a plataforma. O primeiro conector, para conexão com a instância da plataforma MISP, requer alguns parâmetros específicos de configuração para funcionar de maneira adequada. Desse modo, foi necessário definir *tags* de importação, uma organização específica responsável pela criação dos eventos e habilitar a criação de relatórios, observáveis e indicadores a partir dos eventos importados. Além disso, o conector permite a configuração de um tempo, em minutos, para atualização automática das informações compartilhadas entre as plataformas e importação de novos eventos. Já o segundo conector, para conexão com *framework* Mitre ATT&CK, não exige configurações adicionais dado que a ferramenta disponibiliza uma base de conhecimento gratuita e globalmente acessível.

4.2 PRODUÇÃO DE INTELIGÊNCIA DE AMEAÇA

Considerando o processo de evolução de um dado até sua transformação em inteligência, apresentado na Figura 2.2, para garantir a produção de inteligência relevante, é necessário focar nas etapas de processamento e análise, de modo a estruturar, contextualizar e criar relações para os dados de entrada. Além disso, no contexto de produção de inteligência de ameaças cibernéticas, a visualização e extração dos dados de forma intuitiva é fundamental para garantir a inteligência seja oportuna e clara.

4.2.1 Processamento

A etapa de processamento consiste primeiramente, na formatação do dados coletados para um formato padronizado, seguido da combinação desses dados formatados com o objetivo de responder perguntas específicas que auxiliem na produção de informação. Vale ressaltar que a informação produzida deve estar padronizada e estruturada para que seja considerada clara, contextualizada com o cenário no qual está inserida para que seja considerada relevante e produzida rapidamente para que seja oportuna. Desse modo, o processamento de dados exige uma definição de um formato para estruturação do conjunto de dados que será trabalhado e um mecanismo eficiente para processamento e combinação dos dados de modo a inserir contexto e correlacionar os dados em um período de tempo oportuno.

4.2.1.1 Formato dos dados

Com base em uma revisão bibliográfica acerca do tema, constatou-se que STIX e TAXII são considerados os padrões mais promissores no contexto de inteligência de ameaças dado que vêm se destacando nos últimos tempos. Uma grande vantagem desses formatos é o fato se serem os mais populares e, consequentemente, serem compatíveis com a grande maioria das soluções de TI. Além disso, sua arquitetura holística garante sua aplicação em diversos cenários no escopo de CTI.

Considerando o exposto, para a metodologia deste trabalho optou-se pela utilização de conjuntos de dados formatados em STIX/TAXII tanto para o processamento dos dados como para a visualização final da inteligência gerada.

4.2.1.2 Método de processamento

Após a definição do padrão de formatação dos dados, foi necessário estabelecer um método de processamento para os dados estruturados. De acordo com o comparativo descrito pelo trabalho [22], a plataforma MISP foi escolhida para realização dessa etapa.

Primeiramente, por ser uma plataforma focada em coleta, armazenamento e compartilhamento de dados de ameaça, sua performance é bastante eficiente para tratar de quantidades massivas de dados. Seu modelo de dados extremamente flexível e de uso intuitivo permite o armazenamento de dados técnicos e não técnicos, ampliando a gama de dados e metadados que podem ser incluídos na etapa de processamento. Além disso, estão disponíveis diversos módulos aplicáveis a plataforma que auxiliam na importação, exportação e enriquecimento dos dados inseridos. Outra funcionalidade interessante é que a simples inserção de dados na plataforma já habilita mecanismos de correlação automática desses dados e garantem a geração de informação contextualizada. Por fim, a plataforma possui compatibilidade com os padrões STIX e TAXII e possibilidade de integração com outras plataformas de TI.

4.2.2 Análise da informação

O processo de análise de um conjunto de dados estruturado consiste na avaliação desses dados para descobrir padrões e produzir inteligência acionável. Esse processo é de extrema importância pois a in-

formação por si só, mesmo que contextualizada, não é capaz de auxiliar proativamente na prevenção de ameaças.

4.2.2.1 Mecanismo de análise

Para a análise dos dados, foi adotada a plataforma OpenCTI. A escolha foi baseada principalmente no fato de que o objetivo da plataforma é correlacionar dados para obter descrições detalhadas sobre ameaças e identificar relações entre elas. Diferente de outras plataformas, nas quais o foco está nos processos de coleta e processamento de quantidades massivas de dados, a plataforma OpenCTI possui melhor desempenho quando utilizada para analisar conjuntos de dados menores e já estruturados com o objetivo de produzir inteligência de ameaça de qualidade e com visualização intuitiva ao usuário final. O modelo de dados utilizado pela plataforma é baseado no padrão STIXv2 o que garante uma representação holística e legível de ameaças cibernéticas, tanto para humanos como para máquinas. Além disso, os dados são modelados de acordo com a teoria de hipergrafos, permitindo um grande crescimento do número relações para uma dada ameaça. A arquitetura de entidades e relações incluídas no modelo de dados da plataforma garante a correlação de informações de maneira completa, além de disponibilizar a visualização em forma de grafos intuitivos que facilitam o processo de identificação de padrões e possíveis ameaças.

4.2.2.2 Correlação com o *framework* MITRE ATT&CK

Considerando a necessidade de construir um banco de informações dentro da ferramenta OpenCTI para garantir a correlação das informações inseridas, viu-se necessário definir uma fonte de dados confiável e abrangente para realizar a população dessas informações.

MITRE ATT&CK é uma base de informações relacionadas ao contexto de ameaças cibernéticas, que inclui uma ampla gama de técnicas e táticas utilizadas para fins maliciosos [67]. Essa base de dados gratuita se tornou conhecida no contexto de inteligência de ameaça e é utilizada como apoio para o desenvolvimento de metodologias de defesa.

Com base em observações do mundo real, o *framework* MITRE ATT&CK desenvolveu matrizes que sintetizam comportamentos conhecidos de agressores em técnicas e táticas. Por ser uma lista abrangente, ela é bastante útil para executar análises ofensivas e defensivas e, por esse motivo, diversas ferramentas de TI desenvolveram métodos de integração com essa base de conhecimento. Desse modo, escolheu-se a utilização do conector MITRE ATT&CK como fonte de informação para povoar a ferramenta OpenCTI.

4.2.3 Visualização da inteligência produzida

Considerando que a metodologia deste trabalho engloba a análise de qualidade da inteligência gerada pela solução proposta, viu-se necessário considerar uma forma de visualização de dados que permitisse o fácil entendimento e avaliação destes. Dado que a ferramenta OpenCTI, utilizada para desempenhar a análise dos dados, possui como um de seus objetivos e pontos fortes a visualização de informações de TI, optou-se pela sua utilização na etapa de visualização dos dados.

4.2.3.1 Modelo de apresentação

Como mencionado na seção 4.2.2.1, a ferramenta OpenCTI, possui um modelo de dados baseado no padrão STIXv2. Sua visualização é baseada em grafos intuitivos que utilizam os objetos do padrão STIX para representar as entidades e uma diversidade de tipos de relação para criar o esquema de apresentação dos dados. Uma exemplificação desse modelo pode ser vista na Figura 4.2.

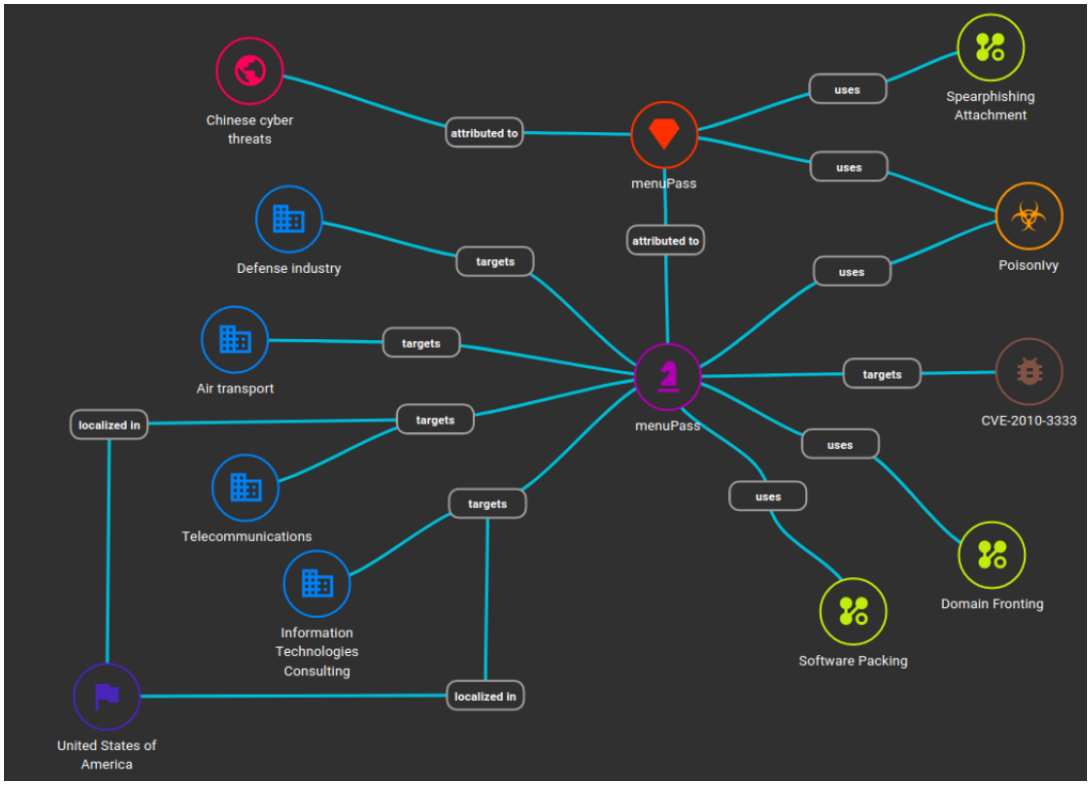


Figura 4.2: Modelo de apresentação de informação plataforma OpenCTI [6]

A partir de informações conjuntas ou de um relatório adicionado na plataforma, de modo automático, as informações inseridas são classificadas e aquelas caracterizadas como uma equivalência à objetos definidos pelo padrão STIXv2 são transformadas em ícones facilmente compreensíveis. Caso já existam na plataforma outras informações sobre algum dos ícones convertidos, é possível clicar nesse ícone e navegar pelas informações relacionadas disponíveis na plataforma. Em seguida, esses ícones são relacionados por meio de linhas de conexão com caixas de texto informando o tipo de relação entre os objetos.

4.3 ANÁLISE DA QUALIDADE

No contexto deste trabalho, considera-se que inteligência de ameaça de qualidade é aquela relevante, oportuna e clara. Para que a inteligência produzida seja capaz de atender a esses requisitos, ela deve descrever uma ameaça, no maior nível de completude possível, de modo que seja possível entendê-la, identificar padrões e inferir novas informações, facilitando a tomada de decisões em relação a uma determinada ameaça. Desse modo, para mensurar a qualidade dos dados construídos com a abordagem proposta, optou-se

por utilizar o método 5W3H como parâmetro de análise.

4.3.1 Análise baseada no Método 5W3H

Tendo em vista a explicação acerca do método 5W3H e sua aplicabilidade no contexto de inteligência de ameças cibernéticas, detalhada na Seção 3.1.2.1, esse método foi escolhido como parâmetro para análise de qualidade das informações produzidas nos cenários experimentais deste trabalho.

Ao aplicar mecanismos de correlação entre todas as informações levantadas a partir da aplicação do método 5W3H, é muito provável que inteligência acionável tenha sido produzida. Desse modo, caso as informações produzidas pela solução proposta sejam capazes de se enquadrar no método, pode-se afirmar que a inteligência obtida possui qualidade e capacidade de auxiliar na tomada de decisões sobre cursos de ação contra ameaças cibernéticas.

4.3.2 Avaliação da qualidade

Após a análise da informação produzida com base no método 5W3H, visando avaliar de forma quantitativa a qualidade dessa informação, alguns valores, dispostos na Tabela 4.1, foram definidos como parâmetro.

Tabela 4.1: Níveis de avaliação

Número de questionamentos respondidos	**Percentual médio de resposta**	**Classificação**
0 - 1	0% - 25%	Insatisfatória
2 - 4	25% - 50%	Pouco Satisfatória
5 - 6	50% - 75%	Satisfatória
7 - 8	75% - 100%	Muito Satisfatória

Considerando que o método compreende um total de oito questionamentos para entender um tópico em sua completude, foram definidos quatro níveis de avaliação baseados em termos percentuais do número de respostas fornecidas pela informação produzida. Em outras palavras, dado o número de questionamentos que a informação produzida é capaz de responder, ela pode ser classificada como insatisfatória, pouco satisfatória, satisfatória ou muito satisfatória.

Vale ressaltar que devido a natureza do que está sendo avaliado e considerando que a avaliação é baseada em uma análise humana, é aceitável que alguns questionamentos sejam respondidos parcialmente. Nesses casos, os valores são mensurados com base em uma média percentual estimada e classificados de acordo com os níveis definidos na Tabela 4.1.

4.4 CENÁRIOS DE TESTE

Com o objetivo de obter resultados práticos acerca da solução desenhada e, consequentemente, entender melhor sua capacidade de aplicação e validar sua eficácia, foram elaborados dois cenários de testes

para simular a metodologia proposta, utilizando conjuntos de dados diferentes.

No primeiro cenário de testes, o objetivo principal foi analisar a capacidade de processamento de uma grande quantidade de dados, em conjunto com a análise desses dados. Para isso, foram utilizados dois métodos de importação diferentes e uma quantidade de indicadores relativamente grande.

No segunda cenário de testes, o objetivo principal era verificar a capacidade de processamento de dados não estruturados, como em relatório ou texto livre, além de verificar os resultados de uma análise de dados bem contextualizados.

4.4.1 Cenário 1: Lista de indicadores de comprometimento (IoCs)

Esse cenário consiste na aplicação da metodologia proposta utilizando um conjunto de dados de entrada baseado em listas de indicadores comuns. Visando avaliar as etapas de processamento e análise, foram consideradas duas abordagens nesse cenário, uma baseada em quantidade de dados e outra no contexto dos indicadores trabalhados.

4.4.1.1 Indicadores Generalizados

Nessa abordagem, o objetivo principal foi analisar a etapa de processamento e avaliar sua performance quando utilizado um grande número de indicadores para formatação e estruturação. Além disso, pretende-se entender o comportamento do processo de análise para um grande volume de dados descontextualizados. Para isso, foi utilizado como conjunto de dados de entrada uma planilha de indicadores, disponibilizada gratuitamente no site da Anomali [68], com indicadores de comprometimento generalizados que englobava IPs, DNS, hostnames, nome de arquivos e *hashs*. A planilha utilizada estava em formato XLXS e possuía 2741 linhas de indicadores para processamento e análise.

4.4.1.2 Indicadores associados ao *Ransomware WannaCry*

Nessa abordagem, o objetivo principal foi analisar a etapa de análise dos indicadores e verificar a capacidade de converter indicadores gerais contextualizados em inteligência de ameaça. Para isso, foi utilizado como conjunto de dados de entrada uma lista, em texto livre, de indicadores relacionados ao conhecido *Ransomware WannaCry*. Essa lista foi obtida gratuitamente na ferramenta OTX da AlienVault [69] e possuía um total de 38 atributos, incluindo domínios, nome de arquivos e *hashs*.

4.4.2 Cenário 2: Relatório contextualizado

Esse cenário consiste na aplicação da metodologia proposta utilizando um conjunto de dados de entrada baseado em um relatório previamente analisado e contextualizado acerca de um tema. O objetivo principal era analisar a capacidade de conversão de informações já analisadas em inteligência de ameaça.

Nessa abordagem foi utilizado um relatório, disponibilizado gratuitamente pela ANSSI [7], descrevendo especificamente o desenvolvimento das atividades desempenhadas pelo grupo criminoso TA505 nos

últimos anos. O relatório estava em formato PDF e inclui as principais informações referentes ao padrão de ataque do grupo criminoso em questão.

4.5 SÍNTESE DO CAPÍTULO

Neste capítulo, foi apresentada a metodologia integrativa proposta para produção de inteligência de ameaças. Foi explicada a abordagem empregada na integração de duas ferramentas de TI de código aberto, embasada nos resultados explicitados no Capítulo 3, assim como as principais configurações realizadas para o funcionamento adequado da solução proposta.

Aqui foram descritos todos os processos adotados para a produção de inteligência de ameaças e explicado o contexto de utilização de cada uma das plataformas nesses processos. Além disso, foi planejado um procedimento de análise da informação produzida com base no método 5W3H e um método para realizar uma avaliação quantitativa de sua qualidade.

Neste capítulo, também foram definidos e apresentados dois cenários de testes, utilizando conjuntos de dados diferentes, visando o entendimento da capacidade de aplicação da metodologia proposta e validação de sua eficácia.

5 RESULTADOS EXPERIMENTAIS

Primeiramente, para a obtenção dos resultados aqui relatados, é importante citar que os dados utilizados para realização dos testes foram obtidos de fontes abertas gratuitas e selecionados de maneira randômica, visando apenas a ilustração da metodologia proposta.

Pontua-se também que os resultados obtidos estão limitados a um escopo relativamente pequeno de dados e informações de inteligência de ameaças. Como o foco deste trabalho não inclui o processo de coleta de dados, a obtenção automática de dados oriundos de *feeds* gratuitos foi desabilitada em ambas as plataformas, limitando à base de conhecimento aos dados inseridos correlacionados com o *framework* Mitre ATT&CK [67].

Além disso, para produzir inteligência de ameaça, as etapas de processamento e análise dos dados foi implementada utilizando-se a integração entre duas ferramentas *open source* de inteligência de ameaças: MISP [61] e OpenCTI [6]. Essa integração é baseada na adição de *tags* aos eventos MISP, com o objetivo de especificar quais eventos que devem ser importados para a plataforma OpenCTI. Desse modo, a *tag import:opencti* foi criada e adicionada em todos os eventos de interesse, possibilitando que os mesmos fossem importados e convertidos para o modelo de dados da plataforma OpenCTI de maneira automática.

Considerando a abordagem proposta por este trabalho para produzir inteligência de ameaça, a avaliação dos resultados considerará a qualidade da inteligência obtida com base no método 5W3H, explicado na seção 3.1.2.1, em dois cenários diferentes: interpretação de relatórios contextualizados e listas generalizadas de indicadores de comprometimento.

Sendo assim, este capítulo traz, na Seção 5.1, a apresentação dos dados utilizados e dos resultados obtidos no primeiro cenário testes, referente às listas de indicadores de comprometimento, e, na Seção 5.2, as informações do segundo cenário de testes, referente aos relatórios previamente contextualizados. Por fim, na Seção 5.4, evidencia-se os resultados do processo como um todo, considerando os diferentes cenários, e avalia-se a metodologia proposta.

5.1 CENÁRIO 1: LISTA DE INDICADORES DE COMPROMETIMENTO (IOCS)

Neste cenário, o conjunto de dados utilizado como entrada para a produção de inteligência de ameaça é composto por indicadores de comprometimento simples e generalizados, incluindo IPs, DNS, Hashes e nomes de arquivos maliciosos. Esses dados foram obtidos gratuitamente a partir de listas de indicadores e observáveis disponibilizadas por instituições envolvidas no contexto de segurança cibernética. No primeiro momento, foram trabalhados dados relacionados especificamente ao *Ransomware WannaCry*, obtidos a partir da ferramenta OTX da AlienVault [69]. Em seguida, foram utilizados dados de ameaças cibernéticas referentes ao tema COVID-19, disponibilizados pela Anomali [68].

5.1.1 Indicadores generalizados

Para processar esse conjunto de dados, uma planilha com os indicadores foi obtida do site oficial da Anomali [68] e esses indicadores foram adicionados como atributos ao respectivo evento na plataforma MISP.

Com o uso do módulo *xlsx-enrich* disponível para a plataforma, foi realizada uma análise de texto livre de uma planilha de indicadores e as informações consideradas possíveis IOCs foram convertidos, de forma automática, em atributos do evento em questão. Como o conjunto de dados utilizado era uma planilha, os indicadores estavam muito bem definidos e não foram gerados falsos positivos no contexto de importação dos dados. O processo desempenhado pelo módulo *xlsx-enrich* gerou um total de 2741 possíveis atributos, número exato de linhas da planilha, em aproximadamente 5 segundos de processamento.

O evento, após o processamento e população dos dados, juntamente com alguns de seus atributos, estão apresentados nas Figuras 5.1 e 5.2. Na descrição do evento é possível observar a *tag* aplicada e a organização criadora do evento, parâmetros importantes para a exportação dos dados pelo conector de integração entre as plataformas. Além disso, por meio do campo atributos, pode-se confirmar o número de atributos gerados pelo processamento desempenhado. Dentre esses atributos, conforme mostra a Figura 5.2, são observadas diferentes categorias de indicadores, incluindo IPs, domínios e *hashes* do tipo MD5.

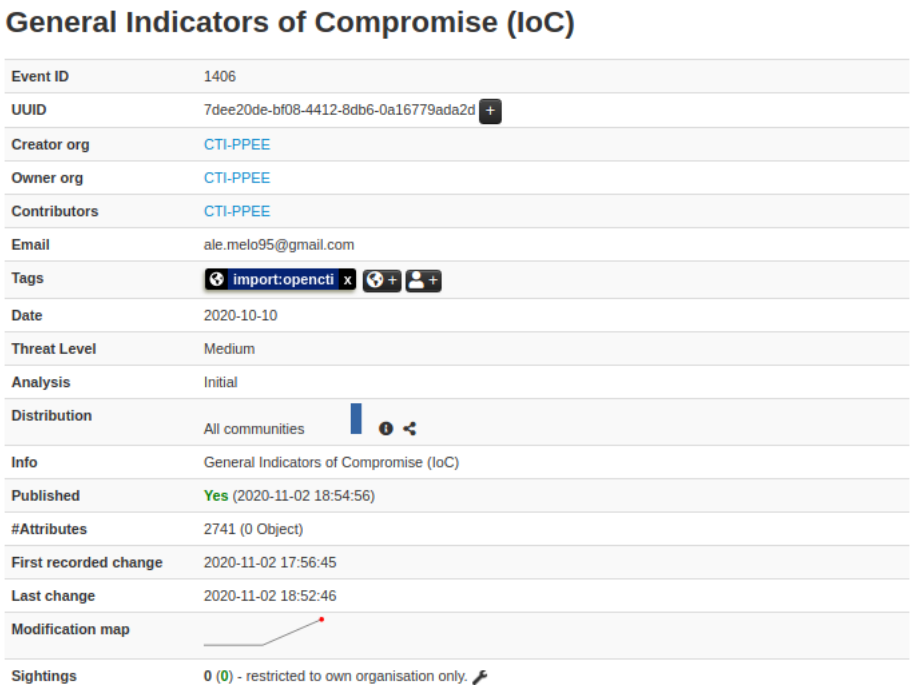
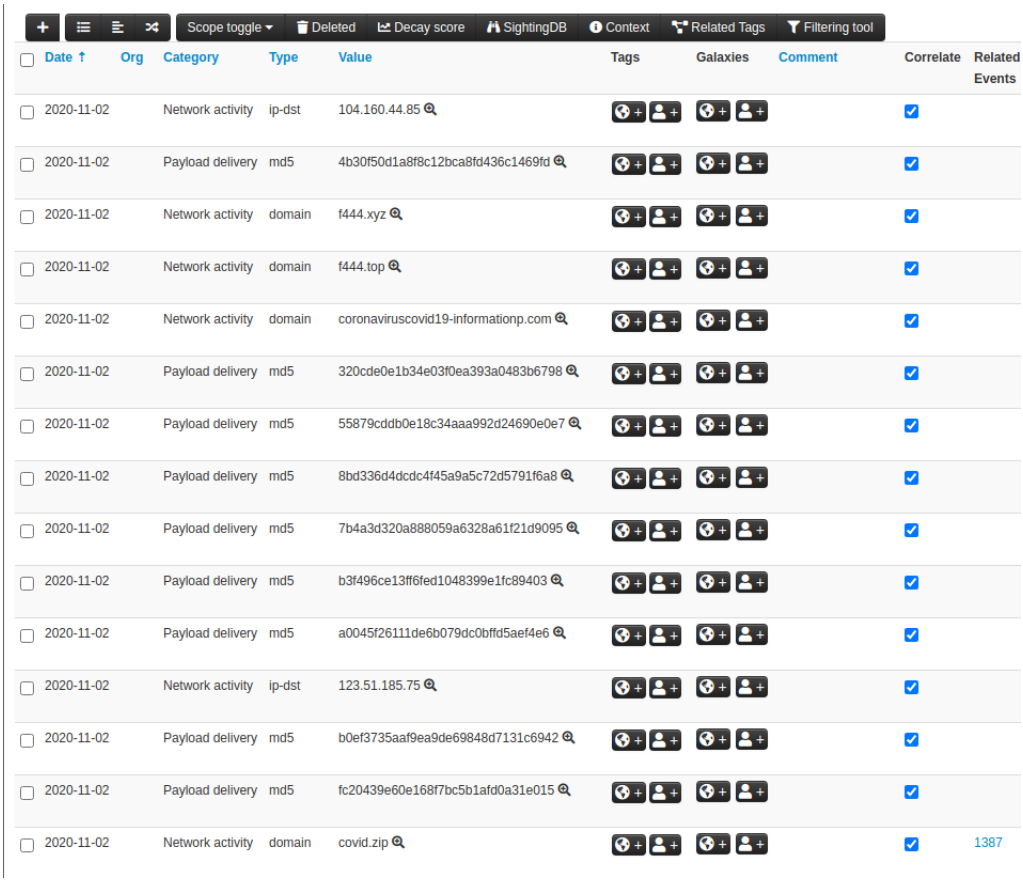


Figura 5.1: Descrição do evento povoado com IoCs generalizados

Em sequência ao processamento dos dados e definição do evento, esse evento foi rotulado e automaticamente importado para a plataforma OpenCTI. Com o objetivo de otimizar a performance na obtenção dos resultados do processo de análise, grande parte dos indicadores foi excluída, mantendo apenas 174 atributos, definidos de forma aleatória.

Para esse modelo de conjunto de dados, a integração não atingiu bons resultados. Por serem indicadores generalizados, apenas relacionados por um tema geral, não existiu conversões de entidades como *threat*



Date ↑	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events
2020-11-02		Network activity	ip-dst	104.160.44.85				✓	
2020-11-02		Payload delivery	md5	4b30f50d1a8f8c12bca8fd436c1469fd				✓	
2020-11-02		Network activity	domain	f444.xyz				✓	
2020-11-02		Network activity	domain	f444.top				✓	
2020-11-02		Network activity	domain	coronaviruscovid19-informationp.com				✓	
2020-11-02		Payload delivery	md5	320cde0e1b34e03f0ea393a0483b6798				✓	
2020-11-02		Payload delivery	md5	55879cddb0e18c34aaa992d24690e0e7				✓	
2020-11-02		Payload delivery	md5	8bd336d4dcdc4f45a9a5c72d5791f6a8				✓	
2020-11-02		Payload delivery	md5	7b4a3d320a888059a6328a61f21d9095				✓	
2020-11-02		Payload delivery	md5	b3f496ce13ff6fed1048399e1fc89403				✓	
2020-11-02		Payload delivery	md5	a0045f26111de6b079dc0bffd5aef4e6				✓	
2020-11-02		Network activity	ip-dst	123.51.185.75				✓	
2020-11-02		Payload delivery	md5	b0ef3735aaf9ea9de69848d7131c6942				✓	
2020-11-02		Payload delivery	md5	fc20439e60e168f7bc5b1afd0a31e015				✓	
2020-11-02		Network activity	domain	covid.zip				✓	1387

Figura 5.2: Parte dos atributos derivados de IoCs generalizados

actor, intrusion set ou *attack pattern*, capazes de agregar informação para converter em inteligência. Desse modo, os dados foram importados para a plataforma OpenCTI apenas como atributos, sem nenhum tipo de relacionamento, mantendo o nível de descrição e completude das informações exatamente igual ao vivenciado com a plataforma MISP. As Figuras 5.3 e 5.4 mostram uma exemplificação da disposição das informações após a importação entre as duas plataformas.

Quando analisadas pelo método 5W3H, pode-se dizer que as informações disponibilizadas são incapazes de responder os questionamentos *what, when, where, who, why how, how long e how much* e por isso devem ser classificadas apenas como informação e não inteligência de ameaça.

5.1.2 Indicadores associados ao *Ransomware WannaCry*

Para processar esse conjunto de dados, os indicadores foram baixados da plataforma OTX [69] e adicionados como atributos ao respectivo evento na plataforma MISP. Esses atributos foram populados por meio do formato de texto livre de importação. Esse formato permite que se digitem textos ou listas de indicadores e os atributos são inferidos automaticamente.

Como o conjunto de dados utilizado era uma planilha em formato .csv, os indicadores estavam muito bem definidos e não foram gerados falsos positivos no contexto de importação dos dados. O evento, após o processamento e população dos dados, juntamente com seus atributos, estão apresentados nas Figuras

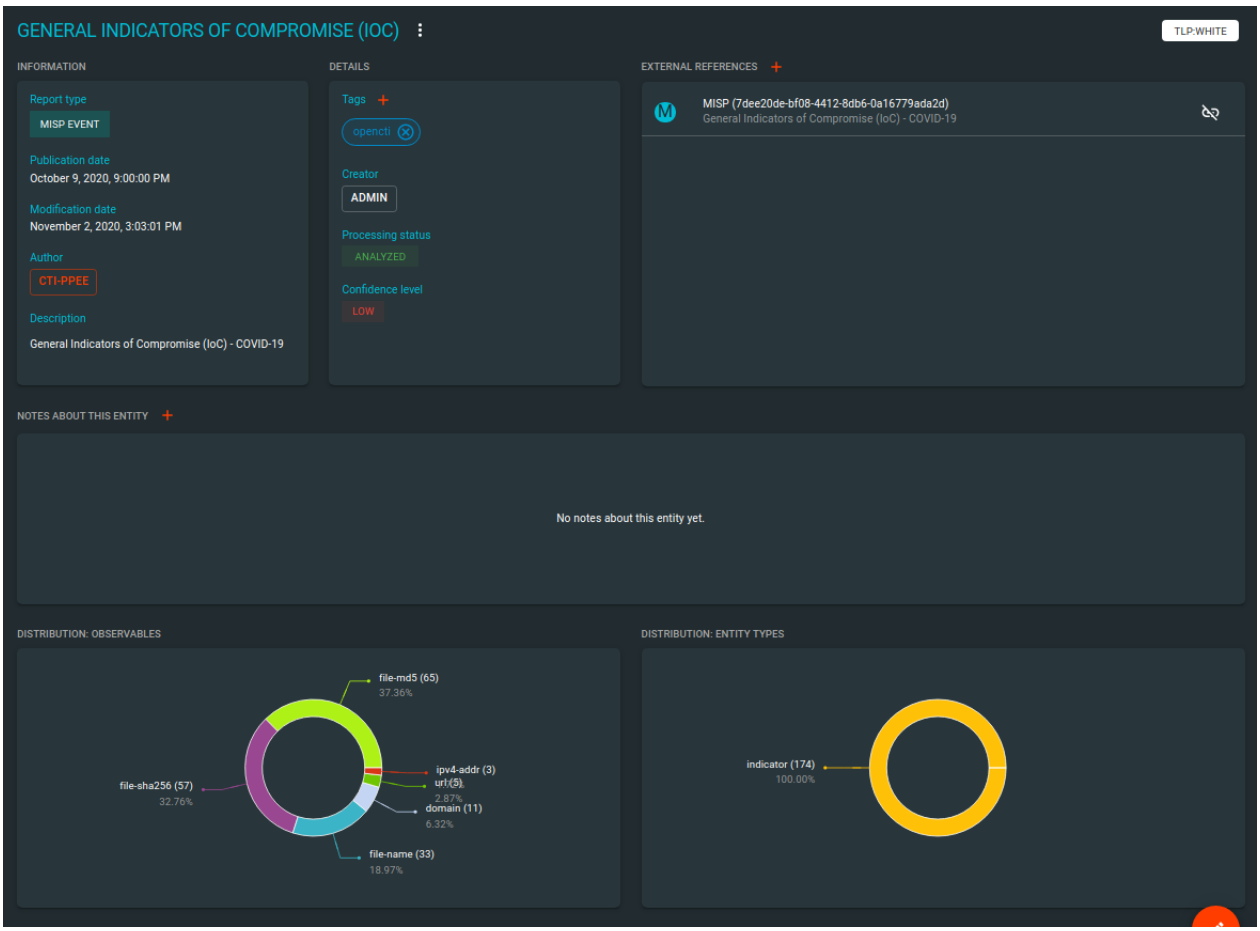
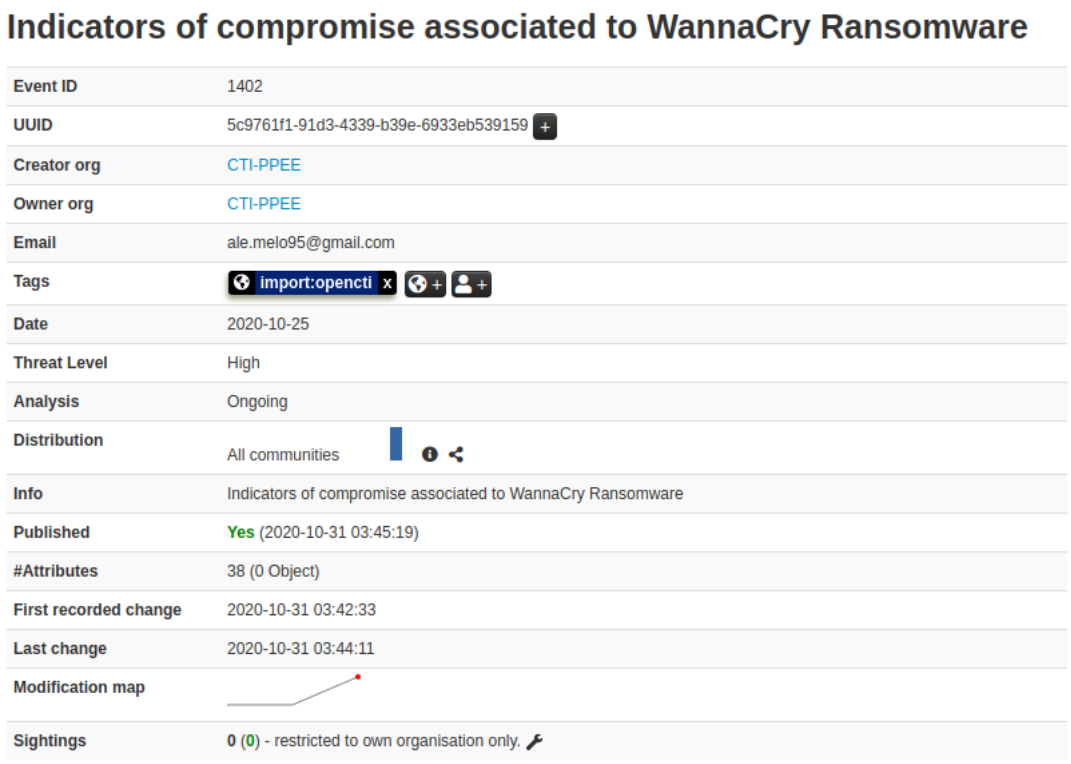


Figura 5.3: Estudo analítico do evento povoado com IoCs generalizados

Figura 5.4: Grafo de relacionamentos do evento populado com IoCs generalizados

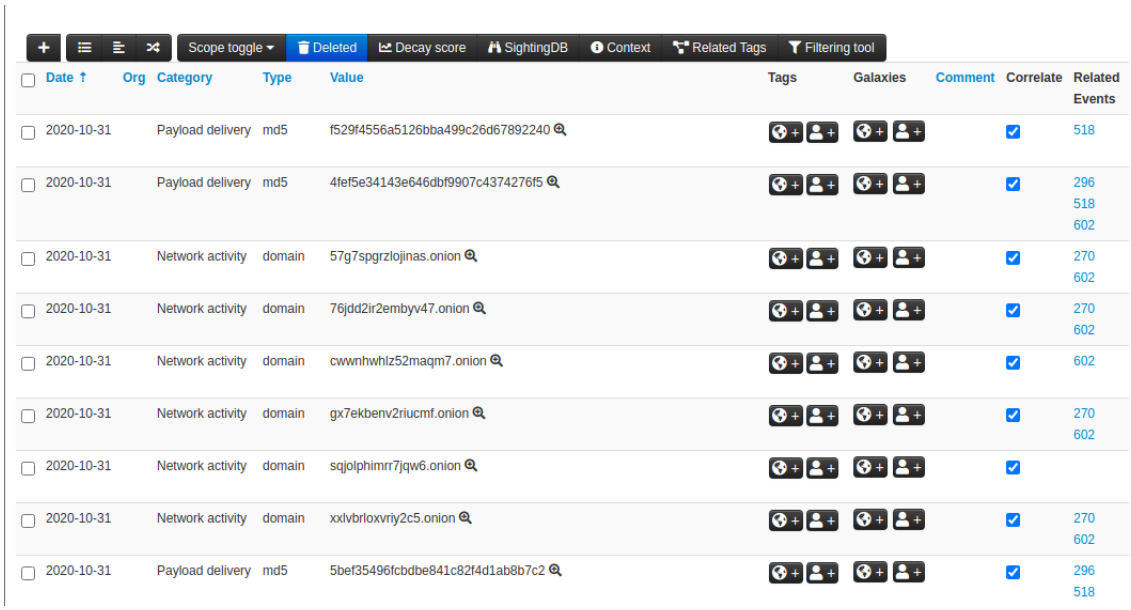
5.5 e 5.6. Na descrição do evento é possível observar a *tag* aplicada e a organização criadora do evento, parâmetros importantes para a exportação dos dados pelo conector de integração entre as plataformas. Além disso, por meio do campo atributos, pode-se confirmar um total de 38 atributos gerados pelo processamento desempenhado. Dentre esses atributos, conforme mostra a Figura 5.6, são observadas duas categorias diferentes de indicadores, incluindo domínios e *hashes* do tipo MD5.



Indicators of compromise associated to WannaCry Ransomware

Event ID	1402
UUID	5c9761f1-91d3-4339-b39e-6933eb539159
Creator org	CTI-PPEE
Owner org	CTI-PPEE
Email	ale.melo95@gmail.com
Tags	import:opencti x
Date	2020-10-25
Threat Level	High
Analysis	Ongoing
Distribution	All communities
Info	Indicators of compromise associated to WannaCry Ransomware
Published	Yes (2020-10-31 03:45:19)
#Attributes	38 (0 Object)
First recorded change	2020-10-31 03:42:33
Last change	2020-10-31 03:44:11
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Figura 5.5: Descrição do evento referente ao *Ransomware WannaCry*



Scope toggle | Deleted | Decay score | SightingDB | Context | Related Tags | Filtering tool

Date	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events
2020-10-31		Payload delivery	md5	f529f4556a5126bba499c26d67892240				✓	518
2020-10-31		Payload delivery	md5	4fef5e34143e646dbf9907c4374276f5				✓	296 518 602
2020-10-31		Network activity	domain	57g7spgrzlojinas.onion				✓	270 602
2020-10-31		Network activity	domain	76jdd2ir2embyv47.onion				✓	270 602
2020-10-31		Network activity	domain	cwwnhwhlz52maqm7.onion				✓	602
2020-10-31		Network activity	domain	gx7ekbenv2riucmf.onion				✓	270 602
2020-10-31		Network activity	domain	sqjolphimrr7jqw6.onion				✓	
2020-10-31		Network activity	domain	xxlvbrloxvriy2c5.onion				✓	270 602
2020-10-31		Payload delivery	md5	5bef35496fcbdbe841c82f4d1ab8b7c2				✓	296 518

Figura 5.6: Atributos associados ao *Ransomware WannaCry* adicionados com o método *free-text import*

Em sequência ao processamento dos dados e definição do evento, esse evento foi rotulado e automaticamente importado para a plataforma OpenCTI. Assim como esperado para a integração, o evento foi analisado e suas principais características apresentadas de maneira intuitiva, ilustrada na Figura 5.7. Além disso, os atributos foram correlacionados e representados por meio de um grafo, conforme mostra a Figura 5.8. Neste grafo, o *Ransomware WannaCry* foi convertido como uma entidade de *Malware* e seus atributos

relacionados a ele por meio da classificação de relação *indicates*.

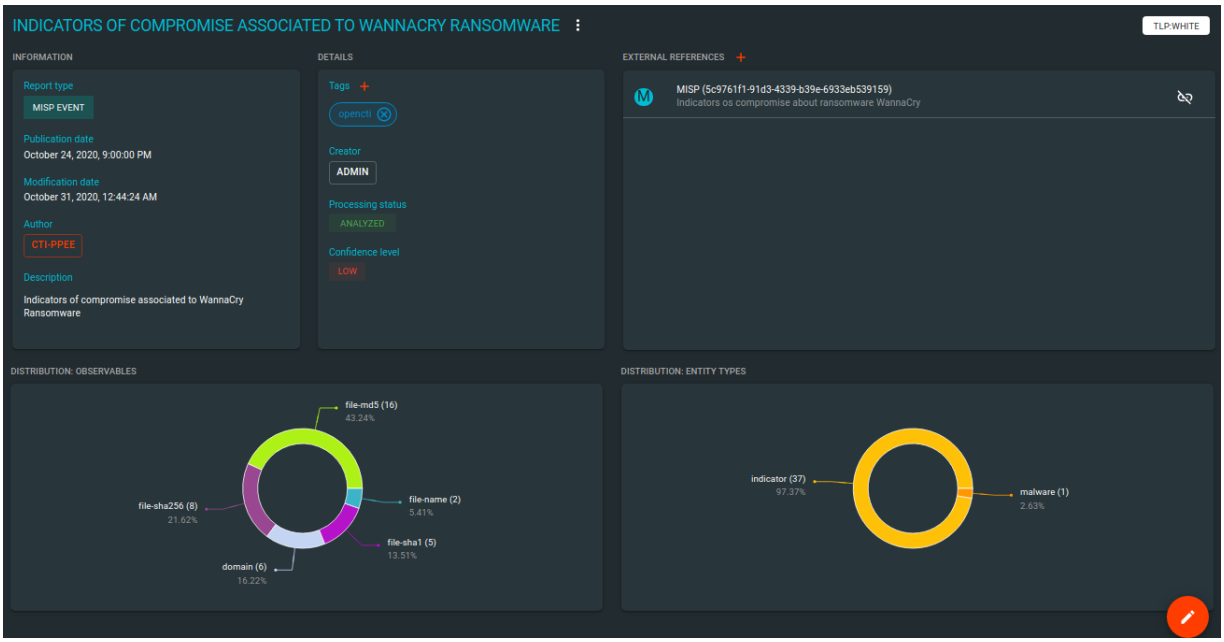


Figura 5.7: Estudo analítico do evento relacionado ao *Ransomware WannaCry*

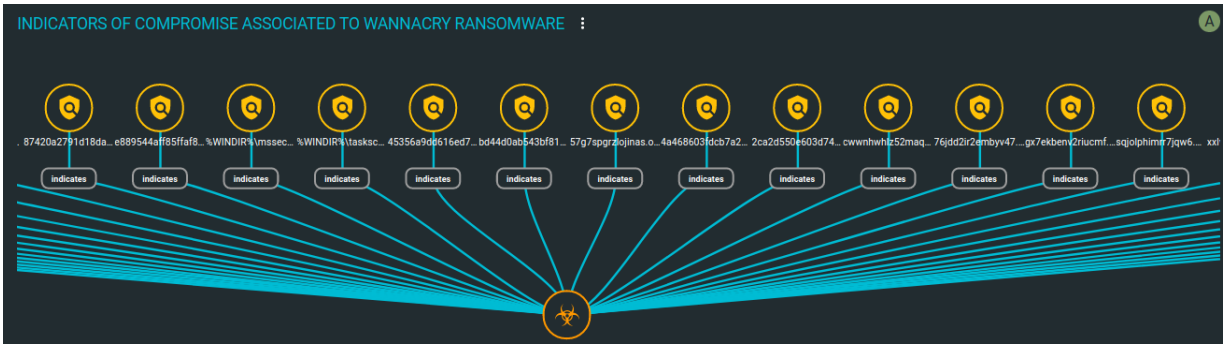


Figura 5.8: Grafo de relacionamentos do evento relacionado ao *Ransomware WannaCry*

Para agregar mais conhecimento ao contexto, é possível navegar pela plataforma e obter novas informações acerca das entidades convertidas. No escopo deste trabalho, as informações adicionais foram obtidas principalmente a partir do *framework* MITRE ATT&CK. Para o *Ransomware WannaCry*, a síntese de conhecimento descreveu 19 relatórios e 41 indicadores, conforme apresentado na Figura 5.9. Dentre essas informações foram identificadas variadas TTPs, uma forma de uso e informações temporais.

5.1.3 Análise dos resultados

Para um conjunto de dados composto por indicadores generalizados, verificou-se que as conversões entre indicadores e entidades não funcionam adequadamente e, por esse motivo, relacionamentos não são construídos. Desse modo, as informações disponibilizadas são incapazes de responder os questionamentos levantados pelo método 5W3H. Assim, para esses dados de entrada, considerando os valores propostos na Tabela 4.1, os processos de processamento e análise resultaram em uma insatisfatória especificidade e completude para os dados trabalhados.

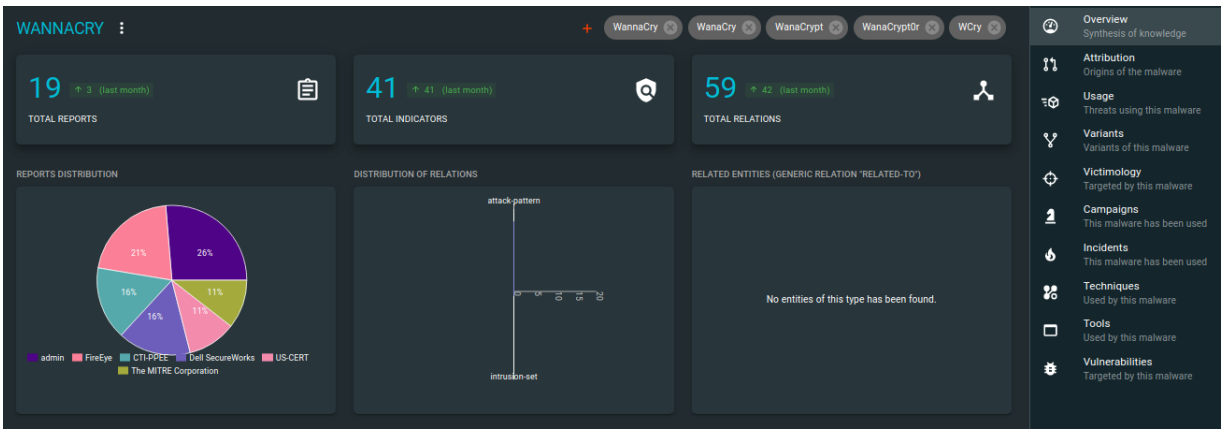


Figura 5.9: Estudo analítico do *Ransomware WannaCry*

Por outro lado, indicadores simples, porém previamente contextualizados, atingiram bons resultados. Ao analisar a informação apresentada na Seção 5.1.2 pelo método 5W3H, pode-se garantir a resposta dos questionamentos *what, when, how e how much* e consegue-se entender parcialmente os questionamentos *where e who*. *What* se refere a entidade em que foi convertido, no caso um *malware*. *When* contempla as informações de primeira e última aparição, além do histórico de modificação. *How* é bem definido a partir da análise das diversas TTPs descritas. *How much* pode ser entendido com as informações retratadas na classe de impacto. *Where* pode ser genericamente descrito considerando o número de países afetados apresentados na síntese. *Who* pode ser inferido quando analisado em conjunto com os *intrusion sets* identificados. Desse modo, a conexão entre os dados analisados, importados da ferramenta MISP, em conjunto com a síntese de conhecimento baseada dos dados oriundos da plataforma MITRE ATT&CK, quando avaliada considerando os valores propostos na Tabela 4.1, resultou em uma satisfatória especificidade e completude de informações.

5.2 CENÁRIO 2: RELATÓRIO CONTEXTUALIZADO

Neste cenário, o conjunto de dados utilizado como entrada para a produção de inteligência de ameaça foi obtido a partir de relatórios previamente estudados e contextualizados acerca de um tema. Esses relatórios foram disponibilizados de forma gratuita por instituições envolvidas no contexto de segurança cibernética.

O relatório aplicado neste cenário de teste foi desenvolvido pela ANSSI e faz um descritivo geral das atividades cibernético maliciosas desempenhadas pelo grupo criminoso denominado TA505 nos últimos anos [7].

Para realizar o processamento dos dados e, consequentemente, a extração de indicadores de comprometimento do relatório, este foi anexado a um evento previamente criado na plataforma MISP, especificamente para essa finalidade. Com o uso do módulo *pdf-enrich* disponível para a plataforma, foi realizada uma análise de texto livre do relatório e as informações consideradas possíveis IOCs foram convertidas, de forma automática, em atributos do evento em questão.

O processo desempenhado pelo módulo *pdf-enrich* gerou um total de 29 possíveis atributos ao evento. Dentro desses atributos, 21 foram identificados como falsos positivos. Isso acontece pois o mecanismo de processamento utilizado pelo módulo *pdf-enrich* é baseado apenas em análise de texto livre fazendo com que algumas URLs e emails oficiais presentes nas referências do relatório, exemplificados na Figura 5.10, fossem identificados como possíveis indicadores de comprometimento.

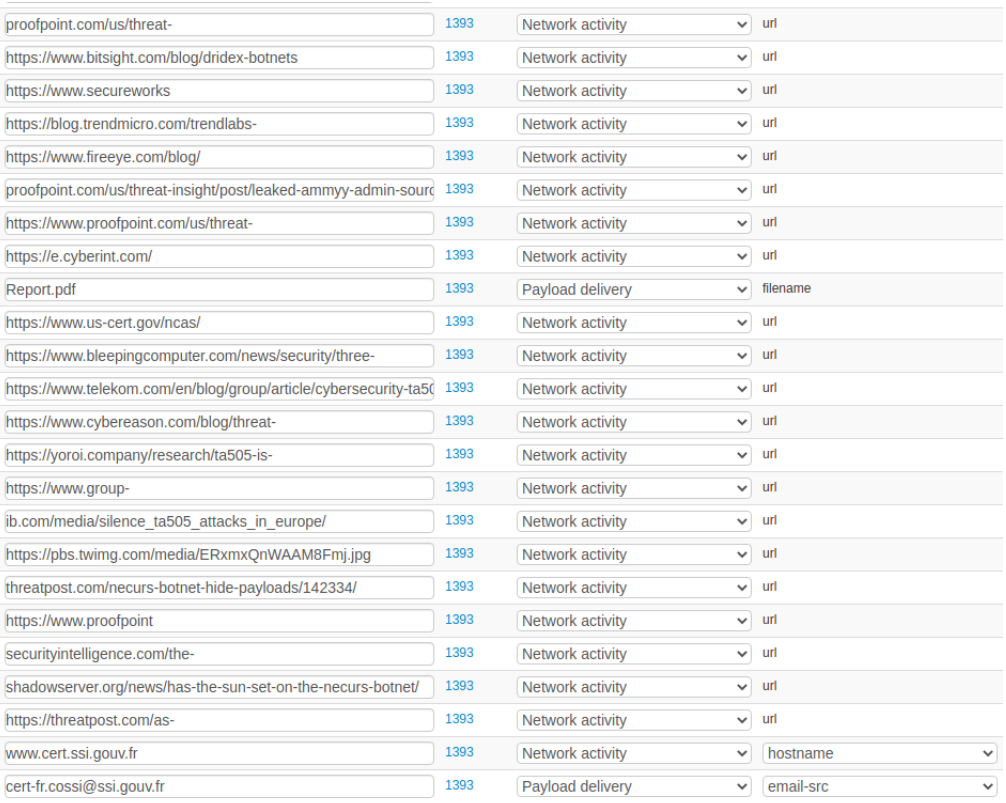


Figura 5.10: Falsos positivos gerados a partir do relatório [7]

Para que não interferissem nos processos posteriores de análise e visualização, os falsos positivos obtidos foram excluídos manualmente dos atributos. O evento, após o processamento e enriquecimentos dos dados, juntamente com seus atributos, estão apresentados nas Figuras 5.11 e 5.12. Na descrição do evento é possível observar a *tag* aplicada e a organização criadora do evento, parâmetros importantes para a exportação dos dados pelo conector de integração entre as plataformas. Além disso, por meio do campo atributos, pode-se confirmar um total de 8 atributos remanescentes para o evento após a exclusão dos falsos positivos.gerados pelo processamento desempenhado. Dentre esses atributos, conforme mostra a Figura 5.12, são observadas duas categorias diferentes de indicadores, incluindo domínios e nome de arquivos, sendo que dois deles são acompanhados por *tags* de *malwares*, fato que auxilia na conversão dos indicadores em entidades na plataforma OpenCTI.

Em seguida, o evento foi rotulado e automaticamente importado para a plataforma OpenCTI. Assim como esperado para a integração, o evento foi analisado e suas principais características apresentadas de maneira intuitiva, ilustradas na Figura 5.13.

Assim como esperado para a integração, os atributos foram correlacionados e representados por meio de um grafo, conforme mostra a Figura 5.14. Neste grafo, o grupo criminoso TA505 foi convertido na

Relevant Threat Actor of Malicious Activity - TA505

Event ID	1404
UUID	4a5de62a-1c76-40bc-8e37-bbcc61a14577
Creator org	CTI-PPEE
Owner org	CTI-PPEE
Email	ale.melo95@gmail.com
Tags	import:opencti x
Date	2020-10-25
Threat Level	High
Analysis	Ongoing
Distribution	All communities
Info	Relevant Threat Actor of Malicious Activity - TA505
Published	Yes (2020-11-02 16:21:26)
#Attributes	8 (0 Object)
First recorded change	2020-11-02 16:14:26
Last change	2020-11-02 16:20:48
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Figura 5.11: Descrição do evento que descreve o relatório [7]

Scope toggle | Deleted | Decay score | SightingDB | Context | Related Tags | Filtering tool

	Date ↑	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events
☐	2020-11-02		Payload delivery	filename	TrickyBot	trickbot x		PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	
☐	2020-11-02		Network activity	domain	iplogger.org			PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	1393
☐	2020-11-02		Payload delivery	filename	sage.exe			PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	1393
☐	2020-11-02		Payload delivery	filename	swaqp.exe			PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	1393
☐	2020-11-02		Payload delivery	filename	Silence.Downloader			PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	1393
☐	2020-11-02		Payload delivery	filename	FlawedAmmyy.downloader			PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	1393
☐	2020-11-02		Payload delivery	filename	Dridex	Dridex x		PDF-to-text from file Report TA505- CERTFR-2020-CTI-009.pdf	☑	309 569 1024
☐	2020-11-02		External analysis	attachment	Report TA505- CERTFR-2020-CTI-009.pdf				☑	1393

Figura 5.12: Atributos enriquecidos a partir do relatório [7]

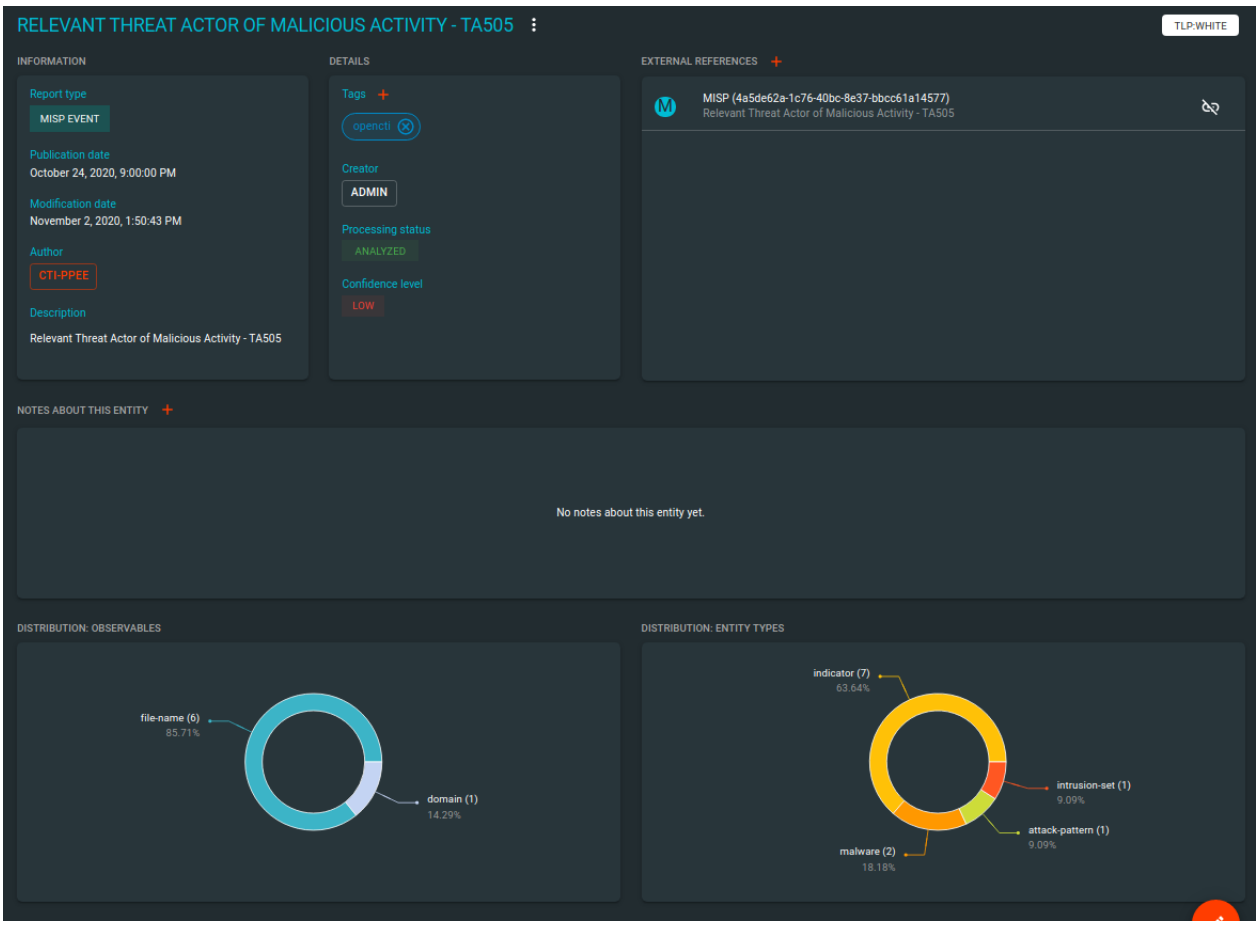


Figura 5.13: Estudo analítico do evento relacionado ao relatório [7]

entidade *Intrusion Set*, seus atributos relacionados a ele por meio da classificação de relação *indicates* e os *Malwares* por meio da classificação de relação *uses*.

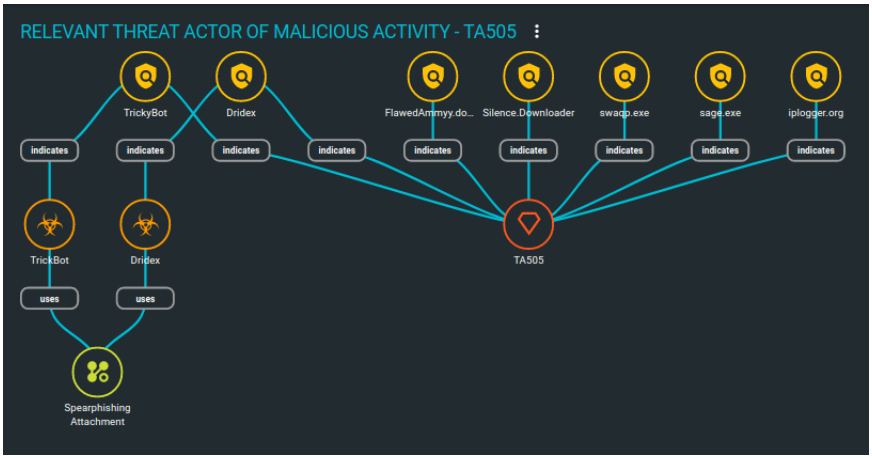


Figura 5.14: Grafo de relacionamentos do evento relacionado ao relatório [7]

Por meio da integração da plataforma OpenCTI com o conector do *framework* MITRE ATT&CK, outras informações foram relacionadas a entidade TA505 e estão apresentadas nas Figuras 5.15 a 5.18. Ao navegar na plataforma, por meio da seleção das entidades de interesse no grafo ilustrado na Figura 5.14, é

possível reunir informações específicas referentes ao ator dos ataques, no caso o grupo criminoso TA505, os principais *Malwares* utilizados por ele e os padrões de ataque utilizados.

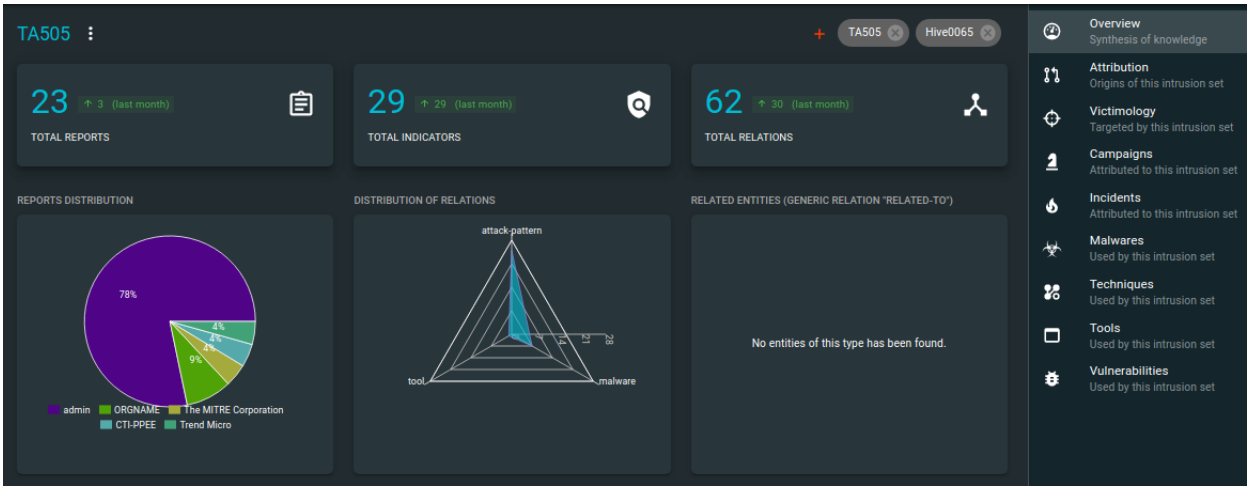


Figura 5.15: Estudo analítico do grupo criminoso TA505

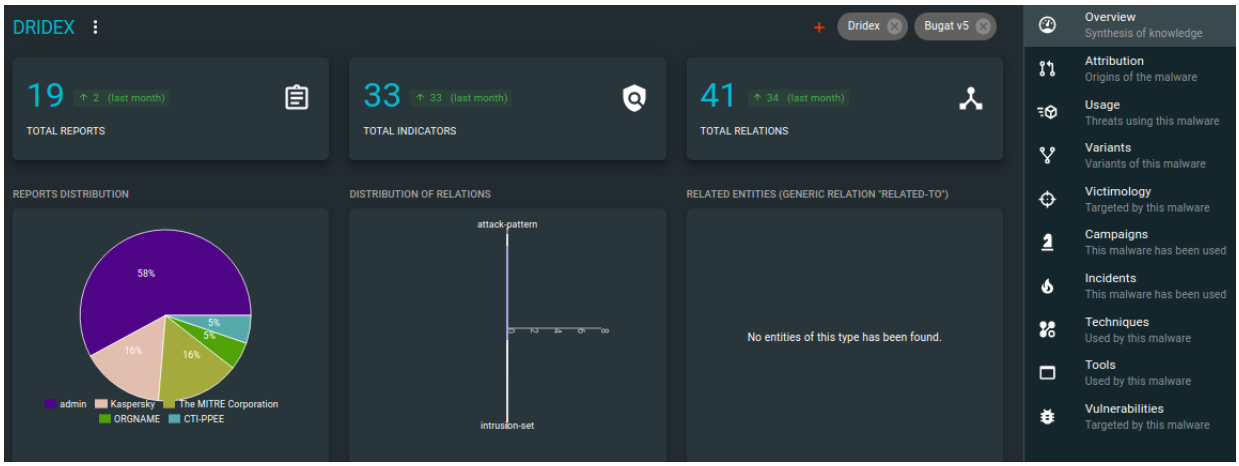


Figura 5.16: Estudo analítico do *Malware* Dridex

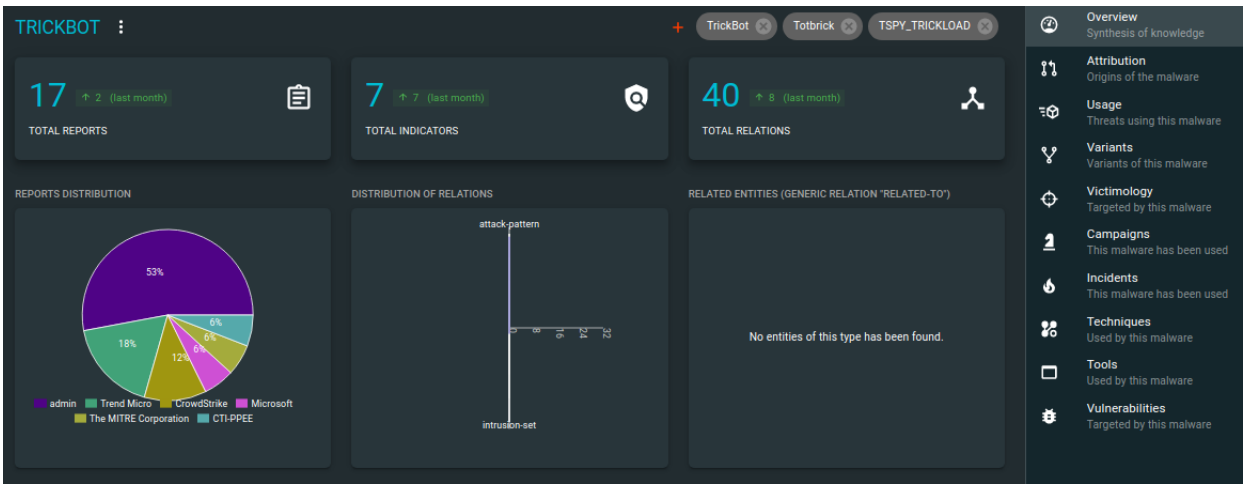


Figura 5.17: Estudo analítico do *Malware* TrickBot

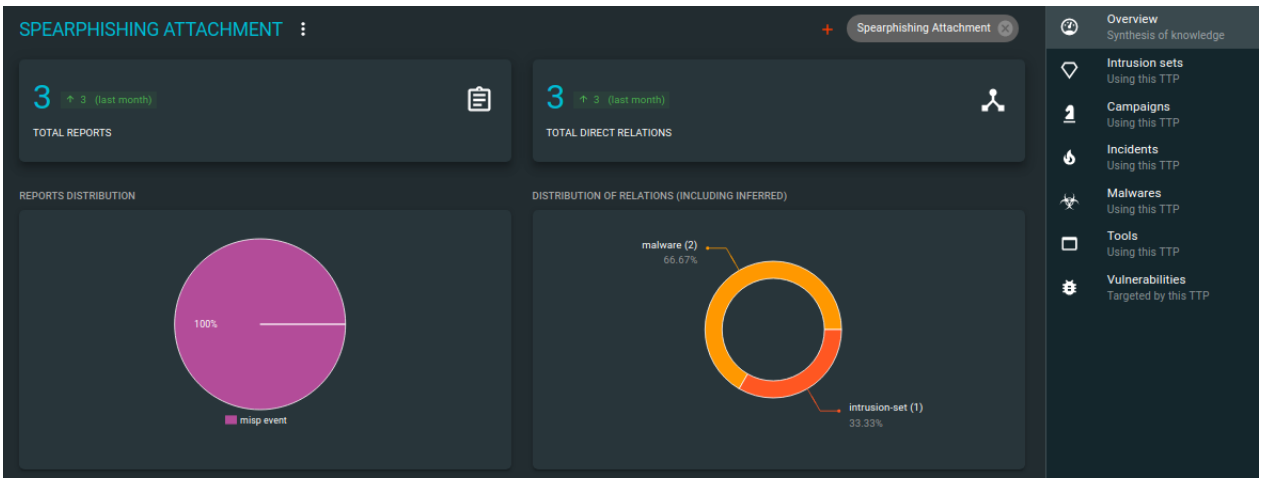


Figura 5.18: Estudo analítico do padrão de ataque *Spearphishing Attachment*

5.2.1 Análise dos resultados

Quando analisada pelo método 5W3H pode-se garantir a resposta dos questionamentos *what, when, who, how e how much* e consegue-se entender parcialmente o questionamento *where*. *What* se refere a entidade em que foi convertido, no caso um *intrusion set*. *When* contempla as informações de primeira e última aparição, além do histórico de modificação. *Who* pode ser descrito dado que a entidade principal é um *intrusion set* amplamente conhecido. *How* é bem definido a partir da análise das diversas TTPs descritas. *How much* pode ser entendido com as informações retratadas na classe de impacto. *Where* pode ser genericamente descrito considerando o número de países afetados apresentados na síntese.

Desse modo, a conexão entre os dados analisados, importados da ferramenta MISP, em conjunto com a síntese de conhecimento baseada dos dados oriundos da plataforma MITRE ATT&CK, quando avaliada considerando os valores propostos na Tabela 4.1, resultou em uma satisfatória especificidade e completude de informações.

5.3 SÍNTESE DOS RESULTADOS

Com base nos resultados obtidos a partir da aplicação dos cenários de testes propostos, algumas conclusões puderam ser tomadas em relação às plataformas utilizados e arquitetura de integração proposta.

5.3.1 Plataforma MISP

Assim como esperado, a plataforma MISP se comporta de forma eficiente no processamento de grandes quantidades de dados. Até mesmo em importações de tabelas com uma quantidade massiva de indicadores, o processo ocorre um pouquíssimos segundos. Em acréscimo, seus diversos módulos de enriquecimento permitem a obtenção de novos atributos relacionados a aqueles já adicionados, o que torna a completude da descrição dos eventos muito maior. Além disso, a inclusão dos dados nessa plataforma permite uma interconexão com diversas outras soluções devido a possibilidade de exportação dos dados em diferentes

formatos. Por outro lado, os mecanismos de correlação e enriquecimento de dados disponíveis na plataforma MISP são em sua maioria baseados em análise de texto livre. Esse tipo de análise pode gerar uma grande quantidade de falsos positivos dependendo do conjunto de dados inicial que é analisado. Ainda, essa característica não permite a criação de relacionamentos totalmente confiáveis entre os eventos presentes na plataforma, dado que não leva em consideração nenhum tipo de contextualização mas apenas a semelhança entre as palavras presentes na definição e descrição dos atributos. Além disso, a visualização das informações não é intuitiva ao ponto de auxiliar na tomada de decisão acerca de um incidente ou ameaça.

Desse modo, a plataforma pode ser considerada como uma boa solução para o processamento de dados não estruturados e indicadores de comprometimento, porém por si só não é capaz de produzir inteligência de ameaça de qualidade, se analisada pela metodologia de avaliação proposta por este trabalho que se baseia no método 5W3H. Sua funcionalidade pode ser amplamente otimizada quando aplicada como um banco de dados povoado com indicadores de comprometimento e dados de ameaça já estruturados, e utilizada como fonte de informação para outras aplicações com foco em análise e visualização.

5.3.2 Plataforma OpenCTI

No contexto deste trabalho, a plataforma OpenCTI foi utilizada para acrescentar análise as informações disponibilizadas e apresentá-las de maneira intuitiva. Nesse sentido, pode-se dizer que a plataforma cumpriu com êxito o seu objetivo. Sua apresentação de conhecimento baseada em grafos, traz de forma bastante clara as relações entre entidades e indicadores envolvidos em um incidente ou ameaça. Além disso, cada um dos observáveis e entidades adicionados à plataforma, são estudados analiticamente, agregando informação e auxiliando na produção de inteligência de ameaça.

Ressalta-se que a aplicação da plataforma é potencializada quando aplicada a contextos mais específicos. Isso porque sua visualização baseada em grafos pode se tornar confusa quando envolve muitas entidades e diferentes relações. Além disso, a inclusão de indicadores totalmente descontextualizados, como por exemplo uma lista de IPs maliciosos aleatórios, pode gerar um grafo totalmente desconexo que não agrega conhecimento.

5.3.3 Integração entre as plataformas

Em relação a integração das plataformas e a complementaridade de suas funções, foram percebidas vantagens e desvantagens ao utilizá-las de forma conjunta. Em resumo, as vantagens e desvantagens observadas foram sintetizadas na Tabela 5.1 e categorizadas com base nos processos envolvidos na metodologia integrativa proposta.

Em relação às vantagens, um primeiro ponto a ser analisado é o processo de integração entre as plataformas. Pode-se dizer que esse processo é eficiente e trouxe os resultados esperados. Os conectores estavam configurados para trabalhar com atualizações de 1 minuto e todas elas ocorriam de forma performática e transparente ao usuário. Além disso, como os conectores são disponibilizados em código livre, alterações podem ser feitas para que realizem funcionalidades personalizadas em cada caso de aplicação. Outro ponto

Tabela 5.1: Síntese de vantagens e desvantagens da metodologia

Processo	Vantagens	Desvantagens
Integração	Fácil instalação	Conector MISP dependente de *tag* e organização criadora
	Conectores com configurações simples	Diferenças nos modelos de dados dificultam a conversão de entidades/indicadores
	Conectores de módulo aberto permitem costumizações	-
Processamento	Boa performance para grande quantidade de dados	Grande quantidade de falsos positivos devido análise de texto livre para identificação de indicadores
	Dados já estruturados são melhor trabalhados na plataforma OpenCTI	-
	Grande diversidade de módulos para importação, exportação e enriquecimento	-
Análise	Mecanismo de relacionamento da plataforma OpenCTI agrega informação aos eventos MISP	Muitos indicadores podem gerar grafos confusos
	Os relacionamentos gerados facilitam a compreensão do tópico quando aplicado o método 5W3H	Dados sem contexto específico não geram relações ou relações pouco relevantes
	Fácil navegação pelos ícones de visualização para obtenção de mais informações para análise	-
Visualização	Dashboards com informações analíticas	-
	Grafos de conhecimento intuitivos	-

é que a integração foi baseada em *tags*, permitindo que apenas dados relevantes fossem compartilhados e analisados.

Além disso, como mencionado anteriormente, inferir conhecimento a partir das informações disponibilizadas na plataforma MISP pode ser complexo, dado o seu modelo de apresentação baseado em listagem. Nesse quesito, a transferência desse processo para a plataforma OpenCTI otimiza a utilização das informações. Em acréscimo, os mecanismos de correlação da plataforma OpenCTI proporcionam uma correlação maior dos indicadores e entidades, não focando apenas em semelhança textual entre os eventos, o que agrega qualidade às informações.

Já na perspectiva de desvantagens, o maior problema encontrado diz respeito às diferenças existentes entre o modelo de dados das duas plataformas. A ferramenta MISP possui um modelo flexível e que pode variar de acordo com o evento analisado. Por outro lado, a plataforma OpenCTI é baseada no formato STIX e possui suas entidades e relações muito bem definidas. Desse modo, nem todas as informações adicionadas na plataforma MISP possuem conversão direta para o modelo de dados baseado em STIX. Essa diferenciação dificulta na conversão de entidades e atributos entre as plataformas e por isso, algumas vezes a análise de eventos MISP realizada na plataforma OpenCTI pode se tornar superficial, não demonstrando toda a potencialidade de correlação da plataforma.

5.4 AVALIAÇÃO GERAL DA METODOLOGIA

Os cenários de testes planejados e executados neste trabalho buscavam a obtenção de resultados para subsidiar a avaliação da eficácia da metodologia proposta para produção e análise de qualidade de inteligência de ameaça. Para garantir que a informação produzida fosse de qualidade, os resultados obtidos foram analisados com base no método 5W3H. Nesse contexto, uma avaliação geral dos resultados obtidos e análises realizadas é descrita a seguir.

Considerando os resultados finais obtidos na produção de inteligência de ameaça, pode-se dizer que na maioria dos casos, os dados e indicadores são melhor relacionados e visualizados a partir da aplicação da metodologia de produção de inteligência proposta. A simples correlação realizada pela plataforma MISP apenas identifica eventos semelhantes textualmente, o que não fornece uma análise suficiente para produção de inteligência de ameaça efetiva. Ao integrar a ferramenta OpenCTI ao processo, os dados são melhor analisados, relacionados e visualizados o que possibilita a produção de inteligência de ameaça acionável.

Na mesma linha, percebeu-se que em cenários mais específicos e contextualizados, a informação analisada e correlacionada com os dados do *framework* Mitre ATT&CK se enquadra dentro das expectativas do método 5W3H, auxiliando a obtenção de respostas diretas acerca de ameaças e incidentes. Desse modo, comprova-se que a metodologia tem potencial para produzir inteligência de ameaça efetiva.

Mesmo com as diferenças entre os modelos de dados das duas plataformas e as dificuldades relacionadas a conversão de atributos e entidades, ainda é possível indicar que a qualidade da informação é otimizada quando aplicada a arquitetura de integração. Isso porque as informações ainda são estudadas analiticamente pela plataforma OpenCTI, agregando conhecimento. Além disso, são indicadas relações dentro da plataforma e a navegação entre entidades e observáveis correlatos dentro da plataforma permite a união de informações para obtenção de inteligência de ameaça.

Por outro lado, notou-se também pontos de falha na aplicação da metodologia proposta. Esses pontos se mostraram significantes em cenários que envolvem uma grande quantidade de dados ou dados sem um contexto específico previamente definido. Em relação à quantidade de dados, em termos de processamento, pode-se dizer que os resultados são satisfatórios. Porém, se tratando do processo de análise, por ser uma metodologia dependente do modelo de visualização das informações, uma grande quantidade de dados relacionadas pode tornar os resultados confusos e poucos intuitivos. Além disso, indicadores sem contexto definido não permitem a geração de relacionamentos diretos capazes de responder os questionamentos levantados pelo método 5W3H, e por isso não podem ser considerados inteligência de ameaça.

Por fim, partindo de todas as considerações supracitadas, é importante ressaltar que dados e informações descontextualizados não são efetivos no contexto de produção de inteligência de ameaça. Quando indicadores de comprometimento são disponibilizados sem contexto, mesmo que esses sejam relevantes, a análise de inferências e relacionamentos se torna muito mais complicada, e a conversão do dado em informação ou inteligência de ameaça praticamente impossível. Desse modo, mesmo adotando uma arquitetura embasada na complementaridade de plataformas de TI bem avaliadas, com cenários de testes bem sucedidos quanto a produção de inteligência, indicadores descontextualizados permaneceram como dados aleatórios.

6 CONCLUSÃO

Neste trabalho foi proposta uma metodologia integrativa para implementar uma solução mais completa no contexto de produção de inteligência de ameaças cibernéticas. A essência da metodologia proposta estava na interoperabilidade entre plataformas de TI de código aberto com características complementares, que em conjunto teriam suas funcionalidades otimizadas. Além disso, este trabalho adotou um método de análise da qualidade da inteligência produzida, objetivando comprovar a eficiência da metodologia proposta.

Para atingir os objetivos supracitados, primeiramente, foi conduzido um estudo sobre o mercado atual de plataformas de TI de código aberto. A partir dos resultados obtidos com esse estudo, foi desenvolvida uma estratégia de seleção das plataformas mais populares e uma metodologia para avaliação das plataformas selecionadas. A avaliação realizada identificou plataformas de TI de código aberto com grande potencial de aplicação e gerou resultados relevantes e imprescindíveis para o desenvolvimento da metodologia integrativa proposta e solução implementada.

A arquitetura implementada para produção de inteligência de ameaças cibernéticas neste trabalho teve como base a integração das plataformas MISP e OpenCTI, e mostrou-se eficiente quando aplicada em cenários previamente contextualizados e que não envolvam uma quantidade massiva de indicadores. A utilização de conjuntos de dados compostos por indicadores de comprometimento gerais e sem contexto não possibilitam a geração de relacionamentos diretos entre eles, sendo esse um mecanismo de análise essencial para o funcionamento do método proposto. Além disso, uma ampla gama de indicadores e dados correlacionados podem gerar grafos pouco intuitivos, dificultando a extração de informação.

O método de análise da qualidade da inteligência produzida adotado, embasado nas informações obtidas a partir dos questionamentos levantados pelo método 5W3H, permitiu verificar a completude da inteligência gerada e foi considerado satisfatório para a execução dessa tarefa. Com ele, foi possível averiguar e atestar a viabilidade de utilização da arquitetura proposta em diferentes cenários de testes.

Em linhas gerais, os resultados obtidos a partir da execução dos cenários de teste propostos permitem a percepção de pontos importantes acerca do contexto de produção de inteligência de ameaça. Primeiramente, mesmo existindo boas plataformas de TI de código aberto disponíveis e havendo vários esforços para o desenvolvimento de módulos de integração entre essas plataformas, o desenho e desenvolvimento de uma ferramenta única, projetada desde seu início para compreender todos os processos de produção de inteligência de ameaça seria a solução ideal para produção de inteligência de qualidade. Além disso, independentemente da plataforma utilizada para a produção de inteligência acionável, é sempre importante que a qualidade dados utilizados e compartilhados seja priorizada em relação a quantidade, de modo a garantir a possibilidade de contextualização e correlação das informações.

6.1 TRABALHOS FUTUROS

Com base nos resultados e conclusões obtidos a partir da realização deste trabalho, alguns prontos que podem ser estudados, testados e/ou implementados são propostos como trabalhos futuros.

Em relação a avaliação comparativa desempenhada sobre plataformas de TI, considera-se relevante a validação da metodologia e dos resultados apresentados com base na execução de testes práticos, envolvendo conjuntos de dados de ameaças cibernéticas reais, para todas as plataformas avaliadas.

Na perspectiva da metodologia integrativa proposta, o aperfeiçoamento do conector utilizado para integrar as plataformas MISP e OpenCTI, no sentido de adicionar equivalências entre os modelos de dados das plataformas e aumentar o espectro de conversão e relacionamento entre indicadores e entidades, poderia auxiliar na produção de inteligência de ameaça em cenários pouco contextualizados.

Outro ponto importante é a necessidade de delinear uma definição padronizada para o conceito e processo de CTI visando auxiliar o desenvolvimento de novos sistemas de inteligência de ameaça otimizados, capazes de compreender todos os processos definidos como necessários para produção de inteligência de ameaça e estabelecer um modelo de defesa eficiente.

Por fim, é sempre interessante a proposição de mais cenários de teste para a metodologia integrativa desenvolvida, incluindo a aplicação em cenários mais próximos de situações reais. Desse modo, novos resultados seriam obtidos, tornando possível uma análise ainda mais robusta do método proposto neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

1 FLOOD, M. D.; LEMIEUX, V. L.; VARGA, M.; WONG, B. L. W. The application of visual analytics to financial stability monitoring. *SSRN Electronic Journal*, Elsevier BV, 2014. Disponível em: <https://doi.org/10.2139/ssrn.2438194>.

2 ABU, M. S.; SELAMAT, S. R.; ARIFFIN, A.; YUSOF, R. Cyber threat intelligence – issue and challenges. *Indonesian Journal of Electrical Engineering and Computer Science*, Institute of Advanced Engineering and Science, v. 10, n. 1, p. 371, 4 2018. ISSN 2502-4752. Disponível em: <https://doi.org/10.11591/ijeecs.v10.i1.pp371-379>.

3 OASIS. *Identifying a Threat Actor Profile*. 2021. <https://oasis-open.github.io/cti-documentation/examples/identifying-a-threat-actor-profile>. [Accessed Jan 2021].

4 KAMPANAKIS, P. *IODEF Usage Guidance*. 2021. <https://tools.ietf.org/id/draft-ietf-mile-iodef-guidance-09.html>. [Accessed Jan 2021].

5 KASPERSKY. *Threat Lookup: Exporting to OpenIOC*. 2021. <https://tip.kaspersky.com/help/Doc_data/ExportingToOpenIOC.htm>. [Accessed Jan 2021].

6 OPENCTI. *OpenCTI documentation 3.0.2*. 2019. <https://github.com/OpenCTI-Platform/opencti>. [Accessed May 2020].

7 ANSSI. *Development of the activity of the TA505 cybercriminal group*. 2020. Disponível em: https://www.cert.ssi.gouv.fr/uploads/CERTFR-2020-CTI-009.pdf. Acesso em Outubro de 2020. Disponível em: <https://www.cert.ssi.gouv.fr/uploads/CERTFR-2020-CTI-009.pdf>.

8 BURGER, E. W.; GOODMAN, M. D.; KAMPANAKIS, P.; ZHU, K. A. Taxonomy model for cyber threat intelligence information exchange technologies. In: *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security - WISCS-14*. ACM Press, 2014. ISBN 978-1-4503-3151-7. Disponível em: <https://doi.org/10.1145/2663876.2663883>.

9 KLOCK, A. C. T.; GASPARINI, I.; PIMENTA, M. S. 5w2h framework. In: *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*. ACM, 2016. Disponível em: <https://doi.org/10.1145/3033701.3033715>.

10 POKORNY, Z. The threat intelligence handbook: Moving toward a security intelligence program. *Annapolis, CyberEdge Group*, 2019.

11 BISSELL, K.; LASALLE, R.; CIN, P. D. *The cost of cybercrime - Ninth annual cost of cybercrime study*. 2019. <https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-study-final.pdf>. [Accessed May 2020].

12 BISSELL, K.; LASALLE, R.; CIN, P. D. *The 2020 Cyber Security Report*. 2020. <https://pages.checkpoint.com/cyber-security-report-2020>. [Accessed May 2020].

13 TOUNSI, W. What is cyber threat intelligence and how is it evolving? In: *Cyber-Vigilance and Digital Trust*. John Wiley & Sons, Inc., 2019. p. 1–49. ISBN 978-1-1196-1839-3. Disponível em: <https://doi.org/10.1002/9781119618393.ch1>.

14 ALSHAMRANI, A.; MYNENI, S.; CHOWDHARY, A.; HUANG, D. A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials*, IEEE, v. 21, n. 2, p. 1851–1877, 2019. ISSN 1553-877X.

15 WU, J. New approaches to cyber defense. In: *Cyberspace Mimic Defense*. [S.l.]: Springer, 2020. p. 113–157. ISBN 978-3-030-29844-9.

16 CHADWICK, D. W.; FAN, W.; COSTANTINO, G.; LEMOS, R. de; CERBO, F. D.; HERWONO, I.; MANEA, M.; MORI, P.; SAJJAD, A.; WANG, X.-S. A cloud-edge based data security architecture for sharing and analysing cyber threat information. *Future Generation Computer Systems*, Elsevier BV, v. 102, p. 710–722, 1 2020. Disponível em: <https://doi.org/10.1016/j.future.2019.06.026>.

17 RAMSDALE, A.; SHIAELES, S.; KOLOKOTRONIS, N. A comparative analysis of cyber-threat intelligence sources, formats and languages. *Electronics*, MDPI AG, v. 9, n. 5, p. 824, 5 2020. Disponível em: <https://doi.org/10.3390/electronics9050824>.

18 BAUER, S.; FISCHER, D.; SAUERWEIN, C.; LATZEL, S.; STELZER, D.; BREU, R. Towards an evaluation framework for threat intelligence sharing platforms. In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 2020. Disponível em: <https://doi.org/10.24251/hicss.2020.239>.

19 SHIN, B.; LOWRY, P. B. A review and theoretical explanation of the 'cyberthreat-intelligence (CTI) capability' that needs to be fostered in information security practitioners and how this can be accomplished. *Computers & Security*, Elsevier BV, v. 92, p. 101761, 5 2020. Disponível em: <https://doi.org/10.1016/j.cose.2020.101761>.

20 RANTOS, K.; SPYROS, A.; PAPANIKOLAOU, A.; KRITSAS, A.; ILIOUDIS, C.; KATOS, V. Interoperability challenges in the cybersecurity information sharing ecosystem. *Computers*, MDPI AG, v. 9, n. 1, p. 18, 3 2020. Disponível em: <https://doi.org/10.3390/computers9010018>.

21 ZHAO, J.; YAN, Q.; LI, J.; SHAO, M.; HE, Z.; LI, B. TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. *Computers & Security*, Elsevier BV, v. 95, p. 101867, 8 2020. Disponível em: <https://doi.org/10.1016/j.cose.2020.101867>.

22 SILVA, A. de Melo e; GONDIM, J. J. C.; ALBUQUERQUE, R. de O.; VILLALBA, L. J. G. A methodology to evaluate standards and platforms within cyber threat intelligence. *Future Internet*, MDPI AG, v. 12, n. 6, p. 108, jun. 2020. Disponível em: <https://doi.org/10.3390/fi12060108>.

23 RIESCO, R.; LARRIVA-NOVO, X.; VILLAGRA, V. A. Cybersecurity threat intelligence knowledge exchange based on blockchain. *Telecommunication Systems*, Springer Science and Business Media LLC, v. 73, n. 2, p. 259–288, 9 2019. Disponível em: <https://doi.org/10.1007/s11235-019-00613-4>.

24 TOUNSI, W.; RAIS, H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, Elsevier BV, v. 72, p. 212–233, 1 2018. ISSN 0167-4048. Disponível em: <https://doi.org/10.1016/j.cose.2017.09.001>.

25 GAO, Y.; LI, X.; PENG, H.; FANG, B.; YU, P. HinCTI: A cyber threat intelligence modeling and identification system based on heterogeneous information network. *IEEE Transactions on Knowledge and Data Engineering*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2020. Disponível em: <https://doi.org/10.1109/tkde.2020.2987019>.

26 FERREIRA, H. G. C.; JUNIOR, R. T. de S. Security analysis of a proposed internet of things middleware. *Cluster Computing*, Springer Science and Business Media LLC, v. 20, n. 1, p. 651–660, 1 2017. Disponível em: <https://doi.org/10.1007/s10586-017-0729-3>.

27 SILLABER, C.; SAUERWEIN, C.; MUSSMANN, A.; BREU, R. Data quality challenges and future research directions in threat intelligence sharing practice. In: *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security - WISCS16*. ACM Press, 2016. ISBN 978-1-4503-4565-1. Disponível em: <https://doi.org/10.1145/2994539.2994546>.

28 BARNUM, S. *Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX)*. 2012. 1–22 p. <https://www.mitre.org/publications/technical-papers/standardizing-cyber-threat-intelligence-information-with-the>. [Accessed March 2020].

29 FRIEDMAN, J.; BOUCHARD, M. *Definitive Guide to Cyber Threat Intelligence: Using Knowledge about Adversaries to Win the War against Targeted Attacks*. [S.l.]: CyberEdge Group, 2015.

30 CHISMON, D.; RUKS, M. Threat intelligence: Collecting, analysing, evaluating. *MWR InfoSecurity Ltd*, 2015.

31 PLANQUE, D. *Cyber Threat Intelligence: From confusion to clarity*. 2017. Disponível em: https://openaccess.leidenuniv.nl/bitstream/handle/1887/64551/. Acesso em Agosto de 2020. Disponível em: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/64551/Planque_D_2017_CS.pdf>.

32 CERT-UK. An introduction to threat intelligence. *CERT-UK Publication*, 2015. [Accessed May 2020].

33 MENGES, F.; PERNUL, G. A comparative analysis of incident reporting formats. *Computers & Security*, Elsevier BV, v. 73, p. 87–101, 3 2018. ISSN 0167-4048. Disponível em: <https://doi.org/10.1016/j.cose.2017.10.009>.

34 ASGARLI, E.; BURGER, E. Semantic ontologies for cyber threat sharing standards. In: *2016 IEEE Symposium on Technologies for Homeland Security (HST)*. IEEE, 2016. ISBN 978-1-5090-0770-7. Disponível em: <https://doi.org/10.1109/ths.2016.7568896>.

35 ENISA. *Exploring the opportunities and limitations of current Threat Intelligence Platforms*. 2018. <https://www.enisa.europa.eu/publications/exploring-the-opportunities-and-limitations-of-current-threat-intelligence-platforms>. [Accessed March 2020].

36 SKOPIK, F.; SETTANNI, G.; FIEDLER, R. A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, Elsevier BV, v. 60, p. 154–176, 7 2016. ISSN 0167-4048. Disponível em: <https://doi.org/10.1016/j.cose.2016.04.003>.

37 POPUTA-CLEAN, P.; STINGLEY, M. *Automated Defense-Using Threat Intelligence to Augment Security*. 2015. <https://www.sans.org/reading-room/whitepapers/threats/paper/35692>. [Accessed March 2020].

38 WAGNER, T. D.; MAHBUB, K.; PALOMAR, E.; ABDALLAH, A. E. Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, Elsevier BV, v. 87, p. 101589, 11 2019. ISSN 0167-4048. Disponível em: <https://doi.org/10.1016/j.cose.2019.101589>.

39 STEINBERGER, J.; SPEROTTO, A.; GOLLING, M.; BAIER, H. How to exchange security events? overview and evaluation of formats and protocols. In: *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2015. ISBN 978-1-4799-8241-7. Disponível em: <https://doi.org/10.1109/inm.2015.7140300>.

40 SAUERWEIN, C.; SILLABER, C.; MUSSMANN, A.; BREU, R. Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. In: *13th International Conference on Wirtschaftsinformatik*. [S.l.: s.n.], 2017.

41 TRUONG, T. C.; ZELINKA, I.; PLUCAR, J.; ČANDÍK, M.; ŠULC, V. Artificial intelligence and cybersecurity: Past, presence, and future. In: *Advances in Intelligent Systems and Computing*. Springer Singapore, 2020. p. 351–363. Disponível em: <https://doi.org/10.1007/978-981-15-0199-9_30>.

42 NOOR, U.; ANWAR, Z.; AMJAD, T.; CHOO, K.-K. R. A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems*, Elsevier BV, v. 96, p. 227–242, 7 2019. Disponível em: <https://doi.org/10.1016/j.future.2019.02.013>.

43 DALTON, A.; AGHAEI, E.; AL-SHAER, E.; BHATIA, A.; CASTILLO, E.; CHENG, Z.; DHADUVAI, S.; DUAN, Q.; ISLAM, M. M.; KARIMI, Y.; MASOUMZADEH, A.; MATHER, B.; SANTHANAM, S.; SHAIKH, S.; STRZALKOWSKI, T.; DORR, B. J. The panacea threat intelligence and active defense platform. *ArXiv*, abs/2004.09662, 2020.

44 MARTINSO, C.; MEDEIROS, I. Generating threat intelligence by classification and association of security events. *DSN Workshop on Data-Centric Dependability and Security*, 2019.

45 KAZATO, Y.; NAKAGAWA, Y.; NAKATANI, Y. Improving maliciousness estimation of indicator of compromise using graph convolutional networks. In: *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2020. Disponível em: <https://doi.org/10.1109/ccnc46108.2020.9045113>.

46 MAVROEIDIS, V.; BROMANDER, S. Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. In: *2017 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, 2017. ISBN 978-1-5386-2385-5. Disponível em: <https://doi.org/10.1109/eisic.2017.20>.

47 SHACKLEFORD, D. Cyber threat intelligence uses, successes and failures: The sans 2017 cti survey. *SANS Institute, InfoSec Reading Room*, 2017.

48 OASIS. *STIX Version 2.0. Part 1: STIX Core Concepts*. 2017. <http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html>. [Accessed May 2020].

49 OASIS. *STIX Version 2.0. Part 2: STIX Objects*. 2017. <http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html>. [Accessed May 2020].

50 CORPORATION, M. *Cyber Observable eXpression (CybOX^{TM}) Archive Website*. 2017. <https://cyboxproject.github.io/>. [Accessed May 2020].

51 TC, O. C. T. I. C. *STIX^{TM} Version 2.0. Part 3: Cyber Observable Core Concepts*. 2017. <http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part3-cyber-observable-core.pdf>. [Accessed May 2020].

52 OASIS. *TAXII Version 2.0*. 2017. <http://docs.oasis-open.org/cti/taxii/v2.0/taxii-v2.0.html>. [Accessed May 2020].

53 DANYLIW, R.; MEIJER, J.; DEMCHENKO, Y. *The Incident Object Description Exchange Format*. 2007. <https://tools.ietf.org/html/rfc5070>. [Accessed May 2020].

54 DANYLIW, R. *The Incident Object Description Exchange Format Version 2*. 2016. <https://tools.ietf.org/html/rfc7970>. [Accessed May 2020].

55 MORIARTY, K. *Real-time Inter-network Defense (RID)*. 2012. <https://tools.ietf.org/html/rfc6545>. [Accessed May 2020].

56 INC., M. *An Introduction to Open IOC*. 2011. <https://www.academia.edu/31820654/An_Introduction_to_Open_IOC>. [Accessed May 2020].

57 WAGNER, T. D.; PALOMAR, E.; MAHBUB, K.; ABDALLAH, A. E. Relevance filtering for shared cyber threat intelligence (short paper). In: *Information Security Practice and Experience*. Springer International Publishing, 2017. p. 576–586. ISBN 978-3-319-72359-4. Disponível em: <https://doi.org/10.1007/978-3-319-72359-4_35>.

58 LIU, R.; ZHAO, Z.; SUN, C.; YANG, X.; GONG, X.; ZHANG, J. A research and analysis method of open source threat intelligence data. In: *Communications in Computer and Information Science*. Springer Singapore, 2017. p. 352–363. ISBN 978-981-10-6385-5. Disponível em: <https://doi.org/10.1007/978-981-10-6385-5_30>.

59 ANSSI. *OpenCTI – The open source solution for processing and sharing threat intelligence knowledge*. 2020. <https://www.ssi.gouv.fr/en/actualite/opencti-the-open-source-solution-for-processing-and-sharing-threat-intelligence-knowledge/>. [Accessed May 2020].

60 GARNIER, F. *CTI & Information Fusion Benefits and Challenges*. 2020. <https://www.enisa.europa.eu/events/2019-cti-eu/presentations/200130-cti-info-fusion-tlp-white>. [Accessed May 2020].

61 PROJECT, M. *MISP - Open Source Threat Intelligence Platform & Open Standards For Threat Information Sharing*. 2020. <https://www.misp-project.org/features.html>. Acesso em: 03 mar 2020.

62 CORPORATION, M. *Welcome to CRITs*. 2020. <https://github.com/crits/crits#readme>. [Accessed May 2020].

63 CORPORATION, M. *Collaborative Research Into Threats*. 2020. <https://crits.github.io/#nav>. [Accessed May 2020].

64 GADGETS, C. *The FASTEST Way to Consume Threat Intelligence. Period.* 2018. <https://csirtgadgets.com/collective-intelligence-framework>. [Accessed May 2020].

65 IOVINO, G. *What is the Collective Intelligence Framework?* 2015. <https://github.com/csirtgadgets/massive-octo-spice/wiki/What-is-the-Collective-Intelligence-Framework%3F>. [Accessed May 2020].

66 ANOMALI. *Anomali STAXX - Installation and Administration Guide*. 2018. <https://update.anomali.com/staxx/docs/Anomali_STAXX_Installation_&_Administration_Guide.pdf>. [Accessed May 2020].

67 CORPORATION, T. M. *MITRE ATT&CK*. 2020. Disponível em: https://attack.mitre.org/. Acesso em Outubro de 2020. Disponível em: <https://attack.mitre.org/>.

68 ANOMALI. *COVID-19 Resource Center*. 2020. Disponível em: https://www.anomali.com/learn/covid-19-resources. Acesso em Outubro de 2020. Disponível em: <https://www.anomali.com/learn/covid-19-resources>.

69 ALIENVAULT. *WannaCry Indicators*. 2020. Disponível em: https://otx.alienvault.com/pulse/5915db384da2585b4feaf2f6. Acesso em Outubro de 2020. Disponível em: <https://otx.alienvault.com/pulse/5915db384da2585b4feaf2f6>.