DISSERTAÇÃO DE MESTRADO PROFISSIONAL

Modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados

Bruce William Percílio Azevedo

Brasília, Agosto de 2020

Programa de Pós-Graduação Profissional em Engenharia Elétrica

DEPARTAMENTO DE ENGENHARIA ELÉTRICA
FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

REFERENCE MODEL FOR THE DEVELOPMENT OF CYBER THREAT INTELLIGENCE APPLICATIONS AND ITS APPLICABILITY FOR DATA SHARING

MODELO DE REFERÊNCIA PARA O DESENVOLVIMENTO DE APLICAÇÕES DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS E SUA APLICABILIDADE PARA O COMPARTILHAMENTO DE DADOS

BRUCE WILLIAM PERCÍLIO AZEVEDO

ORIENTADOR: WILLIAM FERREIRA GIOZZA, DR.
COORIENTADOR: ROBSON DE O. ALBUQUERQUE, DR.

DISSERTAÇÃO DE MESTRADO PROFISSIONAL EM
ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPEE.MP.002
BRASÍLIA/DF: AGOSTO - 2020

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO PROFISSIONAL

Modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados

Bruce William Percílio Azevedo

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. Robson de O. Albuquerque, Ph.D, FT/UnB _____
Orientador

Prof. Georges Daniel Amvame-Nze, Ph.D, FT/UnB _____
examinador interno

Prof. André Ricardo Abed Grégio, Ph.D, UFPR _____
Examinador externo

Prof. Edna Dias Canedo, Ph.D, FT/UnB _____
Suplente

FICHA CATALOGRÁFICA

PERCÍLIO AZEVEDO, BRUCE WILLIAM
Modelo de referência para o desenvolvimento de aplicaçõesde inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados [Distrito Federal] 2020.
xvi, 63 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. Inteligência de ameaças cibernéticas
2. Modelo de referência
3. Compartilhamento de informação
4. Ameaças avançadas

REFERÊNCIA BIBLIOGRÁFICA

AZEVEDO, B. W. P. (2020). *Modelo de referência para o desenvolvimento de aplicaçõesde inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados*. Dissertação de mestrado profissional, Publicação: PPEE.MP.002, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 63 p.

CESSÃO DE DIREITOS

AUTOR: Bruce William Percílio Azevedo
TÍTULO: Modelo de referência para o desenvolvimento de aplicaçõesde inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados.
GRAU: Mestre em Engenharia Elétrica ANO: 2020

Bruce William Percílio Azevedo
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

DEDICATÓRIA

Dedico esse trabalho a todos aqueles que embora envoltos a grandes dificuldades e desilusões, nunca deixaram de buscar os sonhos que carregam.

AGRADECIMENTOS

Antes de tudo e todos, gostaria de agradeças a Deus, pois sem a sua permissão nada seria possível.

Agradeço aos meus pais e minha noiva, pelo apoio incontestável que me deram durante todo esse percurso.

Agradeço aos meus orientadores, por todas as lições valiosas que certamente levarei comigo por onde eu passar.

E por fim, agradeço a todos que direta ou indiretamente, fizeram parte da minha formação.

RESUMO

Título: Modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas e sua aplicabilidade para o compartilhamento de dados
Autor: Bruce William Percílio Azevedo
Orientador: William Ferreira Giozza, Dr.
Coorientador: Robson de O. Albuquerque, Dr.
Programa de Pós-Graduação Profissional em Engenharia Elétrica – Área de Concentração em Segurança Cibernética
Brasília, 3 de agosto de 2020

Devido a evolução das ameaças que assolam o âmbito cibernético, contramedidas convencionais não tem mais a eficácia necessária para pará-las. As ferramentas de CTI (inteligência de ameaças cibernéticas) surgiram como alternativa informacional e de contexto para reduzir a quantidade de ataques bem-sucedidos, de maneira que combatem essas ameaças criando conhecimento operacional e estratégico, tendo como origem a mesma fonte de dados que essas ameaças são fundamentadas, a Internet. O presente trabalho apresenta um modelo de referência voltado para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas, cujo sua estrutura é composta de funcionalidades e caraterísticas inerentes a essa tecnologia. Esta proposta de modelo foi criada baseada em suporte acadêmico, da indústria de segurança cibernética e na análise de outras ferramentas. Esse conjunto de conhecimento serviu como fundamento para o planejamento de uma aplicação de CTI. Essa aplicação foi desenvolvida com propósitos demonstrativos, proporcionando visibilidade de dados estruturados no padrão STIX v2 com mapa de vínculos. O compartilhamento de dados foi implementado utilizando o padrão TAXII, característica esta, restrita a poucas outras aplicações de CTI. Assim esse trabalho proporciona uma revisão sobre CTI e seus padrões, bem como desenvolve um modelo de referência seguido de uma implementação com os padrões STIX v2 e TAXII como prova de conceito de seu funcionamento.

ABSTRACT

Title: Reference model for the development of cyber threat intelligence applications and its applicability for data sharing
Author: Bruce William Percílio Azevedo
Supervisor: William Ferreira Giozza, Dr.
Co-Supervisor: Robson de O. Albuquerque, Dr.
Professional Post-Graduate Program in Electrical Engineering – Cybersecurity Concentration Area
Brasília, August 3, 2020

Due to the evolution of threats that plague the cyber scope, conventional countermeasures no longer have the necessary efficiency to stop them. CTI (cyber threat intelligence) tools have emerged as an informational and contextual alternative to reduce the number of successful attacks, so that they combat these threats by creating operational and strategic knowledge, originating from the same source of data that these threats are, the Internet. The present work introduces a reference model aimed at the development of cyber threat intelligence applications, whose structure is composed of features and characteristics inherent to this technology. This model proposal was created based on academic support, from the cybersecurity industry and on the analysis of other tools. This set of knowledge served as a foundation for planning a CTI application. This application was developed for demonstrative purposes, providing visibility of data structured in the STIX v2 standard with link map. Data sharing was implemented using the TAXII standard, which is restricted to a few other CTI applications. Thus, this work provides a review on CTI and its standards, as well as develops a reference model followed by an implementation with STIX v2 and TAXII standards as proof of concept of its operation.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ACRÔNIMOS

Siglas

A IPA japonesa	Agência de Promoção de Tecnologia da Informação, Japão
API	*Application Programming Interface* (Interface de Programação de Aplicativo)
APP	*Application* (Aplicação)
APT	*Advanced Persistent Threat* (Ameaça Persistente Avançada)
ATT&CK	*Adversarial Tactics, Techniques, and Common Knowledge* (Táticas, Técnicas e Conhecimento Comum Sobre Adversários)
AWS	*Amazon Web Services* (Serviços Web da Amazon)
BTC	*Bitcoin*
CDN	*Computer Network Defense* (Defesa de Rede de Computadores)
CERN	*European Organization for Nuclear Research* (Organização Europeia para Pesquisa Nuclear)
CFM	*Cyber Federated Model* (Modelo cibernético Federado)
CIAWI	*Conferencia Ibero Americana WWW/INTERNET*
CIF	*Collective Intelligence Framework* (Estrutura de Inteligência Coletiva)
CISCP	*Certified International Supply Chain Professional* (Profissional Certificado Internacional de Cadeia de Suprimentos)
CPE	*Common Platform Enumeration* (Plataforma de Enumeração Comum)
CS&C	*Cybersecurity and Communications* (Cibersegurança e Comunicações)
CTI	*Cyber Threat Intelligence* (Inteligência de Ameaças Cibernéticas)
CVE	*Common Vulnerabilities and Exposures* (Vulnerabilidades e Exposições Comuns)
CybOx	*Cyber Observable eXpres-sion* (Expressão cibernética observável)
DHS	*United States Department of Homeland Security* (Departamento de Segurança Interna dos Estados Unidos)
EUA	Estados Unidos da América
GCP	*Google Cloud Platform* (Plataforma de Nuvem do Google)
HTTP	*Hypertext Transfer Protocol*
ID	*Identifier* (Identificador)
INPI	Instituto Nacional de Propriedade Intelectual
IOC	*Indicator of Compromise* (Indicador de Compromisso)
IODEF	*Incident Object Description Exchange Format* (Formato de Troca da Descrição de Objetos de Incidente)
IOT	*Internet of things* (Internet das coisas)

IP	*Internet Protocol* (Protocolo de Internet)
MAEC	*Malware Attribute Enumeration and Characterization* (Enumeração e Caracterização de Atributos de Malware)
NIST	*National Institute of Standards and Technology* (Instituto Nacional de Padrões e Tecnologia)
RPC	Registro de Programa de Computador
SDO	*Stix Domains Objects* (Objetos de Domínios Stix)
SIEM	*Security information and event management* (Gerenciamento e Correlação de Eventos de Segurança)
SRO	*Stix Relationship Objects* (Objetos de relacionamento Stix)
STIX	*Structured Threat Information eXpression* (Expressão de Informações Estruturadas Sobre Ameaças)
TAXII	*Trusted Automated eXchange of Indicator Information* (Troca Automática Confiável de Indicadores)
TIC	Tecnologias da Informação e Comunicação
TTP	*Tactics, Techniques, and Procedures* (Táticas, técnicas e procedimentos)
URL	*Uniform Resource Locator* (Localizador Padrão de Recursos)
V2	Versão 2 (dois)
WWW	*world wide web* (Rede Mundial de Computadores)
XML	*Extensible Markup Language*

1 INTRODUÇÃO

Todos os dias, um grande número de pessoas, empresas e até mesmo nações, trocam a maneira de manipular suas informações, digitalizando e automatizando processos. Esse êxodo rumo a esfera tecnológica é realidade principalmente em países de economia emergente, enquanto que em países desenvolvidos essa abordagem já é considerada comum, mantendo-se em números elevados [1]. Uma das tecnologias chaves para essa taxa elevada de adoção é a rede mundial de computadores, ou Internet. O valor que sua utilização agrega nos mais diversos âmbitos da sociedade, é indiscutível. Porém, juntamente com os benefícios que seu emprego proporciona, também é precioso levar em consideração as suas debilidades. Uma das responsabilidades acarretadas pelo seu uso é a necessidade de gerenciar e proteger, os dados que trafegam pelo meio tecnológico utilizado em seu acesso.

Desde o inicio da Internet acontecem ações criminosas com todo tipo escuso de objetivo. E devido à sua adoção maciça e à dependência gerada com ela, o impacto dessas ações se tornam cada vez mais catastróficos. As cifras anuais de perdas causadas por crimes cibernéticos estão em torno de bilhões de dólares [2]. De olho nesse "mercado"emergente, o perfil dos criminosos também vem mudando, passando de indivíduos isolados, para grupos bem organizados e muitas vezes financiados, aumentando assim consideravelmente os incidentes no ambiento virtual [3].

Juntamente com a evolução do crime cibernético, porém do outro lado da moeda, surgiram empresas com objetivo especifico de atuar no ramo da segurança da informação. Essas empresas, em conjunto com diversas iniciativas individuais, acadêmicas e até governamentais, mapearam técnicas utilizadas pelos criminosos cibernético, e criaram "vacinas"e outros métodos de detecção e contramedidas [4], com o objetivo de suportar de forma segura o processo de desenvolvimento tecnológico. Porém, à medida com que a segurança avança, também avançam as técnicas utilizadas para transpassá-la.

Atualmente, soluções que antes eram suficientes para garantir um nível razoável de segurança no âmbito virtual, já não possuem a mesma efetividade. Por exemplo, soluções de segurança que utilizam assinatura como a base de seu modelo de detecção, não são mais eficazes contra ameaças consideradas avançadas [5], pois a forma como são atualizadas essas soluções é relativamente lenta comparada à velocidade com que são disseminadas novas ameaças.

Para contornar esse problema, foi proposta uma estratégia usando o principal meio usado pelos criminosos cibernéticos, a Internet. Baseado no fato de que a maioria dos softwares malintencionados são desenvolvidos a partir de dados extraídos da Internet (incluindo *deepweb* e *darkweb*) [6], a proposta consiste em utilizar essa mesma fonte para gerar conhecimento estratégico voltado para organizações. Essa estratégia, porém, deparou-se com o problema da quantidade imensa de dados hospedados dentro da Internet, e sendo incrementada a cada dia, impossibilitando análise manual.

A solução encontrada para o problema da quantidade de dados existentes, foi criar meios de automatizar a análise dos dados, surgindo ferramentas e técnicas de *Cyber Threat Intelligence* (inteligência de ameaças cibernéticas), ou apenas CTI, que têm como objetivo realizar a coleta de dados nos mais diversos locais, analisá-los, e apresentá-los de forma pragmática [7]. Essas ferramentas além de servir como base de dados de ameaças e/ou indicadores, permitem promover mecanismos para que possa ser gerada inteligência a partir desses dados, proporcionado uma visão holística de cenários e contextos [8].

Técnicas e aplicações de CTI dependem da qualidade das fontes de dados que elas consomem. Esses dados devem ser extraídos e analisados, a partir de fontes de dados que já passaram por esse processo, sendo a funcionalidade que vincula esses ciclos, o compartilhamento. Que desde o principio da Internet é uma forma bem-sucedida de minimizar os *gaps* de conhecimento [9].

Ferramentas de CTI têm o objetivo de realizar a extração de fontes de dados, analisá-las e gerar conhecimento operacional e estratégico, geralmente criadas por empresas com fins lucrativos [10]. O compartilhamento desse conhecimento gerado, é raramente encontrado nesse tipo de aplicação, e quando o é, ele é vedado somente a quem possuir ferramentas pertencentes à mesma empresa, ou empresas parceiras a essa. Ferramentas de inteligência de ameaças cibernéticas livres têm a mesma tendência das ferramentas das empresas privadas, compartilhando dados apenas com quem faz parte da comunidade, ou muitas vezes implementando limitações de acesso a qualquer um [11].

CTI é um assunto predominantemente tratado no âmbito do mercado [12], pois se trata de um tópico que já possui uma quantidade considerável de aplicações em produção, sendo vendidas e utilizadas. Considerando-se embasamento acadêmico sobre o tema, constata-se que as pesquisas e artigos, que tratam o assunto, voltam sua atenção para caraterísticas singulares desse contexto, como por exemplo a inserção de dados, o modo como gerar visibilidade, ou a necessidade de realizar o compartilhamento de resultados.

1.1 MOTIVAÇÃO

Devido ao seu uso potencial no enfrentamento de ameaças avançadas, inteligência de ameaças cibernéticas é um tema importante a ser entendido e disseminado. Porém constata-se a inexistência de um estruturara de referência para o desenvolvimento de aplicações, que incorpore características e funcionalidades inerentes ao assunto.

Sendo assim, este trabalho pretende contribuir com uma proposta de um modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas, visando responder as seguintes questões, entre outras:

I. Quais as funcionalidades inerentes ao contexto de inteligência de ameaças cibernéticas?

II. Em que abordagem elas fazem sentido?

III. Quais funcionalidades as ferramentas devem implementar?

IV. Como as diversas facilidades podem trabalhar juntas para conduzir o funcionamento geral da ferramenta?

Além disso, nesse trabalho foi desenvolvido uma aplicação demonstrativa que atua no âmbito de inteligência de ameaças cibernéticas visando respaldar o modelo proposto. O objetivo principal da aplicação é realizar o compartilhamento de dados utilizando protocolo aberto, e com restrições controladas apenas pelo usuário, fomentando assim o uso mais amplo desse recurso essencial na luta contra ameaças mais complexas.

1.2 OBJETIVOS

Esse trabalho foi idealizado com os seguintes objetivos:

Objetivo geral: Propor um modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas.

Objetivos específicos:

I. Utilizar modelo para planejar aplicação de CTI, visando evidenciar efetividade do modelo proposto;

II. Desenvolver aplicação demonstrativa;

III. Realizar testes das funcionalidades implementadas;

IV. Comparar aplicação desenvolvida com outras existentes.

1.3 METODOLOGIA

A metodologia utilizada consiste nos seguintes passos:

I. Realizar pesquisa bibliográfica buscando referências que descrevam a utilização, aplicabilidade, funcionamento e comportamento, de métodos e ferramentas de inteligência de ameaças cibernéticas.

II. Realizar testes em ferramentas e métodos de CTI.

III. Utilizar material coletado e conhecimento adquirido, para conceber uma proposta de modelo de referência voltado para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas.

IV. A partir do modelo proposto, realizar o planejamento e o desenvolvimento de aplicação demonstrativa de inteligência de ameaças cibernéticas.

V. Realizar experimentos com a aplicação demonstrativa e comparar com resultados de outras aplicações existentes.

1.4 PRINCIPAIS CONTRIBUIÇÕES

A seguir é apresentado as principais contribuições desse trabalho:

- Disponibilidade de um modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas, que introduz uma estrutura que unifica caraterísticas e funcionalidades inerentes ao contexto, objetivando seu entendimento contextualizado de forma holística.
- Publicação do artigo "Proposta de modelo de referência de inteligência de ameaças"(51), na 17º Conferencia Ibero Americana (CIAWI), que ocorreu em Lisboa, Portugal, em 2019.
- Disponibilidade de aplicação de inteligência de ameaças cibernéticas demonstrativa, objetivando fomentar o compartilhamento livre de dados.
- Registro de Programa de Computador (RPC) no Instituto Nacional de Propriedade Intelectual (INPI), com o número do processo "512020000871-2".

1.5 ORGANIZAÇÃO

O trabalho a seguir está dividido em 5 capítulos, sendo a introdução o capítulo 1. O capítulo 2 traz o embasamento teórico e trabalhos correlatos, enquanto que o capítulo 3 apresenta a proposta de modelo de referência para desenvolvimento de aplicação de inteligência de ameaças cibernéticas, bem como a estrutura da aplicação desenvolvida como prova de conceito. O Capítulo 4 apresenta os resultados obtidos com o uso da aplicação de inteligência de ameaças cibernéticas implementada, na forma de telas e agregadores de indicadores, e comparando-os com resultados de outras ferramentas. E por fim, o Capítulo 5 apresenta conclusões desse trabalho e apresenta trabalhos futuros.

2 REFERENCIAL TEÓRICO E TRABALHOS CORRELATOS

Nos dias de hoje, pode-se observar que o mundo vem se tornando cada vez mais interconectado, tanto economicamente quanto socialmente. A adoção de tecnologias, como a Internet, é um dos pilares desse crescimento, que tem como resultado o auxilio ao progresso humano. Um estudo [1] com fins voltados a monitorar essa adoção, identificou o crescimento expressivo do percentual de utilização desse tipo de tecnologia, principalmente em países com a economia emergente, enquanto que em países desenvolvidos, esse padrão de utilização já está estabilizado num alto patamar.

A rede mundial de computadores (Internet) teve sua origem nos Estados Unidos da América no ano de 1969, com o nome de "Arpanet"[13]. Ela foi criada com o objetivo de interligar laboratórios de pesquisa, durante o período da "guerra fria". A sua utilização foi expandida a partir de 1982, quando passou a ser adotada dentro do âmbito acadêmico nos EUA e posteriormente em países como Holanda, Dinamarca e Suécia, passando a ser conhecida como Internet. A partir de 1992, empresas e outras organizações começaram a investir na sua ampliação e difusão, juntamente com a invenção da *World Wide Web* (www) pelo Laboratório Europeu de Física de Partículas (CERN), dando origem a Internet "comercial", resultando desde então em seu emprego de forma maciça.

O uso generalizado desse tipo de tecnologia engloba diversos âmbitos da sociedade, tendendo a gerar mais eficiência e eficácia, no desempenho de suas atividades. Porém esse uso generalizado causa dependência, pois uma vez automatizados os processos, é necessário adaptar técnicas e metodologias [14]. Assim sendo, a indisponibilidade total ou parcial, dessa tecnologia, pode ser nociva a seus usuários. Estima-se, por exemplo, que os prejuízos causados pelo criminosos cibernéticos, que afeta diretamente a utilização desses recursos, esteja na casa dos bilhões de dólares todos os anos [2].

2.1 CRIMES CIBERNÉTICOS

Crimes cibernéticos acontecem dentro do ciberespaço desde a criação da rede mundial de computadores, e vêm crescendo junto com sua popularização. Em constante expansão e desenvolvimento, novas tecnologias regularmente são incorporadas à Internet, e para suporta-las, sua estrutura se torna cada vez mais complexa. Essa complexidade vem acarretando na distribuição de sua estrutura entre provedores de serviços e seus usuários, abrindo (ainda mais) brechas [15], pela necessidade dessa coparticipação.

Junto com o incorporeamente de novas tecnologias à Internet, novos recursos são gerados. E

visando sua utilização com foco em seus benefícios, sua adoção vem crescendo em seguimentos vitais a pessoas, empresas e até nações. Acarretando na evolução dos ataques praticados, saindo de pequenos e solitários agentes em seus "porões", para um negócio diversificado, segmentado e altamente profissionalizado [16], que gera bilhões de dólares.

O crime cibernético é algo tão prejudicial na atualidade, que foram surgindo empresas que tem foco exclusivo de seu negocio voltado a auxiliar a proteção cibernética de outrem, e até mesmo nações possuem departamentos estratégicos nesse foco. Porém, apesar da conscientização e esforços, nenhuma proteção é definitiva, ao passo que as proteções evoluem os ataques também o fazem.

2.1.1 Ciclo de vida de ataque comum

No cenário atual do combate ao crime cibernético, existe uma vasta documentação de ataques categorizados cujos comportamentos foram mapeados por órgãos/empresas que atuam nesse meio. Porém os ataques são modificados constantemente, e são reinventados com mais velocidade do que a capacidade atual de análise e tipificação. Um exemplo bem sucedido de um ataque, foi denominado como *Dyre Banking* [17], e teve bancos como alvos. Veja detalhes de seu ciclo de vida na Tabela 2.1.

Tabela 2.1: ciclo de vida do ataque cibernético *Dyre Banking*

Data	Descrição
Fev. 2015	Microsoft identifica uma vulnerabilidade (MS15- 010/CVE 2015-0057) que permite a execução de código remotamente no *Windows*. Até o momento não existia nenhum *exploit* conhecido para essa vulnerabilidade.
Abr. 2015	Um *exploit* (MS15-010/CVE 2015-0057) que fazia uso dessa vulnerabilidade, foi encontrado a venda no *darknet market* por 48 BTC (cerca de $10,000-15,000).
Jul. 2015	A FireEye identificou que o Trojan Dyre Banking, desenhado para roubar dados de cartões de crédito, fazia uso dessa vulnerabilidade.

Baseado nas informações contidas na Tabela 2.1, observa-se que em fevereiro de 2015, a *Microsoft* identificou uma falha de segurança que afetava o sistema operacional *Windows*. Essa falha permitia a execução remota de código malicioso e, até o momento da divulgação, não haviam sido encontrados softwares que explorassem tal vulnerabilidade. Dois meses depois, em abril de 2015, um software malicioso que se aproveitava dessa vulnerabilidade foi localizado sendo vendido no mercado negro da *darknet* (por cerca de U$10.000). Algum tempo depois, em julho de 2015, a empresa de segurança *FireEye* identificou um trojan (tipo de programa malicioso), que foi batizado de *Dyre Banking*. Ele era capaz de roubar números de cartões de crédito, explorando a vulnerabilidade descoberta pela *Microsoft* 5 meses atrás. A exposição global média desse trojan foi de 57,3%, ou seja, quase 6 em 10 organizações no mundo foram afetadas. Segundo o relatório da *Blueliv* [18] voltado para o *Dyre Banking*, foram infectados organizações em mais de 100

países, com um prejuízo incalculável.

Analisando-se a estrutura disposta na Tabela 2.1, observa-se um padrão translúcido, onde os criminosos cibernéticos estão se aproveitando-se de vulnerabilidades conhecidas (divulgadas publicamente), objetivando prejudicar infraestruturas vitais para o seguimento das funções desempenhadas na esfera tecnológica. Em menos de 3 meses foi possível identificar uma vulnerabilidade com grande potencial de exploração, desenvolver um software, e por fim explorar remotamente (e com sucesso) a falha exposta.

2.1.2 Ataques avançados

A técnica utilizada para executar o *Dyre Banking* foi a criação de um Trojan, que é um tipo de programa malicioso que se disfarça de um programa legitimo, para causar prejuízos [19]. Como já evidenciado, ele pode ser nefasto e causar grande prejuízos, porém no cenário atual do cibercrime, existem técnicas ainda mais nocivas sendo empregadas. Um exemplo, e que merece o destaque, é o *Advanced Persistent Threats* (APT), que vem chamando muita atenção da comunidade e da indústria da segurança da informação [20]. Segundo a *National Institute of Standards and Technology* (NIST), o APT pode ser definido como [21]: "Um adversário que possui níveis sofisticados de especialização e recursos significativos que permitem criar oportunidades para atingir seus objetivos usando vários vetores de ataque (como cibernético, físico e engenharia social). Esses objetivos geralmente incluem estabelecer e estender pontos de apoio na infraestrutura de tecnologia da informação das organizações-alvo para fins de filtrar informações, minar ou impedir aspectos críticos de uma missão, programa ou organização; ou se posicionar para realizar esses objetivos no futuro. É uma ameaça persistente avançada: (i) persegue seus objetivos repetidamente durante um período prolongado de tempo; (ii) adapta-se aos esforços dos defensores para resistir; e (iii) está determinado a manter o nível de interação necessário para executar seus objetivos".

Ataques como o APT normalmente são realizados por grupos bem organizados, focados e muitas vezes financiados. São extremamente complicados de serem parados, pois fazem uso de diversas técnicas para atingir seu objetivo. Estendendo-as e alternando-as segundo estratégia e abordagem escolhidas, para realizar a invasão.

Segundo estudo realizado pela *Positive Technologies* [22], no mercado negro é possível se encontrar diversos tipos de APT à venda, variando o preço dependendo do nível de complexidade, e do tempo de serviço. Por exemplo, a criação de um ataque simples como *phishing*, custa em torno de 300 dólares americanos, porém quando exige estudo de ambiente e criação de software direcionado, o valor sobe para algo em torno de 10.000 dólares. Outras técnicas que exigem maior conhecimento técnico, como a criação de um *zero-day* (software criado sem uso de outros códigos conhecidos, e direcionados para um fim especifico [23]), podem chegar a custar um milhão de dólares. Um exemplo bem-sucedido de ataque, o APT38, levou 155 dias até que o agente malicioso conseguisse realizar a intrusão. Foram aplicados 26 tipos diferentes de técnicas, e o preço estimado do ataque excede meio milhão de dólares. O prejuízo estimado do dano

causado por esse APT foi de 41 milhões de dólares.

2.2 TÉCNICAS PARA CONTENÇÃO DE AMEAÇAS

Para interromper, ou frear ataques avançados, é preciso realizar um processo constante de fechamento de brechas e aplicação de estratégias de diminuição de impacto de ataques, até o ponto em que fique mais caro continuar com a sua execução, do que o atacante pretende arcar [24]. Inicialmente os responsáveis por orquestrar a defensa devem buscar entender o fluxo de funcionamento desse tipo de ataque, juntamente com as principais técnicas que eles empregam [25].

Uma das organizações a frente da criação de iniciativas que buscam realizar análises de ameaças e documentá-las para o público em geral, é a *The Mitre Coporation*[1], organização sem fins lucrativos, que trabalha junto com o governo dos EUA, e tem sedes em vários estados norte-americanos. Atualmente é uma das maiores referências em se tratando de crimes cibernéticos, sendo autora de várias iniciativas bem-sucedidas nesse campo de conhecimento [26].

2.2.1 Kill Chain

Objetivando lidar com ameaças de alto nível (como APTs), a Mitre foi autora de uma iniciativa que organiza ataques em fases, buscando uma visão holística do processo de intrusão, consolidando sua estrutura em etapas necessárias até atingir o objetivo. Essa iniciativa gerou um modelo de cadeia de análises, a *Kill Chain* [27], apresentado na Figura 2.1:

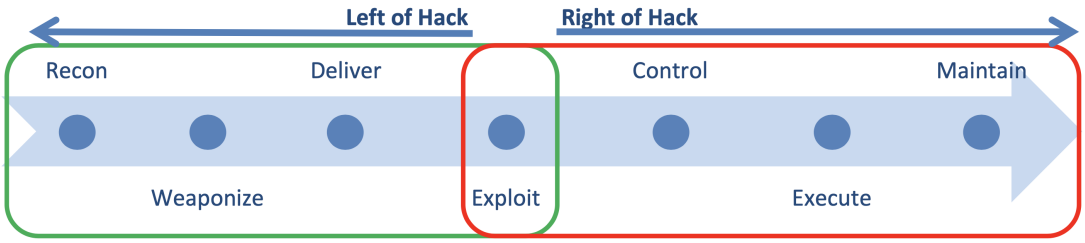

Figura 2.1: Modelo de cadeia de análises *Kill Chain* da Mitre

Como é possível se observar na Figura 2.1, a estrutura desse modelo de cadeia de análises é dividido em duas partes, um lado esquerdo (*left of hack*) que acontece antes do procedimento de intrusão (*exploit*) e o lado direito (*right of hack*), que acontece após a intrusão. A tabela 2.3 apresenta a descrição de cada uma das etapas da *Kill Chain*:

[1]<https://www.mitre.org/>

Tabela 2.2: Ciclo de vida de ataque

Fase	Descrição
Recon (Reconhecimento)	O adversário identifica e investiga o alvo. Como por exemplo, mineração de dados em *websites* e lista de participantes em conferencias online.
Weaponize (Armamento)	O conjunto de ferramentas de ataque é empacotado para entrega e execução no computador/rede da vítima. Como por exemplo, o adversário cria um arquivo PDF "trojanizado"contendo suas ferramentas de ataque.
Deliver (Entrega)	A ferramenta ou ferramentas de ataque empacotadas são entregues ao(s) alvo(s). Como por exemplo, o adversário envia um e-mail com *spearphishing* contendo o arquivo PDF "trojanizado"para sua lista de alvos.
Exploit (Intrusão)	O ataque inicial ao alvo é executado. Como por exemplo, o usuário alvo abre o arquivo PDF malicioso e o *malware* é executado.
Control (Controle)	O adversário começa a direcionar o(s) sistema(s) da vítima para tomar ações. Como por exemplo, o adversário instala ferramentas adicionais no(s) sistema(s) da vítima.
Execute (Execução)	O adversário começa a cumprir os requisitos de sua missão. Como por exemplo, O adversário começa a obter os dados desejados, geralmente usando o sistema da vítima como ponto de partida para obter acesso interno e à rede ao sistema interno.
Maintain (Manutenção)	Acesso a longo prazo é alcançado. Como por exemplo, o adversário estabeleceu *backdoors* ocultos na rede de destino para permitir a reentrada regular.

O modelo *Kill Chain* é utilizado para auxiliar no processo de análise de (possíveis) intrusões, buscando clarificar as etapas que os (possíveis) invasores já percorreram e/ou em qual delas estão. Cada uma das etapas analisadas pela *Kill Chain* tem suas próprias características, abordagens e metodologias de consolidação [28]. Sendo assim, para analisar cada uma delas, também é necessária a utilização de técnicas especificas visando o reconhecimento de padrões e indícios, proveniente de técnicas utilizados no (possível) ataque.

2.2.2 Framework ATT&CK

Para complementar o processo de análise juntamente com a *Kill Chain*, a Mitre foi autora do framework "*Adversarial Tactics, Techniques, and Common Knowledge*"(ATT&CK)[2], uma base de conhecimento em modelo de curadoria, para o comportamento de adversários cibernéticos. A base do *framework ATT&CK* contém a descrição de *Tactics, Techniques and Procedures* (Táticas, Técnicas e Procedimentos), ou apenas TTP, que se enquadram em cada uma das etapas apresentadas na *Kill Chain*, trazendo abordagens confiáveis para auxiliar nas verificações das tecnologias implementadas na infraestrutura a ser protegida. Os principais tópicos que compõem a sua estrutura, juntamente com alguns exemplos de técnicas, são apresentados pela Tabela 2.3:

[2]<https://attack.mitre.org/>

Tabela 2.3: Técnicas descritas na base do *framework ATT&CK*

Tópico	Exemplos de técnicas
Initial access	*drive-by compromise*, *exploit public-facing application* e *external remote services*
Execution	*appleScript*, *CMSTP* e *command-Line Interface*
Persistence	*bash_profile and bashrc*, *accessibility features*, *account manipulation*
Privilege escalation	*access token manipulation*, *accessibility features*, *appCert DLLs*
Defense evasion	*access token manipulation*, *binary padding*, *bits jobs*
Credential access	*account manipulation*, *bash history*, *brute force*
Discovery	*remote system discovery*, *account discovery*, *application window discovery*
Lateral movement	*appleScript*, *application deployment Software*, *application deployment software*
Collection	*audio capture*, *automated collection*, *clipboard data*
Command and control	*commonly used port*, *communication through removable media*, *connection proxy*
Exfiltration	*automated exfiltration*, *data compressed*, *data encrypted*
Impact	*account access removal*, *data destruction*, *data encrypted for impact*

As técnicas apresentadas na Tabela 2.3, juntamente com muitas outras, são todas descritas pelo *framework ATT&CK*. Essa descrição consiste em detalhamento de vulnerabilidades, técnicas para consumar ataques e os softwares desenvolvidos para realizar a intrusão (quando for o caso). É possível portanto basear-se em indícios e artefatos mapeados, para identificar TTP's que estão sendo utilizadas por agentes maliciosos, e a partir dessas TTP's é possível identificar a fase do ataque, utilizando a *Kill Chain*. Isso permite posicionar-se em relação à defesa da infraestrutura, elaborando uma visão holística do cenário, baseada nos indícios e artefatos mapeados, objetivando a tomada de decisões.

2.2.3 Utilização prática de técnicas para contenção de ameaças

Para exemplificar o funcionamento da identificação do posicionamento do atacante dentro da *Kill Chain*, pode-se destacar uma das técnicas mapeadas pelo *framework ATT&CK*, pertencente no tópico de *Discovery*, a "*remote system discovery*"[3]. Veja a seguir o detalhamento dessa técnica:

- Descrição: Os adversários provavelmente tentarão obter uma lista de outros sistemas por endereço IP, nome do *host* ou outro identificador lógico em uma rede que possa ser usada para o Movimento Lateral (*Lateral Movement*) no sistema atual. A funcionalidade pode existir dentro das ferramentas de acesso remoto para permitir isso, mas os utilitários disponíveis no sistema operacional também podem ser usados. Os adversários também podem

[3]<https://attack.mitre.org/techniques/T1018/>

usar arquivos de *host* local para descobrir os mapeamentos de nome dehost para endereço IP de sistemas remotos.

- Plataformas: *Linux*, *macOS*, *Windows*, *GCP*, *Azure*, *AWS*
- Mitigações: Esse tipo de técnica de ataque não pode ser facilmente mitigado com controles preventivos, pois se baseia no abuso de recursos do sistema.
- Detecção: As técnicas de descoberta de sistemas e redes normalmente ocorrem durante uma operação, à medida que o adversário aprende o ambiente. Dados e eventos não devem ser vistos isoladamente, mas como parte de uma cadeia de comportamento que pode levar a outras atividades, como o Movimento Lateral (*Lateral Movement*), com base nas informações obtidas.

Após aplicação prática da "remote system discovery", e uma vez detectada a ocorrência do ataque, é possível identificar em qual fase ele se encontra, utilizando a *Kill Chain*. No caso dessa técnica, o ataque estaria na fase de *Recognition* (Reconhecimento). Uma vez com as evidencias analisadas e as características dos (possíveis) atacantes, como por exemplo endereço IP que originou ataque, as ferramentas utilizadas e/ou brecha explorada, é possível identificar a situação real da infraestrutura (baseado em dados e métodos) e realizar as tomadas de decisões, visando frear o ataque, ou minimizar os seus prejuízos.

Em suma, para aumentar a segurança da infraestrutura vigente, é necessário gerar dados estratégicos investigativos utilizando indícios das ações realizadas por agentes mal-intencionados, evidenciando-os com bases de conhecimento e projetando-os em modelos, que auxiliam na compreensão holística do cenário corrente.

Essa abordagem, embora comum e corrente, possui uma série de falhas. Primeiramente a infraestrutura precisa ser alvo de um ataque, para somente então, ser possível a reunião de dados estratégicos, para que se possam ser tomadas decisões visando sua melhoria. E depois, após a tomada decisão, não existe prospecção alguma da utilização desses dados, como parte de um planejamento futuro, em que eles também podem ser utilizados. Além disso, a consciência situacional gerada apenas indica brechas em ativos que foram atacados e a equipe de segurança teve êxito em evidenciar, podendo existir brechas em outros ativos. Por fim, dificilmente seria possível compartilhar a experiência obtida com esse ataque, com interessados, que por sua vez poderiam apoiar-se nelas para realizar o processo de tomada de decisão.

2.3 INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

Tendo em vista o grande número de tecnologias necessárias para o funcionamento de uma infraestrutura cibernética, juntamente com a propagação em iguais proporções de ameaças e brechas, é impossível protege-la de forma isolada, realizando manualmente todas as análises requeridas. Se faz necessário uma solução que busque dados estratégicos (brechas, TTP's, IPs, etc), de

meios similares aos dos atacantes, realizar uma análise sobre esses dados e gerar conhecimento estratégico de forma proativa e automatizada, para que a sua defesa ande em compasso com a capacidade criativa dos atacantes.

Nesse sentido, foi estruturado um campo de conhecimento nomeado como CTI (Inteligência de ameaças cibernéticas) e definido por Rob McMillan [29], como o conhecimento baseado em evidências, que incluem o contexto, mecanismos, indicadores, implicações e aconselhamento sobre ameaças existentes ou emergentes que podem trazer danos a ativos. Nesse trabalho adota-se essa definição de CTI.

2.3.1 Necessidade de sua utilização

As técnicas e soluções de CTI, são implementadas visando oferecer os mecanismos necessários para suprir as seguintes necessidades:

- Armazenamento e gerenciamento de dados de ameaças e brechas;
- Manipulação de dados na tratativa de gerar conhecimento;
- Prover visibilidade adequada sobre os dados armazenados;
- Devido a grande quantidade de ameaças, poder consumir dados de análises realizadas por outrem.

2.3.2 Contextualização sobre dados, informação e conhecimento

Antes de entrar em mais detalhes sobre CTI, se faz necessário realizar uma definição de dados, informação e conhecimento (ou inteligência), na visão deste trabalho. Esses três termos são usados algumas vezes sem muito cuidado. Por exemplo, algumas fontes de ameaças são anunciadas como inteligência quando na verdade são apenas pacotes de dados. A Tabela 2.4 apresenta a definição adotada neste trabalho para os termos dados, informação e conhecimento/inteligência [30]:

Tabela 2.4: Dados x Informação x Conhecimento/inteligência

Tipo	Descrição	Cibersegurança
Dado	Consiste em fatos e estatísticas discretas, reunidos como base para uma análise mais aprofundada.	Os dados geralmente são apenas indicadores, como endereços IP, URL ou hash. Os dados não nos dizem muito sem análise.
Informação	São vários pontos ou conjunto de dados combinados para responder a perguntas específicas.	As informações respondem a perguntas como: "Quantas vezes minha organização foi mencionada nas mídias sociais este mês?"Embora seja uma saída muito mais útil do que os dados brutos, ainda não informa diretamente uma ação específica.
Conhecimento	Analisa dados e informações para descobrir padrões e histórias que orientam a tomada de decisão.	O conhecimento é o produto de um ciclo de identificação de perguntas e objetivos, coleta de dados relevantes, processamento e análise desses dados, resultando na produção de inteligência que proporciona ações e compartilhamento.

2.3.3 Papel de ferramentas de CTI

As ferramentas e técnicas de CTI são facilitadoras para a gestão de conhecimento de ameaças. Elas estão munidas com a possibilidade de se realizar a inserção de dados provenientes de várias fontes (e.g., análises como a descrita usando *Kill Chain* e o *framework ATT&CK*), tanto aquelas que foram criadas a partir da própria gestão de segurança da organização que a aplica, quanto aquelas criadas por outros. São capazes de realizar a sua análise e enriquecimento dos dados e informações, buscando pontos comuns e de interesse da organização que a estuda, visando responder perguntas e gerar conhecimento operacional e estratégico a partir dessa análise. E por fim, as ferramentas e técnicas de CTI devem permitir o compartilhamento de suas análises com interessados, visando um repasse proveitoso de experiência.

A inteligência de ameaças cibernéticos é um conhecimento que permite impedir e mitigar ataques a sistemas digitais. Com base nos dados obtidos, a CTI fornece contextos que podem auxiliar na descoberta de conhecimento crítico para o combate ao ciber criminosos, como por exemplo quem é o atacante, quais são as motivações e capacidades, e quais são os indicadores de comprometimento (IOCs) da infraestrutura atacada. Esse tipo de conhecimento permite avaliar o cenário atual do negocio e serve como base na tomada decisões.

2.3.4 Principais beneficiados

Basicamente todos os membros de uma organização podem ser beneficiados com o uso de CTI, veja alguns exemplos na Tabela 2.5:

Tabela 2.5: Beneficiados com o uso de inteligência de ameaças cibernéticas

Beneficiado	Situação
As equipes de operações de segurança	São rotineiramente incapazes de processar o fluxo avassalador de alertas que recebem. A CTI pode ser integrada às soluções de segurança que eles já usam, ajudando-os a priorizar e filtrar automaticamente alertas e outras ameaças.
As equipes de gerenciamento de vulnerabilidades	Precisam priorizar com precisão as vulnerabilidades mais importantes. A CTI fornece acesso a *insights* e contextos externos que os ajudam a diferenciar ameaças imediatas à suas empresas específicas das ameaças meramente potenciais.
Equipes de prevenção de fraudes, de análise de riscos e outras equipes de segurança de alto nível	São desafiadas a entender o atual cenário de ameaças. A CTI fornece informações importantes sobre os atores de ameaças, suas intenções e alvos e suas táticas, técnicas e procedimentos (TTP's).

2.3.5 Abordagem pragmática sobre a utilização de CTI

Atualmente existem ferramentas e técnicas desenvolvidas especificamente para atuar no campo da CTI. Elas buscam proporcionar uma capacidade de gerar/manipular conhecimento estratégico, voltado para formar entendimento holístico da ameaça representada pelo adversário, permitindo apoiar de forma mais eficaz a tomada de decisão e a priorização de cursos de ação. Proporcionam também uma potencial oportunidade de afetar fundamentalmente o equilíbrio de poder entre o defensor e o adversário, pondo a inteligência da organização ombro a ombro com agentes maliciosos, possibilitando a aquisição de experiência de outrem, para apoiar no combate a ameaças dentro da própria organização.

Segundo Hutchins, Cloppert, e Amin [31], O efeito da CND (defesa de rede de computadores) baseada em inteligência é uma postura de segurança mais resiliente. Os atores do APT, por sua natureza, realizam tentativas de invasão sucessivas vezes, ajustando suas operações com base no sucesso ou falha de cada tentativa. Em um modelo de cadeia de análises, como a *Kill Chain*, apenas uma mitigação quebra a cadeia e frustra o adversário; portanto, qualquer repetição do adversário é uma responsabilidade que os defensores devem reconhecer e alavancar. Através de análises baseado em CTI, os defensores podem desenvolver mitigações resilientes contra intrusos e priorizar de forma inteligente os investimentos em novas tecnologias ou processos. Se os defensores implementam contramedidas mais rapidamente do que os adversários evoluem, isso aumenta os custos que um adversário deve gastar para alcançar seus objetivos. Mostrando que com a utilização de ferramentas de CTI, ao contrário do que se pensa, que esses agressores não

têm vantagem inerente sobre os defensores.

2.3.6 Necessidade de compartilhamento no âmbito de inteligência de ameaças cibernéticas

A implementação de ferramentas/métodos de CTI dentro das organizações é um desafio, pois nenhuma organização, por si só, possui a capacidade de gerar conhecimento significativo em todos os escopos relevantes. Uma maneira de superar essa limitação é através do compartilhamento de informações relevantes sobre ameaças cibernéticas entre parceiros e comunidades confiáveis.

O compartilhamento de informações sobre ameaças pode ser utilizado de forma eficiente, estratégica e eficaz contra ameaças emergentes. Malik [9], afirma que O compartilhamento de dados voltados a CTI é a única maneira de combater o crescente *gap* de habilidades. Na prática o compartilhamento de dados e informações sobre ameaças é usado para comunicar experiência operacional, visando com que os participantes possam aprimorar suas defesas contra-ataques contínuos, criando abordagens proativas nesse contexto.

Dentro desse contexto, as ferramentas de CTI têm que enfrentar o desafio de lidar com diversos tipos de fontes de informações sobre ameaças, visando absorver conhecimento pré-gerado. Juntamente com a necessidade de padronizar o conhecimento absorvido, em formato compreensível e compartilhável.

2.4 FONTES DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

Dentro do contexto de CTI, o nome atribuído a locais em que são extraídos conteúdo utilizados para munir as ferramentas, é amplamente conhecido como "*Feed*". Nesse contexto não existe realmente um padrão de dados, qualquer fonte pode conter dados, que dependendo do contexto e da abordagem, podem gerar valor investigativo. Alguns *feeds*, inclusive, são consideradas não usuais em outros contextos, como por exemplo, blogs, fóruns, e informações dispostas em domínios pertencentes a estrutura da *darknet*.

Dentro dos *feeds* estudados nesse trabalho, foi encontrada uma divisão em duas categorias, "privados"e "abertos". Os *feeds* "privados"possuem autoria e em sua maioria, objetivos lucrativos. Por exemplo, empresas como McAfee[4] e Symantec[5] constroem bases de dados de CTI, e vendem seu conteúdo (ou parte dele).

Feeds "abertos"são dados que podem ser acessados de forma gratuita e que possuem autoria. Sempre é interessante verificar a confiabilidade desse tipo de fonte, pois existem ataques que disponibilizam falsos positivos, visando invalidar bases de dados inteiras, devido a partes não confiáveis. Como exemplos de *feeds* "abertos", tem-se relatórios de pesquisadores de segurança

[4] <https://www.mcafee.com/enterprise/pt-br/threat-center/global-threat-intelligence-technology.html>
[5] <https://www.symantec.com/services/cyber-security-services/>

da informação (e afins), blogs de fornecedores de tecnologia, listas de bloqueio, reputação e URLs (disponíveis publicamente), e órgãos públicos que visam seguranças de Estados (nações). Além disso, existem algumas empresas que oferecem *feeds* de forma gratuita, como Anomali[6], Hailataxii[7] e Alienvault[8].

Existe também possibilidade de se criar os próprios *feeds*, realizando a inserção manual de determinado objeto dentro da base de dados pertencente à ferramenta de CTI, ou utilizando dados criados por ativos pertencentes à organização, visando gerar dados de forma automática. Soluções como SIEM (*Security information and event managment*) para gerenciamento e correlação de eventos de segurança [32] e *sandbox* que replicam ambientes e executam softwares visando sua análise de forma local [33], podem fomentar ferramentas de CTI, de forma customizada.

2.5 PADRÕES DESCRITIVOS E SUAS PRINCIPAIS CARACTERÍSTICAS

Dados são consumidos de *feeds* de CTI com objetivo de serem incorporados dentro de um escopo criado por organizações, empresas, pesquisadores, etc. A fim de gerar conhecimento operacional e estratégico, para ser usado para diversos fins. A seguir veja alguns exemplos de dados que podem encontrados em *feeds*:

- Observáveis existentes no espaço cibernético
- Indicadores
- Incidentes
- Táticas, técnicas e procedimentos adversos (incluindo padrões de ataque, malware, explorações, ferramentas infraestrutura, vítimas sendo almejadas, etc.)
- Alvos explorados (por exemplo, vulnerabilidades, pontos fracos ou configurações)
- Cursos de ação (por exemplo, resposta a incidentes ou soluções ou mitigações de vulnerabilidades/fraquezas)
- Campanhas de ataques cibernéticos
- Atores de ameaças cibernéticas

Esse tipo de dado é disponibilizado em diversos formatos, não seguindo padrão algum. Os responsáveis por gera-los, normalmente disponibilizam eles da maneira que lhes convém, pois pensam somente na relevância de seu conteúdo, deixando de lado a necessidade de sua aplicabilidade em conjunto. Para que o processo de gerar conhecimento seja possível, se faz necessária

[6] <https://www.anomali.com/>
[7] <http://hailataxii.com/>
[8] <https://www.alienvault.com/>

a normalização desses dados em formato padrão unificado. Pois somente assim, é possível realizar o processo de análise nos dados armazenados com os mesmos mecanismos e utilizando as mesmas técnicas [34].

O processo de análise de dados é ainda mais valioso junto com o processo de compartilhamento de dados, pois realizar o processo de compartilhamento de dados é a forma mais significava de troca de experiência, acarretando em nivelamento de *gaps* de conhecimento. Porém, devido à falta de padronização, o compartilhamento de dados em diferentes formatos, é difícil de ser implementado e utilizado.

Mesmo sem os padrões necessários, os processos de análise e compartilhamento de dados, sempre foram realizados, principalmente no âmbito da industria de segurança cibernética. Nesse sentido foram criados alguns padrões de disposição de dados, mas direcionados para ramos específicos dentro do ciberespaço. Por exemplo, o OpenIOC[9], é um *framework* criado com base no padrão XML, visando a documentação de características de táticas para identificar ameaças e metologias de ataques. O OpenIOC utiliza o formato IODEF (*Incident object description exchange format*) [35] para realizar o compartilhamento dados. Outro exemplo de *framework* criado com esse objetivo, é o CIF (*Collective Intelligence Framework*)[10], uma plataforma aberta criada para para o compartilhamento de dados de CTI. O CIF também utiliza o IODEF e seu principal foco é o registro de dados para identificação de atacantes, como por exemplo IP, domínio, URL, etc.

Visando a idealização de um padrão que consolida as áreas necessárias do ciberespaço, para a documentação de forma satisfatória de todos os itens necessários para a análise e compartilhamento de dados voltados para a CTI, foram criados pela MITRE os padrões STIX (*Structured Threat Information eXpression*) [36]) e TAXII (*Trusted Automated eXchange of Indicator Information*)[37] .

2.5.1 STIX

Focando no padrão atual de criação de dados construídos a partir de arquiteturas homogêneas, mas provenientes de um conjunto diversificado de produtos e sistemas diferentes. A organização MITRE foi a criadora de uma iniciativa, aberta e colaborativa, de um padrão de disposição de informações de CTI. O STIX (*Structured Threat Information eXpression*) [36], um padrão de disposição de informações de CTI cuja estrutura funciona como uma linguagem para padronizar especificações, capturas, caracterizações e a comunicação de eventos.

O padrão STIX foi criado de forma aberta e colaborativa com a comunidade de segurança da informação. No processo de sua criação foram separadas diversas características comumente expostas em relatórios. Essas características foram conceituadas e separadas em forma de objetos, que foram utilizados para formular um formato compreensível textualmente e visivelmente, gerando uma cadeia de objetos.

[9] <www.openioc.org>
[10] <https://csirtgadgets.com/collective-intelligence-framework>

A estrutura do STIX foi desenhada para aceitar uma grande quantidade de disposições diferentes de dados, sendo totalmente flexível e customizável, possuindo diversos tipos diferentes de objetos. O STIX é um padrão que não é focado em apenas uma classe de objetos dentro da CTI, sendo flexível o bastante para armazenar qualquer tipo possível objeto dentro desse contexto (com base nas abordagens atuais).

O padrão STIX descrito no artigo original [36] foi reestruturado durante sua aplicação pratica no decorrer dos anos, evoluindo para da versão 1, depois para a versão 1.1, e finalmente para a versão 2, a versão atual. Um dos locais em que essa versão foi documentada foi o na OASIS[11], sendo esse padrão dividido em dois tipos de objetos, o SDO (*Stix Domains Objects*) e o SRO (*Stix Relationship Objects*). Os SDO são objetos utilizados para documentar dados pertencentes ao contexto de CTI. A Tabela 2.6 apresenta os objetos pertencentes ao grupo dos SDO.

Tabela 2.6: Objetos SDO no padrão STIX v2

Stix - SDO	
Nome do SDO	Descrição
Attack Pattern	Um tipo de tática, técnica e procedimento (TTP) que descreve as maneiras pelas quais os agentes de ameaças tentam comprometer os alvos.
Campaing	Um agrupamento de comportamentos suspeitos que descreve um conjunto de atividades ou atividades mal-intencionadas que ocorrem durante um período de tempo contra um conjunto específico de destinos.
Course of Action	Uma ação tomada para impedir ou responder a um ataque.
Identity	Indivíduos, organizações ou grupos, bem como classes de indivíduos, organizações ou grupos.
Indicator	Contém um padrão que pode ser usado para detectar atividades cibernéticas suspeitas ou maliciosas.
Intrusion Set	Um conjunto agrupado de comportamentos e recursos suspeitos, com propriedades comuns que acredita-se serem orquestradas por um único ator de ameaça.
Malware	Um tipo de TTP, também conhecido como código malicioso e software malicioso, usado para comprometer a confidencialidade, integridade ou disponibilidade dos dados ou sistema de uma vítima.
Observed Data	Transmite informações observadas em um sistema ou rede (por exemplo, um endereço IP).
Report	Coleções de CTI focadas em um ou mais tópicos, como uma descrição de um agente de ameaças, malware ou técnica de ataque, incluindo detalhes contextuais.
Threat Actor	Indivíduos, grupos ou organizações que acredita-se estarem operando com intenção maliciosa.
Tool	Software legítimo que pode ser usado por atores de ameaças para realizar ataques.
Vulnerability	Um erro no software que pode ser usado diretamente por um hacker para obter acesso a um sistema ou rede.

[11] <https://oasis-open.github.io/cti-documentation/stix/compare>

Os SRO são utilizados para aclarar relacionamentos existentes entre os SDO. A Tabela 2.7 apresenta os objetos pertencentes ao grupo dos SRO.

Tabela 2.7: Objetos SRO no padrão STIX v2

Stix - SRO	
Nome do SRO	Descrição
Relationship	Usado para vincular dois SDOs e descrever como eles se relacionam.
Sighting	Indica a crença de que um elemento de CTI foi visto (por exemplo, indicador, malware).

Em cada objeto disposto do padrão STIX v2, estão incluídas também as *tags* do "criador"desse objeto, bem como uma estrutura flexível de marcação de dados baseada em ID's, para fornecer um contexto útil e facilitar a implementação das restrições de manipulação de informações. O campo "criador", incluído em cada construção STIX, identifica a fonte das informações de ameaça, por exemplo, nome e organização do analista. Embora o campo "criador"possa ser usado para atribuir localmente confiança aos dados compartilhados, também pode facilitar a implementação de restrições de manipulação baseadas na organização.

Para maximizar a compatibilidade e a facilidade de adoção, o STIX propõe a adequação de normas descritivas com alguns padrões existentes, como CybOX[12] (Cyber Observable eXpression), a MAEC[13] (Malware Attribute Enumeration and Characterization), o CVE[14] (Vulnerabilidades e exposições comuns), e o CPE[15] (Common Platform Enumeration). Veja a seguir algumas iniciativas e maneira como elas foram empregadas:

- O Departamento de Segurança Interna dos EUA (DHS) está utilizando STIX em várias áreas críticas, incluindo o esforço de Troca automatizada confiável de informações sobre indicadores (TAXII), que permite ao Escritório de segurança cibernética e comunicações (CS&C) e seus parceiros no governo e no setor privado, realizar a troca de elementos de dados e relacionamentos definidos pela STIX usando mecanismos automatizados seguros. Com o uso do STIX, eles buscam permitir a rápida detecção, prevenção e mitigação de ameaças cibernéticas e, sempre que possível, automatizam os principais elementos desse processo. Os esforços iniciais de prova de conceito para o TAXII estão em andamento [38].

- O Programa de Compartilhamento e Colaboração de Informações Cibernéticas do DHS (CISCP) está atualmente utilizando o STIX para a publicação de todas as suas informações de ameaças operacionais aos parceiros do programa [39].

- A IPA japonesa (Agência de Promoção de Tecnologia da Informação, Japão) está atualmente realizando um estudo de viabilidade ativo da aplicação de elementos da arquitetura

[12] <https://cyboxproject.github.io/>
[13] <http://maec.mitre.org/>
[14] <http://cve.mitre.org/>
[15] <http://cpe.mitre.org/>

STIX (CybOX, MAEC, etc.) como um formato de intercâmbio internacional para observáveis cibernéticos e informações sobre ameaças [40].

O padrão STIX vem conquistando um espaço significativo dentro de organizações e comunidades que visam enfrentar os desafios de empreender e/ou apoiar na na evolução do registro padronizado e compartilhamento de dados voltados a ameaças cibernéticas.

2.5.2 TAXII

Como já exposto, o compartilhamento de conhecimento e evidências de ameaças cibernéticas é vital para o combate a crimes cibernéticos atuais, de forma eficaz. Felizmente essa ação é realizada desde o inicio, existindo alguns métodos de compartilhamento de informação amplamente utilizados. Conforme disposto a seguir:

- Listas de E-mail
- Discussões em portais protegidos
- Wiki / Edição colaborativa
- Repositórios de dados
- Feed de dados / notificações de dados
- Chat / Comunicação em tempo real

Entretanto, mesmo com seu uso, a maioria desses métodos (se não todos) não permite à inserção e consumo, de forma automatizada de informações no âmbito de ameaças cibernéticas. A maioria dos agentes que consomem esses dados, rotineiramente coletam informações brutas (não estruturada) de ameaças cibernéticas de *feeds* como os citados, e as sintetiza (normalizando-as) em seu próprio banco de dados interno sobre ameaças. Dessa forma, a opção ainda mais utilizada hoje em dia, é a extração e a inserção manual de conteúdo, por parte de um ativo humano.

A única solução viável para o problema da necessidade universal de compartilhamento, tendo em vista a possibilidade de consumir e compartilhar dados, é uma abordagem com uma arquitetura aberta, baseada em padrões de informações em uso atualmente. Houve algumas tentativas de criação de estrutura de compartilhamento, o REN-ISAC, por exemplo, emprega um sistema de compartilhamento de indicadores baseado na plataforma CIF (*Collective Intelligence Framework*) como parte de seu *Security Event System* [41]. O Laboratório Nacional de Argonne, parte do Departamento de Energia dos EUA, desenvolveu o *Cyber Federated Model* (CFM) [42] para compartilhar listas de bloqueio com comunidades federadas de parceiros. Os consórcios regionais e setoriais enérgicos, também desenvolveram modelos baseados em padrões para compartilhar indicadores dentro de sua comunidade. No entanto, todos os esforços foram idealizados para satisfazer apenas as necessidades próprias de cada comunidades, nenhum desses esforços resultou em um padrão entre comunidades para o compartilhamento interoperável de indicadores.

2.5.2.1 Necessidade de solução de compartilhamento de dados

Tendo como base o cenário atual da segurança da cibernética, se faz necessária uma solução mais ampla de compartilhamento de informações sobre ameaças cibernéticas, que abranja diferentes comunidades e modelos de compartilhamento, permita diferentes métodos de compartilhamento e ofereça suporte a uma ampla gama de dados de proveniente de ameaças. Em particular, os objetivos gerais de uma solução ideal de compartilhamento de informações sobre ameaças cibernéticas são:

I. Permitir compartilhamento de informações de ameaças cibernéticas mais rápido e preciso;

II. Reduzir as atividades tediosas de analistas humanos, por exemplo, entrada de dados, e liberar tempo do analista para o trabalho de análise, mais valioso;

III. Mover ameaças mais bem compreendidas da análise humana para o processamento da máquina;

IV. Permitir o compartilhamento automatizado de uma ampla gama de dados de ameaças, além de indicadores atômicos simples, para permitir a defesa ativa;

V. Proteger o compartilhamento de dados de ameaças confidenciais;

VI. Permitir a inserção automática de dados de ameaças compartilhadas nas bases de conhecimento de ameaças locais, mas com contexto e discrição, exigindo menos olhos de analista necessários para tal;

VII. Permitir a colaboração de analistas entre organizações nas questões verdadeiramente desafiadoras.

Somente com a utilização de formatos de dados de ameaças padronizados, as implementações de compartilhamento de informações sobre ameaças cibernéticas poderão atingir esses objetivos. Conforme observado em [43], um roteiro para a segurança da informação orientada pela inteligência: "Sistemas automatizados de troca de dados precisam ser estabelecidos para remover a dependência de pessoas específicas. Além disso, padrões harmonizados para representar informações de ataque em formato legível por máquina, fornecê-las com segurança e consumi-las em tempo real ajudariam a habilitar a automação."

Com esse objetivo, a Mitre criou uma iniciativa, de forma aberta e colaborativa, o TAXII (*Trusted Automated eXchange of Indicator Information*) [37] que corresponde a um conjunto de especificações técnicas e documentação de suporte para a troca segura e independente de plataforma de informações de ameaças cibernéticas. As especificações do TAXII foram projetadas para aprimorar a interoperabilidade de diferentes soluções de segurança cibernética, em vez de adotar uma tecnologia ou produto específico, e os fornecedores são incentivados a incorporar suporte às especificações do TAXII em seus produtos e serviços de segurança cibernética. Ao

oferecer suporte às especificações TAXII, os fornecedores aumentam o valor de suas soluções, permitindo que seus clientes aproveitem a inteligência acionável de várias fontes.

A primeira parte da automação da troca de informações sobre ameaças cibernéticas é estabelecer consenso sobre o que está sendo compartilhado. O TAXII usa uma linguagem padronizada para expressar informações sobre ameaças cibernéticas o padrão STIX. O STIX fornece mecanismos para a separação e manipulação por protocolos de transmissão de dados, como o TLP (Traffic Light Protocol) [16], utilizado pelo TAXII. A Figura 2.2 ilustra a troca de dados automatizada pelo TAXII.

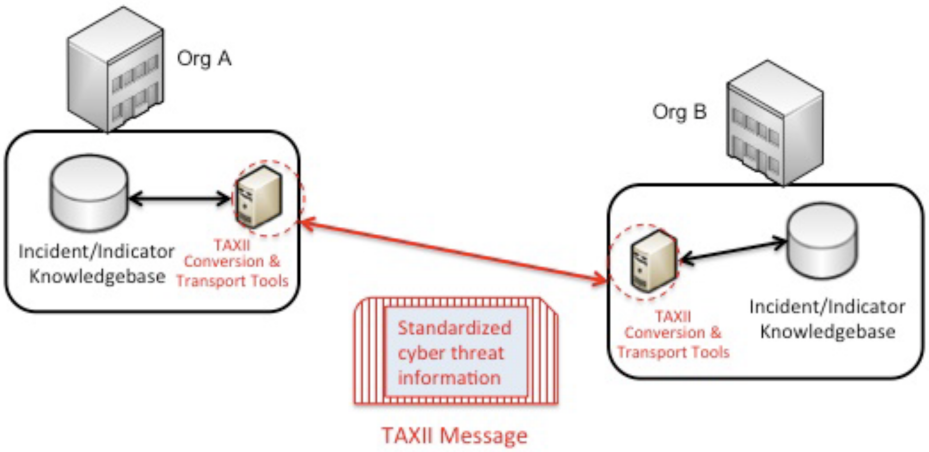

Figura 2.2: Fluxograma Taxii

2.5.2.2 Principais objetivos

Os principais objetivos do padrão TAXII são:

- Permitir o compartilhamento oportuno e seguro de informações sobre ameaças dentro e entre comunidades de defensores cibernéticos;
- Alavancar padrões de consenso para permitir o compartilhamento de indicadores acionáveis e muito mais através das fronteiras da organização e do produto / serviço;
- Ampliar o compartilhamento de indicadores para permitir trocas robustas, seguras e de alto volume de conjuntos significativamente mais expressivos de informações sobre ameaças cibernéticas;
- Apoiar uma ampla gama de casos de uso e práticas comuns às comunidades de compartilhamento de informações de ameaças cibernéticas;
- Aproveitar os padrões maduros existentes, quando apropriado;
- Eventual adoção por uma ou mais organizações internacionais de padrões.

[16] <http://www.us-cert.gov/tlp/>

2.5.2.3 Benefícios propostos

O padrão TAXII foi criado visando trazer os seguintes benefícios:

- Alertas e avisos públicos: Avisos para o público geral, transmitidos ou publicados a assinantes. Esses alertas são de natureza tão ampla que nenhuma criptografia, acesso ou autorização especial é necessária, porém uma assinatura digital, no entanto, é importante para garantir a autenticidade.
- Alertas privados: Os alertas privados são semelhantes aos alertas públicos, exceto que as informações de ameaças compartilhadas são confidenciais e restritas ao compartilhamento de parceiros. Dependendo da natureza dos comunicados, diretrizes explícitas de manuseio ou marcações para restrições ao compartilhamento de dados podem ser necessárias.
- Suporte a consultas: Essa funcionalidade permite ao servidor contestar com a informação solicitada, quando um cliente anteriormente cadastrado a solicitar.
- Transferência em massa: Várias organizações de compartilhamento de informações sobre ameaças cibernéticas podem ter novos membros, possivelmente empresas membros ou outras organizações. O novo membro pode exigir um "despejo de dados"do repositório de dados de ameaças da organização. Portanto, espera-se que um modo de “transferência em massa” precise ser suportado pelo TAXII.

O TAXII utiliza uma estrutura cliente/servidor [44]. Aquele que deseja disponibilizar os dados a partir de uma estrutura TAXII, deve implementar os aspectos referentes à abordagem de servidor. E em contrapartida, aquele que deseja consumir os dados disponibilizados no padrão TAXII, deve implementar aspectos referentes à abordagem de um cliente.

2.6 TRABALHOS CORRELATOS

Inteligência de ameaças cibernéticas é um assunto predominantemente discutido no âmbito da indústria de tecnologia da informação e comunicação (TIC). Ao se buscar o assunto nesse âmbito, é possível se encontrar uma variedade interessante de ferramentas que trabalham diretamente com CTI. Porém ao se realizar buscas sobre o assunto em locais vinculados diretamente ao meio acadêmico, os resultados não são tão expressivos.

Nos resultados das buscas realizadas no âmbito acadêmico, obtidos no *google scholar*[17] por consultas utilizando *queries* como: "*Cyber Threat Intelligence*", "*sharing*", "*develop model*", "*tools*", entre outras. Foram encontrados trabalhos cujo o conteúdo é direcionado para o meio corporativo. É o caso por exemplo do artigo de Joseph C.Magee, AlisonM.Andrews, Mark W. Nicholson, Jonathon Lance James, Henry C. Li, Christopher L. Steverson e Joel Lathrop [45],

[17]<http://www.us-cert.gov/tlp/>

que relata o registro de uma patente, voltada para a coleta automatizada de dados na Internet. O sistema de coleta realiza o refinamento dos dados em forma de cascata, onde um mesmo dado é refinado varias vezes, porém agregando informações de outros *feeds*, visando gerar conhecimento estratégico de forma automatizada.

Outro artigo interessante encontrado, com autoria de Cynthia Wagner, Alexandre Dulaunoy, Gérard Wagener e Andras Iklody [46], detalha o projeto e implementação de uma ferramenta de CTI para combater ameaças. Esse trabalho destaca a importância de realizar a separação de criticidade de informações, juntamente a necessidade de realizar seu compartilhamento.

Um artigo com foco estritamente direcionados ao meio acadêmico, de autoria de Sarah Brown, Oscar Serrano e Joep Gommers [47], detalhando tópicos específicos pertencentes ao contexto de CTI, descreve a falta de efetividade de ferramentas de detecção de ameaças que se baseiam em assinaturas, implicando assim na necessidade do uso de ferramentas capazes de coletar dados de várias fontes e contextualizar essa informação adquirida ao ambiente cuja proteção está na responsabilidade do analista que está configurando a ferramenta. Também destaca que é importante auxiliar a ferramenta na busca do objetivo de extrair dados e gerar conhecimento realmente útil e valioso, para aquela organização, sempre colocando foco na importância da padronização, para a continuidade do desenvolvimento desse conhecimento.

Outro trabalho sobre o tema, com foco direcionado ao compartilhamento de dados, foi desenvolvido por Aziz Mohaisen, Omar Al-Ibrahin, Kevin Kwiat, Charles Kamhoua e Laurent Njilla [48]. Este trabalho conclui que a verificação de focos de intrusão de forma manual para todos os ativos em infraestrutura, se tornou impossível, sendo a única saída viável o compartilhamento desses de dados de CTI. Isso se deve ao fato da imensa quantidade de novos hardwares e software e da predominância de modelos de desenvolvimento flexíveis como o da *Internet of Things* (IOT) que deixam o mercado propício a pequenos desenvolvedores/criadores. Em razão da criticidade dos dados, o trabalho propõe a criação de um modelo de infraestrutura voltado ao compartilhamento, submetido a riscos, escolhendo uma estrutura de dados aberta e independente de fabricante, mas com controle da qualidade dos dados que são compartilhados. Dessa forma procura-se evitar que agentes mal-intencionados possam se conscientizar sobre o nível de conhecimento de ameaças, ou venha de alguma forma alterar os dados, gerando informação comprometida.

Outro trabalho voltado à descrição de ameaças que necessitam de ferramentas como as de CTI para serem combatidas, foi desenvolvido por Wiem Tounsi e Helmi Rais [49]. Esse artigo destaca a inutilidade de ferramentas usuais de segurança da informação, contra o que ele chama de "nova geração de ameaças", devido ao fato delas serem mais evasivas, resilientes e complexas. Além de ameaça APT, os seguintes exemplos foram apresentados como pertencentes a essa nova geração de ameaças:

- *Polymorphic threats*: ameaças polimórficas são ataques cibernéticos, como vírus, worms ou cavalos de Tróia, que mudam ("morph") constantemente [50]. A evolução de ameaças polimórficas pode ocorrer de diferentes maneiras (por exemplo, alterações de nome de arquivo e compactação de arquivo). Apesar da aparência variável do código em uma ameaça polimór-

fica após cada mutação, a função essencial geralmente permanece a mesma. Por exemplo, um malware destinado a atuar como um registrador de chaves continuará a executar essa função mesmo que sua assinatura tenha sido alterada.

- *Zero-day threats*: ameaças de dia zero são ameaças cibernéticas em uma vulnerabilidade publicamente desconhecida de um sistema operacional ou aplicativo. É assim chamado porque o ataque é lançado no "dia zero" ou antes da conscientização pública da vulnerabilidade e, em muitos casos, antes que o fornecedor saiba [23]. Em alguns casos, o fornecedor já está ciente da vulnerabilidade, mas não a divulgou publicamente porque a vulnerabilidade ainda não foi corrigida.

- *Composite threats*: Os ataques cibernéticos podem ser classificados como ataques sintáticos ou semânticos. Uma combinação dessas duas abordagens é conhecida como ataques compostos ou ataques combinados. Os ataques sintáticos exploram vulnerabilidades técnicas em software e/ou hardware, por exemplo, uma instalação de malware para roubar dados; enquanto ataques semânticos exploram vulnerabilidades sociais para obter informações pessoais, por exemplo, solicitações de fraude [51]. Nos últimos anos, houve progresso usando as duas abordagens para realizar ataques compostos.

O trabalho [49] relata que para com que essas ameaças sejam combatidas de forma satisfatória, a cadeia de informações consumidas pela ferramenta de CTI, precisa ser atualizada frequentemente, tendo como melhor opção, a automatização.

2.6.1 Ferramentas correlatas

Além dos trabalhos que pormenorizaram o tema, é importante destacar as ferramentas existentes.

Uma das ferramentas de CTI que vem ganhando espaço no mercado, é o Luminar da Verint [18]. O Luminar é vendido em formato de serviço, em que o usuário coloca termos e dados de interesse, como por exemplo IPS, nomes de marcas e tipos de tecnologia, e a ferramenta lhe proporciona um conhecimento estratégico disponível sobre eles. O objetivo da ferramenta é isentar o usuário de manipular dados e da preocupação com processos e a sapiência necessária para a produção desse tipo de conhecimento. Para suportar essa estratégia de funcionamento, a Verint tem uma base de dados fomentada diariamente com conteúdo selecionado e processado, provenientes de fontes diversas.

Outra ferramenta interessante de ser comentada é a *Cyber Threat Intelligence Exchange* da McAfee[19], que tem o objetivo de utilizar os dados provenientes da infraestrutura de seu usuário, como ponto de função para a localização de conhecimento de CTI que seja estratégico para essa organização. A ferramenta utiliza esses dados também para auxiliar na descoberta e validação

[18] <https://cis.verint.com/product/cyber-security/luminar/>
[19] <https://www.mcafee.com/enterprise/en-us/products/threat-intelligence-exchange.html>

de ameaças, como por exemplo, localizando IPs que estão agindo de forma maliciosa em mais de uma organização. Todo conhecimento gerado a partir dos dados dos usuários podem ser compartilhados com outras organizações, utilizando as diretivas da própria ferramenta. Seu objetivo principal é criar uma cadeia de conhecimento limitada a quem utiliza esse serviço.

A Anomalli oferece uma ferramenta de CTI gratuita, a STAXX[20]. A STAXX é uma ferramenta que possui um *feed* de conhecimento próprio e pré-mapeado, sendo possível solicitar sua coleta e ter visibilidade imediata de seu conteúdo, juntamente com a possibilidade de realizar buscas por palavras chaves e termos. Esse *feed* é uma versão demonstrativa do *feed* vendido pela empresa. A STAXX permite também a seleção e extração de *feeds* TAXII, e a indexação manual de dados no formato STIX v2. A visibilidade gerada pela ferramenta é composta de gráficos quantitativos, que apresentam o número total de tipos de ameaças mapeadas, oferecendo uma visão superficial de seu conteúdo.

Existem também ferramentas criadas em código aberto para fins de CTI, como o OpenCTI[21] criada pela MITRE. Ela é uma ferramenta que tem como objetivo principal o auxílio na identificação e confirmação de ameaças, possuindo extensões que realizam consultas em mecanismos que possuem o objetivo de concentrar informações sobre ameaças, como por exemplo o virustotal[22]. Além das extensões para consultas, o OpenCTI possui vários *feeds* mapeados e cadastros, sendo apenas necessário habilitar a importação de *feeds* para que ela os importe periodicamente. Além do mais, ela permite a importação de dados no formato STIX v1 e v2, permitindo assim armazenar esse conteúdo para utilização para em futuras buscas (por termos e palavras chaves), e também consultar a veracidade ou realizar a confirmação, desses dados importados.

Outro exemplo de ferramenta de código aberto é a MISP[23]. Ela contém *feeds* já mapeados, sendo necessário apenas requisitar sua inserção para que esses dados sejam coletados e armazenados localmente. Ela permite a inserção de manual de *feeds* de dados no formato TAXII e o registro manual de eventos singulares no formato STIX v1 (como ataques e indícios). A MISP permite a visibilidade dos dados armazenados em gráfico de mapa de vínculos, um dos formatos mais interessantes para esse tipo de dado. A funcionalidade mais singular dessa ferramenta, é a possibilidade de realizar o compartilhamento de dados armazenados em formato TAXII, padrão aberto. Porém ela não oferece mecanismos para que sejam realizados filtros nos dados que se deseja compartilhar.

Uma ultima ferramenta de código aberto analisada foi a YETI [24], uma ferramenta criada utilizando como base estrutural o padrão STIX v1. A YETI possui *feeds* previamente mapeados e registrados, apenas sendo necessário a solicitação de coleta de seu conteúdo. Uma funcionalidade interessante oferecida, é a possibilidade de registar objetos STIX manualmente, permitindo o registro previamente de conhecimento criado fora da ferramenta. Porém a funcionalidade que

[20]<https://www.anomali.com/community>
[21]<https://www.opencti.io/en/>
[22]<https://www.virustotal.com>
[23]<https://www.misp-project.org/>
[24]<https://yeti-platform.github.io/>

mais de destaca na YETI é a criação da visibilidade dos dados armazenados pela luz do "mapa de vínculo", um tipo de gráfico que permite uma visão holística sobre os dados armazenados, mostrando seus relacionamentos.

2.6.2 Síntese de ferramentas

A Tabela 2.8 apresenta uma síntese das ferramentas de inteligência de ameaças cibernéticas analisadas.

Tabela 2.8: Síntese de ferramentas analisadas

Nome	Tipo	Tecnologia	Principal funcionalidade
Luminar (Verint)	Privada	Privada	Consulta de itens chave (IP, hostname, marcas) dentro de base dados da ferramenta.
CTI Exchange (McAfee)	Privada	Privada	Uso de dados da infraestrutura vigente para produzir inteligência.
STAXX (Anomalli)	Gratuita	STIX v2	Coleta de dados provenientes de servidores que respondem padrão TAXII.
OpenCTI	Gratuita	CyBoX	Pesquisa de dados em outras plataformas, visando enriquecer seu uso.
YETI	Gratuita	STIX v1	Visibilidade de dados utilizando mapa de vínculos.
MISP	Gratuita	STIX v2	Compartilhamento de dados usando o padrão TAXII, enviando a pilha inteira de dados armazenados.
Minerva	Gratuita	STIX v2	Compartilhamento de dados usando o padrão TAXII, limitando acesso a dados específicos à usuários específicos.

3 PROPOSTA DE MODELO DE REFERÊNCIA

Como já visto, no âmbito industrial existem muitas ferramentas criadas exclusivamente para atuar no contexto de CTI. Enquanto que no contexto acadêmico, as principais contribuições relatam a utilização de métodos e ferramentas em contextos específicos, detalhando sua necessidade ou pormenorizam funcionalidades específicas, propondo melhorias, outras maneiras de implementar, ou mesmo, explicando os ganhos que podem ser gerados com seu uso.

Na pesquisa realizada no *google scholar*[1], não foi encontrado nenhum documento que busque detalhar quais são as caraterísticas e funcionalidades inerentes a uma aplicação atuante no contexto de CTI. Isso resulta em parte devido ao fato das ferramentas surgirem antes de uma fundamentação teórica adequada, propiciando a criação de novas abordagens à funcionalidades existentes [52], ou explanando o motivo de determinado comportamento, para melhorar ou consolidar o uso do que já foi criado, ao invés de redigir uma base para novas ferramentas, ou a auditoria das que já existem.

Neste trabalho é proposto um modelo de referência para o desenvolvimento de aplicações de CTI, e para provar a sua eficácia, uma ferramenta foi desenvolvida utilizando-o como fundamento.

3.1 MODELO DE REFERÊNCIA PARA O DESENVOLVIMENTO DE APLICAÇÕES DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

O modelo de referência para o desenvolvimento de aplicações proposto para inteligência de ameaças cibernéticas, ilustrado na Figura 3.1, pretende incorporar as funcionalidades inerentes ao contexto de CTI, encontradas em ferramentas e métodos, criados no âmbito industrial, bem como aqueles descritos em documentos acadêmicos.

Para organizar e facilitar a compreensão do modelo de referência proposto, seu conteúdo foi dividido em 6 camadas, Gerenciamento, Armazenamento, Coleta, Geração, Pesquisa e Compartilhamento. Cada uma delas representando conceitos, características e funcionalidades, encontrados nos documentos e ferramentas analisadas. Cada camada foi planejada como parte inseparável de um modelo unificado, sendo pormenorizadas suas funções e objetivos. Embora não exista a necessidade de uma ferramenta implementar todas as camadas, é sempre imprescindível conhecer o relacionamento entre elas.

A seguir veja a descrição de cada uma das camadas.

[1] <http://www.us-cert.gov/tlp/>

Figura 3.1: Modelo de referência para o desenvolvimento de aplicações de inteligência de ameaças cibernéticas

3.1.1 Gerenciamento

A camada de Gerenciamento é responsável por controlar a aplicação e suas funcionalidades. Suas atribuições incluem o controle dos usuários que acessam a aplicação, o controle voltado à manipulação das funcionalidades oferecidas e o gerenciamento da interação entre essas funcionalidades.

A camada de Gerenciamento relaciona-se com todas as outras camadas, apoiando o controle de entrada de dados junto à camada de Coleta e a camada de Geração. também auxilia no gerenciamento de dados armazenados junto a camada de Armazenamento e trabalha junto da camada de Pesquisa, gerando controle de acesso a visibilidades sobre os dados. Além disso, a camada de Gerenciamento controla o outorgamento da funcionalidade de compartilhamento, para usuários específicos acessarem de fora da aplicação à dados específicos, trabalhando junto com a camada de compartilhamento e a camada de Armazenamento.

A implementação da camada de Gerenciamento por parte da aplicação deve fundamentar a gerência de permissões. Essa funcionalidade deve ser vinculada a todas ações que podem ser desempenhadas pelos usuários, como a criação de solicitações para coletar dados, a escolha da localização de armazenamento de dados, a manipulação de dados armazenados, o compartilhamento de dados, entre outras.

É aconselhável que a base de dados utilizada na implementação da camada de Gerenciamento, seja exclusiva para esse fim, pois requer uma estrutura que suporte todas as funcionalidades, sem a possibilidade de dependência de recursos.

O gerenciamento de aplicações de CTI segue as diretivas de gerenciamento de aplicações de risco, considerando o nível de criticidade e sigilo, contidos nos dados armazenados. A estrutura

da camada de Gerenciamento dentro da aplicação, é implementada segundo diretivas das organizações que a desenvolvem. Um exemplo disso é apresentado em [53], onde a estrutura de uma aplicação de CTI é divida em políticas de acesso e separada por ativos, visando proteger e garantir, a confiabilidade dos dados.

Essa funcionalidade foi implementada comumente em todas as aplicações analisadas, como o YETI, OpenCTI e STAXX, pelo menos parcialmente. Um exemplo disso, é que todas as aplicações citadas possuem controle de acesso às funcionalidades propostas pelas ferramentas, sendo exigido cadastro prévio de usuário para esse fim.

3.1.2 Armazenamento

A camada de Armazenamento deve trabalhar junto com a camada de Gerenciamento, e realizar o controle dos dados de CTI que entram, saem e permanecem na aplicação. Para suportar a funcionalidade de armazenamento de dados, a aplicação deve possuir uma estrutura de armazenamento capaz de sustentar a quantidade de entrada e saída de dados almejada para o bom funcionamento das aplicações. As principais estruturas de armazenamento atualmente são bancos de dados relacionais, bancos de dados não relacionais e estruturas customizadas, utilizadas em softwares proprietários.

A estrutura implementada para suportar a camada de Armazenamento deve ser capaz de armazenar dados enviados pelas camadas de Coleta e de Geração, e disponibilizar dados no formato exigido pelas camadas de Pesquisa e de Compartilhamento.

Existem diversas alternativas de bases de dados que podem ser utilizados pra implementar uma ferramenta de inteligência de ameaças cibernéticas [54]. Devido às características de melhor gestão de inserção de dados, e mais eficacia no controle de grandes quantidades de dados, os melhores tipos de bases de dados para essas ferramentas, são bancos dados não relacionais.

A camada de Armazenamento é comumente implementada por todas as aplicações, porém a disponibilização de seus recursos para usuários finais é legada somente por aplicações que lhes fornecem a possibilidade de inserção de dados (camadas de Coleta e de Geração).

3.1.3 Coleta

A camada de Coleta engloba a capacidade de coletar dados provenientes de fontes externas à infraestrutura que suporta a aplicação de CTI, normalizá-los (quando necessário) e enviá-los para a camada de Armazenamento para serem internalizados. A camada de Coleta deve trabalhar junto com a camada de Gerenciamento para controlar da melhor forma possível os *feeds* de interesse, juntamente com a constância da coleta.

Aplicações que implementam a camada de Coleta, devem oferecer funcionalidades que utilizem mecanismos voltados para a coleta de dados, que são disponibilizados em diversas estruturas e padrões, como pode exemplo *wikis*, comunidades de segurança e servidores padrão TAXII.

Devem oferecer também métodos que padronizem esses dados em um único formato, sendo obrigatório a compatibilidade com o formato requerido pela estrutura implementada pela camada de armazenamento, como por exemplo o padrão STIX.

Essas funcionalidades propostas na camada de coleta são vitais para a utilização de uma aplicação de CTI, pois permite contextualizar a informação coletada, com os ativos da infraestrutura a ser monitorada [45]. O modo como a coleta deve ser realizado buscando tirar melhor proveito deve ser uma conjectura entre a infraestrutura e a técnicas de coleta elegida pelo desenvolver da aplicação, como exemplo o uso de robôs, coletas de *feed* em padrões diversos, etc.

A maioria das ferramentas agregam as funcionalidades propostas na camada de Coleta em sua lista, porém são poucas as que permitem com que o usuário final possam manipulá-la, tendo a capacidade de inserir os *feeds* que sejam de seu interesse particular. A Lumir, ferramenta de inteligência de ameaças cibernéticas da Verint[2], por exemplo, apenas permite que o usuário realize consultas de interesse em sua interface, mas não permite ao usuário inserir nenhum dado já conhecido, nem alterar os existentes. Por outro lado, a ferramenta Anomali[3], já é mais flexível, permitindo a inserção de dados (em formato próprio da ferramenta) e sua manipulação por parte dos analistas.

3.1.4 Geração

A camada de Geração contempla funcionalidades voltadas para a coleta de dados que são gerados por ativos que pertencem à infraestrutura interna das organizações que fazem uso das ferramentas de CTI. Essa camada deve trabalhar junto com a camada de Gerenciamento para controlar da melhor forma possível os ativos que enviam dados para a aplicação, juntamente com a frequência de seu recebimento.

Aplicações que implementam as funcionalidades da camada de Geração, devem possuir mecanismos que voltados para padronizar a estrutura dos dados recebidos, visando unificá-los em formato compatível com o padrão requerido pela camada de Armazenamento.

A camada de Geração visa proporcionar a geração de valor para a organização que a implementa, usando ativos da infraestrutura computacional vigente como *feed*. Ativos de segurança geram dados sobre as ameaças que combatem, e catalogam seus detalhes. Esses dados gerados podem ser contextualizados em informação útil e utilizados para CTI.

Essa capacidade de gerar dados é extremamente complicada de ser implementada por ferramentas de CTI. Em geral as ferramentas de CTI consomem os dados disponibilizados por ferramentas de SIEM (*Security Information and Event Managment*). Uma exemplo de ferramenta que implementa as funcionalidades da camada de Geração é o *Threat Itelligence Sharing* da McAfee[4], enquanto que um exemplo de ferramenta que consome o dados diretamente de um SIEM é

[2] <https://cis.verint.com/product/cyber-security/luminar/>
[3] <https://www.anomali.com/>
[4] <https://www.mcafee.com/enterprise/en-us/products/threat-intelligence-exchange.html>

a OpenCTI[5], que possui plug-in nativo para o Splunk[6].

3.1.5 Pesquisa

A camada de Pesquisa é responsável por agregar mecanismos e métodos visando a manipulação, exploração e enriquecimento de dados armazenados. Essa camada deve trabalhar junto com as camadas de armazenamento e gerenciamento, pois os dados devem ser controlados pela camada de Gerenciamento e enviados para a camada de Pesquisa pela camada de Armazenamento.

A ferramenta que implementar a camada de Pesquisa, deve oferecer mecanismos que sejam capazes de ler os dados nos padrões disponibilizados pela camada de Armazenamento, de gerar visibilidade adequada desses dados, em padrão compreensível pelos analistas que usam a ferramenta, segundo a necessidade de visualização dentro de um escopo escolhido pelo fabricante/desenvolvedor. A visibilidade referida não é baseada apenas de forma gráfica, porém também no detalhamento dos dados em forma tabelar, *feedbacks* de autogerenciamento e alertas.

A visibilidade gerada baseada sobre a base de dados é uma estrutura dependente do diretamente do desenvolvedor da aplicação. Em [55], observa-se que visibilidade é interessante a partir do momento que pode ser vinculada a algum modelo de referencia na busca por descoberta por ameaças ou agentes mal-intencionados, como por exemplo a *Kill Chain*.

A visibilidade gerada baseado nos dados é uma funcionalidade amplamente implementada. Ferramentas como a Luminar da Verint[7] somente possibilitam ao usuário final a visualização dos resultados finais, enquanto que ferramentas como a MISP[8], permitem a manipulação dos dados e a construção de visibilidade customizada.

3.1.6 Compartilhamento

A camada de compartilhamento tem como objetivo agrupar funcionalidades que visam o compartilhamento de dados com requerentes fora da estrutura da aplicação. O trabalho junto com a camada de Gerenciamento é fundamental para essa camada, pois a criação de regras de negócio voltadas à permissividade, é essencial para o gerenciamento de forma granular.

É aconselhável que ferramentas que implementem a camada de Compartilhamento, por motivos de segurança, possuam uma estrutura de compartilhamento separada, com acesso controlado e limitado somente à camada de Armazenamento.

A funcionalidade de compartilhamento permite compartilhar o conhecimento produzido com terceiros, sendo a principal arma contra o combate a ameaças modernas (como o APT).

A maioria das aplicações que implementam essa funcionalidade realizam o compartilhamento

[5]<https://www.opencti.io/en/>
[6]<https://www.splunk.com/>
[7]<https://cis.verint.com/product/cyber-security/luminar/>
[8]<https://www.misp-project.org/>

em formato proprietário, atingindo somente uma comunidade alvo, como é o caso da *Threat Intelligence Exchange* da McAfee[9]. Um exemplo de ferramenta que implementa essa funcionalidade com padrão aberto é a MISP[10][11], porém seu gerenciamento não permite ser feito de forma granular.

3.2 IMPLEMENTAÇÃO DA FERRAMENTA DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

Utilizando como fundamento o modelo de referência proposto para o desenvolvimento de aplicações de CTI, (sessão item 3.1), foi desenvolvido uma aplicação, que possui fins estritamente demonstrativos. Todo seu código e estrutura, foram produzidos unicamente dentro do contexto acadêmico, diferentemente do padrão de desenvolvimento com fins lucrativos, característico dessa área do conhecimento.

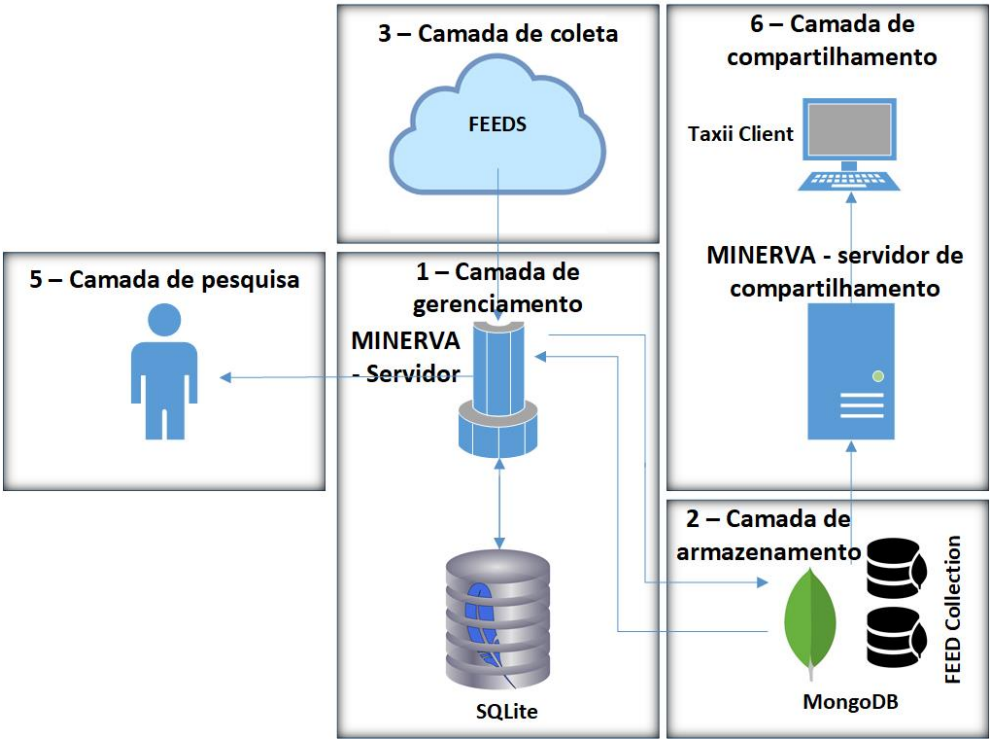

Figura 3.2: Estrutura da aplicação demonstrativa

A Figura 3.2 apresenta o desenho das camadas implementadas, ressaltando-se que sua estrutura contempla cinco, das seis camadas do modelo de referência proposto. As camadas que foram implementadas na primeira fase do desenvolvimento da aplicação foram: A camada de Gerenciamento, a camada de Armazenamento, a camada de Coleta, a camada de Pesquisa e a camada de Compartilhamento. Devido a necessidade de subtração de dados de uma infraestrutura

[9] <https://www.mcafee.com/enterprise/en-us/products/threat-intelligence-exchange.html>
[10] <https://www.misp-project.org/>
[11] <https://github.com/MISP/MISP>

computacional, a camada de Geração não foi implementada.

3.2.1 Componentes utilizados no desenvolvimento

Para o desenvolvimento da aplicação de inteligência demonstrativa foram utilizados os componentes apresentados na Tabela 3.1.

Tabela 3.1: componentes utilizados no desenvolvimento da aplicação demonstrativa

Componente	Versão	função
Python	2.7	Linguagem de programação que suporta toda a aplicação
Flask	1.1	Servidor Web
SqlLite	3.27.2	Utilizado para gerencia
MongoDB	4.2.2	Utilizado para armazenar dados coletados e criados
Bootstrap	4	Utilizado no design da aplicação
d3	4	Utilizado para gerar gráficos
stix2viz	1.1	Utilizado para gerar mapa de vínculos
Chart.js	2.0	Utilizado para gerar gráficos
STIX	2.0	Padrão utilizando da discriminação dos dados
TAXII	2.1	Padrão utilizando no compartilhamento de dados

3.2.2 Planejamento da aplicação

Para suportar todas as funcionalidades recomendadas pelo modelo, foi escolhida uma linguagem de programação ampla e robusta, a Python 2.x. Python permite a flexibilização de objetos, deixando-os mais próximos da ideia original da aplicação, como por exemplo a classe de usuário, criada para suportar as funcionalidades descritas na camada de Gerenciamento, conforme sua estrutura mostrada no Listing 3.1:

Listing 3.1: Classe de usuário

```
1  from flask_login import UserMixin
2  class User(UserMixin):
3      id = ""
4      username = ""
5      name = ""
6      email = ""
7      password = ""
8      share = ""
9
10     def __init__(self):
11         self.id = ""
12         self.username = ""
13         self.name = ""
```

```
14          self.email = ""
15          self.password = ""
16          self.share = ""
17
18      def __init__(self, id, username, name, email, password, share):
19          self.id = id
20          self.username = username
21          self.name = name
22          self.email = email
23          self.password = password
24          self.share = share
25
26      def setFullUser(self, id, username, name, email, password, share):
27          self.id = id
28          self.username = username
29          self.name = name
30          self.email = email
31          self.password = password
32          self.share = share
33
34      def is_authenticated(self):
35          return True
36
37      def is_active(self):
38          return True
39
40      def is_anonymous(self):
41          return False
42
43      def get_id(self):
44          return unicode(self.id)
```

Observa-se no Listing 3.1 que a linha 1 é importada a biblioteca "UserMixin", necessária para o gerenciamento de sessões de usuários, pela aplicação. Na linha 2 é criada a classe "Usuário". Nas linhas 3 a 8, são declaradas as variáveis pertencentes à classe: "id", "username", "name", "e-mail", "password"e "share". Nas linhas 10 a 32, estão as funções que inicializam as instâncias da classe. Nas linhas 34 a 44, estão as funções obrigatórias para se realizar o controle das sessões.

Para permitir o acesso aos recursos da aplicação, foi escolhido o HTTP (Hypertext Transfer Protocol), protocolo base do acesso à web, como meio de acesso a aplicação. O servidor utilizado para receber e interpretar as requisições, e responder adequadamente, foi o Flask, que é um servidor web desenvolvido totalmente em Python, funcionando como uma biblioteca. O Flask também é utilizado pela camada de Compartilhamento, porém como serviço separado, como recomendado pelo modelo proposto. A implementação do Flask como serviço dentro do servidor, é divida em 2 arquivos, run.py e __init__.py, conforme ilustrado nos Listing 3.2 e 3.3.

a) Arquivo: run.py

Listing 3.2: Código de arquivo run.py

```
1 from app import app
2
3 if __name__ == "__main__":
4   app.run(port=8080)
```

Na linha 1 do Listing 3.2 é importada a classe "app", estruturando vinculando o comportamento da aplicação ao entendimento de uma "app"a luz do Python. Nas linhas 3 e 4 é verificada se aquela é a função principal da estrutura que compõe a aplicação, se for ele inicializa é inicializada na porta "8080".

b) Arquivo: __init__.py

Listing 3.3: código de arquivo __init__.py

```
1 from flask import Flask
2
3 app = Flask(__name__)
4
5 from app.controllers.server import default
```

Na linha 1 do Listing 3.3 é importada a biblioteca Flask, habilitando a aplicação a trabalhar com todas as funcionalidades que ela oferece. Na linha 3 toda a estrutura da biblioteca Flask é legada à aplicação, permitindo o seu uso. Na linha 5 é importada toda estrutura que está no diretório "controllers.server", permitindo que eles chamem as funções da biblioteca Flask, previamente importada.

A visibilidade foi refinada utilizando BootStrap, D3 e Chart.js, de modo a gerar visibilidade adequada dos dados, utilizando gráficos, tabelas e outras formas customizadas, conforme sugerido pela camada de Pesquisa. Porém, o acesso a esses recursos deve ser controlado pela camada de Gerenciamento, que foi implementada baseada no banco de dados SQLite. A linguagem Python tem uma biblioteca específica para suportar o gerenciamento dessa base de dados. No Listing 3.4 é apresentado o código de consulta de usuários, usado para verificação de login.

Listing 3.4: código de consulta de usuários

```
1 from app.controllers.classes.user import User
2 import sqlite3
3
4 def listUserDB(id):
5
6     conn = sqlite3.connect("db/taxi_register.db")
7     usersConn = conn.execute("""select * from user where id = ?""",(id,))
8
9     for userConn in usersConn.fetchall():
```

```
10          user = User(userConn[0],userConn[1], userConn[2], userConn[3], userConn[4],
11          userConn[5])
12      conn.close()
13      return user
```

Na linha 1 do Listing 3.4 é importada a classe "User", criada para suportar a estratégia de controle de usuário. Enquanto que na linha 2 é importada a biblioteca "sqlite3", utilizada para realizar comunicação com o banco de dados SQLite. Na linha 6 é estabelecida a conexão com o banco "taxi_register.db"e na linha 7 é executada a *query* de consulta dos usuários. Nas linhas 9, 10 e 11, o resultado da consultada é inserido dentro de uma variável e, nas linhas 12 e 13, a conexão com o banco de dados é encerrada e a variável é retornada a quem chamar a função.

Para suportar o armazenamento de dados em formato STIX v2, em conformidade com o modelo de referência proposto para a camada de armazenamento, foi escolhido a base de dados MongoDB. A linguagem Python também tem uma biblioteca especifica para manipular objetos vinculados essas bases de dados noSQL. O Listing 3.5 apresenta o código responsável pela conexão com o banco.

Listing 3.5: Código de conexão com o Mongodb

```
1 from pymongo import MongoClient
2
3 def selectDatabaseMongoDb(nameDB, urlDB):
4     conn = MongoClient(urlDB)
5     mydb = conn[nameDB]
6     return mydb
```

Observerva-se no Listing 3.5, que na a linha 1 é importada a biblioteca "MongoClient", que permite a comunicação com o MongoDB. Na linha 3 é declarada a função "SelectDatabaseMongoDb", sendo necessário o nome do banco e a url de conexão com o mongoDB, como parâmetros. Na linha 4 é estabelecida a conexão com o MongoDB, enquanto que na linha 5 é selecionado o banco passado como argumento, para manipulação. Por fim, na linha 6 a conexão válida é retornada para quem chamar a função.

3.2.3 Casos de uso da ferramenta

A Aplicação com proposito demonstrativo foi planejada e desenvolvida para atuar com um contexto limitado de funcionalidades. A Figura 3.3 apresenta os casos de uso da ferramenta em seu estado atual: inserção de dados, gerenciamento de armazenamento de dados, manipulação de dados, análise de dados e compartilhamento.

A seguir é feita uma descrição de cada um dos casos de uso.

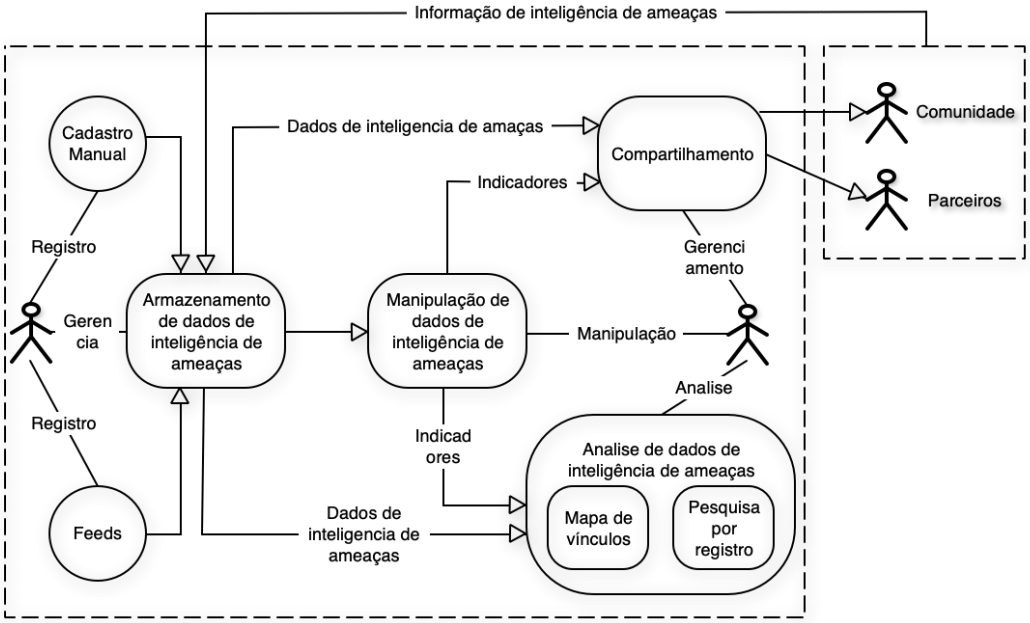

Figura 3.3: Casos de uso

3.2.3.1 Inserção de dados

No status atual da aplicação, a inserção de dados pode ser realizada das seguintes formas:

I. A ferramenta permite que o usuário registre um *feed* de dados que trabalhe no padrão STIX/-TAXII. Ela vai iniciar automaticamente a coleta de todos os dados localizados no *feed*, e vai inseri-los dentro do MongoDB, opção utilizada quando já se tem um *feeds* de CTI de interesse, previamente mapeado.

II. Outra forma de inserção de dados é através de robôs, não existindo a necessidade do usuário inserir nenhuma informação referente à origem do *feed*, somente solicitar sua coleta, pois já estão mapeados. Sua estrutura já foi verificada, e os dados coletado serão normalizados para o padrão STIX v2 e inseridos no MongoDB, automaticamente.

III. Também possui a forma de inserção de dados manualmente. A aplicação disponibiliza a função de cadastro de eventos no padrão STIX v2. Existe um formulário diferente para cada um dos 12 tipos de SDO's e 2 tipos SRO'2, existentes na estrutura do STIX v2. É uma opção utilizada quando já se possui dados de CTI, criados pelo próprio analista ou extraído de fontes, cujo sua inserção não é automaticamente realizada pela aplicação

IV. Pode-se também inserir uma string JSON no padrão STIX v2, que é automaticamente reconhecida e armazenada pela aplicação. Por exemplo, algumas ferramentas de análise de malware já possuem a possibilidade de exportar seus resultados em formato STIX, de modo que basta copiar esses resultados e colar dentro da aplicação.

V. A última forma de inserção de dados é selecionar um arquivo preenchidos com dados no padrão STIX v2 a partir do computador do usuário, e a aplicação vai automaticamente reconhecer os dados e armazená-los. Usando como base o exemplo anterior, se a ferramenta de análise de malware exportar um arquivo no formato STIX v2, basta selecioná-lo e a aplicação importará seu conteúdo automaticamente.

A inserção de dados é vital para o funcionalmente da aplicação, pois o seu funcionamento é baseado na dinâmica da análise desses dados.

3.2.3.2 Gerenciar armazenamento de dados

A aplicação foi desenvolvida para funcionar com uma base dados, porém em estruturas de dados separadas, que o mongoDB as chama de índices. Para que sejam realizadas inserções ou compartilhamento de dados, é necessário especificar um índice. Cabe ao usuário o gerenciamento desses índices e dos dados que lá estão. As seguintes ações podem ser desempenhadas pelos usuário em um índice:

I. Criação de índices;

II. Verificação do tamanho ocupado (segundo a quantidade de dados armazenadas);

III. Remoção de índices.

Para facilitar a gestão, a ferramenta disponibiliza visibilidade em forma de gráfico, baseado na quantidade de dados armazenados em cada índice.

3.2.3.3 Manipulação de dados

Para suportar o padrão STIX v2, os dados armazenados dentro da aplicação são divididos em coleções e objetos, um objeto é um SDO ou SRO, e uma coleção é um conjunto desses objetos, que recebem um ID e um nome. A ferramenta possibilita o usuário a manipular tanto os objetos, quanto as coleções. As seguintes ações podem ser desempenhadas:

I. Mover objeto entre coleções dentro do mesmo índice;

II. Mover objeto entre coleções entre índices diferentes;

III. Mover coleções entre índices;

IV. Apagar objeto;

V. Apagar coleção;

VI. Criar coleção;

VII. Inserir objeto dentro de coleção (vinculado a funcionalidade de inserção);

VIII. Visualizar coleções;

IX. Visualizar objetos de forma detalhada.

3.2.3.4 Análise de dados

A análise de dados armazenados dentro da aplicação é uma das funcionalidades fundamentais para a extração de valor em ferramentas de CTI. A aplicação em sua forma atual, proporciona duas estruturas para análise de dados:

I. Para uma forma mais precisa de localização de dados armazenados dentro da aplicação, o usuário tem a possibilidade de realizar a busca por um registro específico. Essa busca utiliza o ID do objeto que se deseja localizar, e o apresenta de forma detalhada, sendo útil para verificação de duplicidade, existência e localização rápida de registros.

II. Com o objetivo de mostrar melhor o relacionamento dos registros armazenados, a aplicação proporciona uma visibilidade conhecida como "mapa de vínculos", que vincula objetos SDO, a patir de objetos SRO's.

A visibilidade proporcionada pela aplicação é utilizada no contexto de CTI para apresentar de forma clara a cadeia de eventos em que se baseou a decisão de vincular agentes como ameaças. Por exemplo, objetos SDO identificam um agente, outros objetos SDO as ações maliciosas desse agente, e outros objetos SRO são essas ações, essa visibilidade apresenta tudo junto e vinculado. Essa visibilidade apresenta todos os objetos contidos em uma coleção. Se houver a necessidade de que outro objeto faça parte dessa visibilidade, será necessário movê-lo para a coleção em específico.

3.2.3.5 Compartilhamento

A aplicação proporciona a o compartilhamento de dados armazenados, essa funcionalidade é baseada em índices, quando um usuário solicita a coleta de dados de dentro da aplicação, ele vai enviar o conteúdo de um índice inteiro. Para enviar apenas registros específicos, cabe ao usuário criar ao menos um índice especifico e colocar a coleção e os objetos, que ele deseja compartilhar.

A gestão do compartilhamento permite ao usuário escolher o índice que ele deseja que seus dados sejam compartilhados, e quais usuários tenham acesso a esses dados. Sendo necessário um cadastro prévio desses usuários dentro da aplicação.

Como já discutido previamente, essa funcionalidade é essencial para que a transferência de conhecimento, adquirida através de experiência em ferramentas e tecnologias, e registrada aqui em formato STIX v2, seja transferida para parceiros e aliados, através do padrão TAXII.

4 RESULTADOS

Objetivando evidenciar o valor obtido com o uso das funcionalidades propostas pela a aplicação desenvolvida, esse capítulo apresenta resultados com a utilização dessas funcionalidades, juntamente com sua comparação com outras ferramentas.

4.1 FERRAMENTA DE INTELIGÊNCIA DE AMEAÇAS CIBERNÉTICAS

A ferramenta de CTI desenvolvida neste trabalho, com propósito estritamente demonstrativo, foi denominada Minerva. Sua estrutura e funcionalidades, foram baseadas no modelo de referência proposto.

Visando uma abordagem mais organizada, suas funcionalidades foram divididas nos seguintes campos de interesse:

4.1.1 Gerenciamento de usuários e acessos

Devido a criticidade dos dados que são gerenciados pela ferramenta, foi implementado a funcionalidade de gerenciamento de usuários, que realiza o controle de acesso e permissividade. Toda a implementação foi realizada com o banco SQLite. A função pertencente ao código responsável por registrar um usuário dentro do banco dados, é mostrada no Listing 4.1.

Listing 4.1: Função de registro de usuário no SQLite

```
1  from app.controllers.classes.user import User
2  import sqlite3
3
4  def registerUserDB(user):
5
6      conn = sqlite3.connect("db/taxi_register.db")
7      conn.execute("""insert into user (username, name, email, password, share)
8      values (?,?,?,?, ?)""",(user.username, user.name, user.email, user.password,
9      user.share))
10     conn.commit()
11     conn.close()
```

Observa-se no Listing 4.1 que na linha 1 é importada a classe "User", criada para diluir o objeto usuário dentro do código e na linha 2 é importada a biblioteca "sqlite3", utilizada na comunicação com o banco de dados SQLite. Na linha 6 é realizada a conexão com o banco "taxi_register.db"e na linha 7 é feita a inserção dos dados dentro da tabela "user". Na linha 10 as

alterações são salvas e por fim na linha 11 a conexão é encerrada.

O gerenciamento de usuários é realizo a partir da inserção de dados específicos de uma pessoa, caracterizando um usuário válido dentro do sistema. A Figura 4.2 apresenta a tela de cadastro de usuário da aplicação.

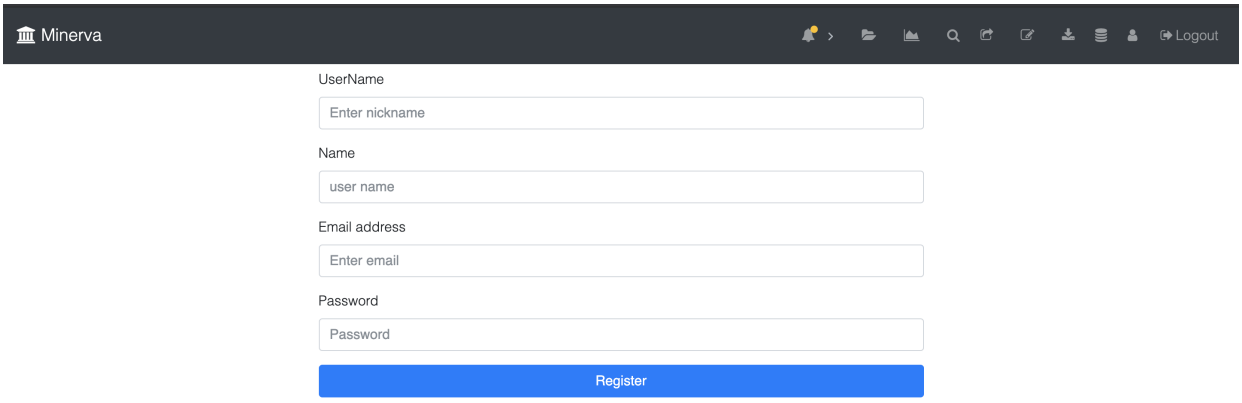

Figura 4.1: Cadastro de usuário no Minerva

Com o intuito de realizar o gerenciamento de usuários que acessam a ferramenta, foi implementando o login, feito a partir do controle de *cookies* aplicados ao gerenciamento de sessões. Essa funcionalidade foi desenvolvida visando a obrigatoriedade do usuário de possuir credenciais válidas, previamente cadastradas na aplicação. A Figura 4.2 apresenta a tela de login na aplicação.

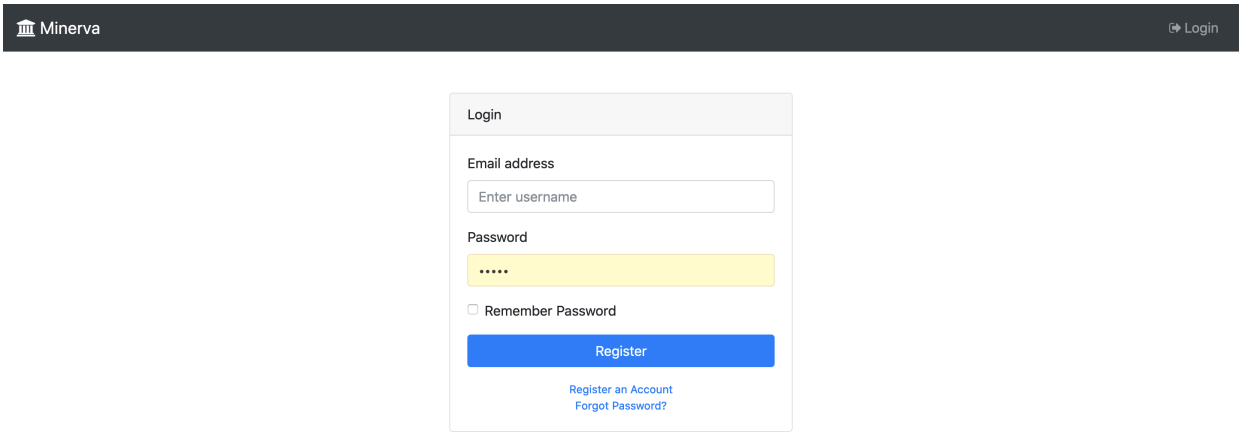

Figura 4.2: Tela de login no Minerva

4.1.2 Gerenciamento de bases de dados

O armazenamento de dados provenientes de *feeds*, é realizado pelo MongoDB. Cada base de dados utilizada para se realizar armazenamento é denominada "índice". A aplicação outorga ao usuário a capacidade/obrigação de gerenciar esses índices. Assim como no gerenciamento de usuários, essa gerencia de dados é realizada pelo SQLite. A função pertencente ao código responsável por registrar índices dentro do banco de dados do Minerva é apresentada no Listing

Listing 4.2: Função de registro de índices no SQLite

```
1  from app.controllers.classes.index import Index
2  import sqlite3
3
4  def registerIndexDB(index):
5
6      conn = sqlite3.connect("db/taxi_register.db")
7      conn.execute("""INSERT INTO indexs (indexName, description, inserted)
8      VALUES (?,?,?)""", (index.indexName, index.description, index.inserted))
9      conn.commit()
10     conn.close()
```

Observa-se no Listing 4.2 que na linha 1 é importada a classe "Index", criada para diluir o objeto usuário dentro do código. Na linha 2 é importada a biblioteca "sqlite3", utilizada na comunicação com o banco de dados SQLite. Na linha 6 é realizada a conexão com o banco "taxi_register.db". Na linha 7 é feita a inserção dos dados dentro da tabela "index". Na linha 10 as alterações são salvas. E por fim na linha 11 a conexão é encerrada.

Devido ao modo como a aplicação foi planejada, esses índices são diretamente relacionados à inserção, manipulação e ao compartilhamento de dados. Sendo necessário a criação de um vínculo com índice existente dentro da aplicação, para realizar qualquer uma das ações especificadas. Cabendo ao usuário criar ou excluir esses índices. A Figura 4.3 apresenta a tela de criação de um índice no Minerva.

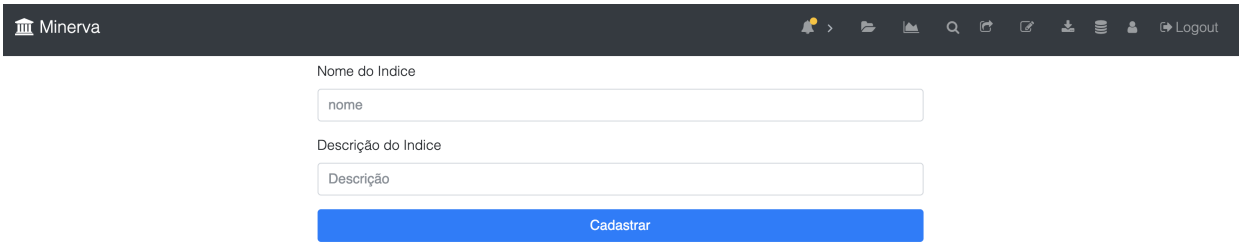

Figura 4.3: Criação de índice no Minerva

4.1.3 Inserção de dados

A inserção de dados consiste em coletar/extrair dados provenientes de fora da infraestrutura vigente da aplicação, normalizá-las (quando preciso) no padrão STIX v2, e inseri-las dentro do índice selecionado (gerenciado conforme subseção anterior). No Minerva, os dados são inseridos diretamente no MongoDB. O código função responsável por extrair dados a partir de um servidor TAXII e inseri-los um índice dentro do mongoDB é mostrado no Listing 4.3.

Listing 4.3: Função de registro de dados dentro do MongoDB

```
1  def insertDataMongoDB(indexName ,urlDB, urlTaxi, user, password):
2
3      server = Server(urlTaxi, user=user, password=password)
4
5      mydb = selectDatabaseMongoDb(indexName, urlDB)
6
7      for api in server.api_roots:
8          for collection in api.collections:
9
10             mydb["collections"].insert_one(
11                 {
12                     "id": collection.id,
13                     "can_read": collection.can_read,
14                     "can_write": collection.can_write,
15                     "description": collection.description,
16                     "title": collection.title
17                 }
18             )
19
20             try:
21
22                 objects = collection.get_objects()
23                 for object in objects[ objects ]:
24
25                     try:
26                         mydb["manifests"].insert_one({
27                                 "id": object[ id ],
28                                 "date_added" : strftime("%Y-%m-%dT%H:%M:%S.%fZ",
29                                 gmtime()),
30                                 "versions": [strftime("%Y-%m-%dT%H:%M:%S.%fZ",
31                                 gmtime())],
32                                 "media_types": ["application/vnd.oasis.stix+json;
33                                 version=2.0"],
34                                 "_collection_id": collection.id,
35                                 "_type": object[ type ]
36                             })
37                     except:
38
39                         try:
40                             mydb["manifests"].insert_one({
41                                 "id": object[ id ],
42                                 "date_added": strftime("%Y-%m-%dT%H:%M:%S.%fZ",
43                                 gmtime()),
44                                 "media_types": ["application/vnd.oasis.stix+json;
45                                 version=2.0"],
46                                 "_collection_id": collection.id,
47                                 "_type": object[ type ]
```

```
48                                  })
49                              except:
50                                  print "Objects Error"
51
52                          obj = json.dumps(object)[:-1] +  , "_collection_id": "  +
53                          str(collection.id) +  "}
54                          soon = json.loads(obj)
55                          mydb["objects"].insert_one(soon)
```

Observa-se na Listing 4.3 que na linha 1 está a função "insertDataMongoDB"que recebe os parâmetros: nome do índice, url do mongoDB, url do servidor TAXII, usuário e senha, para acessar esse servidor. Na linha 3 é realizada uma solicitação de extração de dados no servidor TAXII passado como argumento, sendo a solicitação feita através de um cliente TAXII e os dados recebidos enviados para a variável "servidor". Na linha 5 é realizada uma conexão no MongoDB, no índice e pela url enviados como parâmetros. As linhas 7 e 8 são estruturas repetitivas implementadas para acessar os dados extraídos do servidor TAXII. Da linha 10 à linha 17 são extraídos os dados pertencentes às coleções e inseridas dentro da base "collection". Da linha 22 à linha 50 são extraídos os dados pertencentes aos objetos. Da linha 52 à linha 55 esses objetos são inseridos dentro da base "objects".

A ferramenta proporciona alguns mecanismos de realizar a inserção de dados: Servidor STIX v2 e Inserção de objeto SDO e SRO.

4.1.3.1 Servidor STIX

Para ter acesso a dados provenientes do *feed* padrão TAXII, é necessário ter o IP (ou URL) do *feed* e credenciais válidos dentro da aplicação desejada. Passando esses dados e especificando um índice para armazenar os registros contidos no *feed* de interesse, a aplicação, utilizando um cliente TAXII, vai consumir o conteúdo desse *feed* de forma automática. A Figura 4.4 apresenta a tela de inserção por servidor TAXII pelo Minerva.

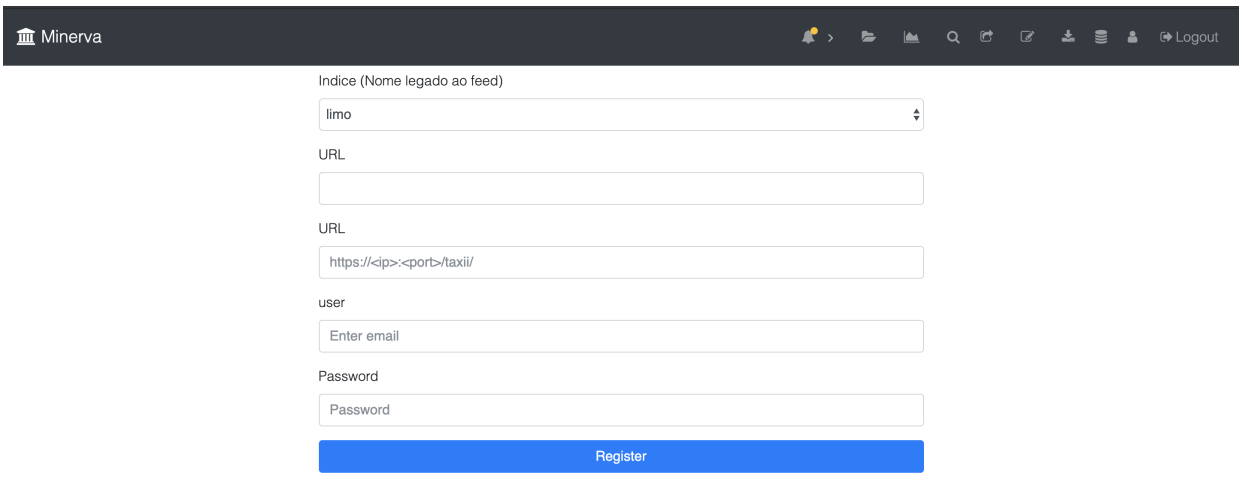

Figura 4.4: Inserção de dados provenientes de servidor TAXII pelo Minerva

4.1.3.2 Inserção de objeto SDO e SRO

A ferramenta também concede ao analista a possibilidade de inserir seus próprios objetos SDO ou SRO. Uma vez que o usuário já possua esses dados, ele será capaz de inseri-los manualmente. A Figura 4.5 apresenta uma tela de registro do SDO de "*attack*".

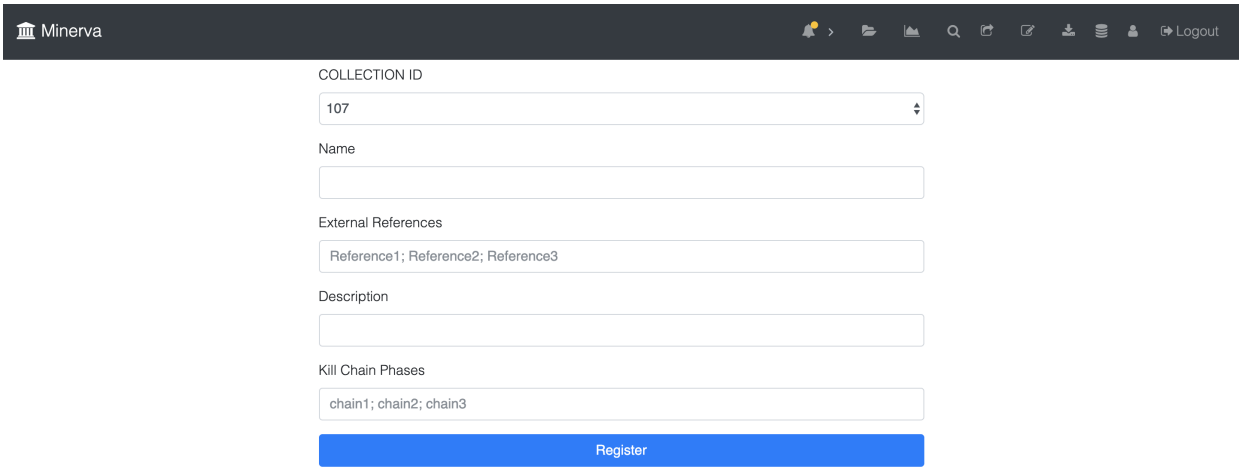

Figura 4.5: Cadastro de SDO de *Attack* no Minerva

4.1.4 Manipulação de dados

Essa funcionalidade foi implementada visando disponibilizar um mecanismo que permite a manipulação de dados vinculado ao padrão STIX v2, o formato que os dados estão armazenados. O Padrão STIX v2 divide os dados armazenados em 2 tipos, coleções e objetos, ambos possuindo mecanismos próprios para seu manuseio.

4.1.4.1 coleções

As coleções são um conjunto de objetos. O Minerva possui a capacidade de manipular essa estrutura. O código para manipulação de coleções armazenadas no MongoDB é mostrado no Listing 4.4.

Listing 4.4: Função de coleta de dados por URL e inserção no MongoDB

```
1 def getAllCollectionFromAPI(urlDB, nameAPI, offset):
2
3     mydb = selectDatabaseMongoDb(nameAPI, urlDB)
4
5     starting_id = mydb[ collections ].find().sort( _id , pymongo.ASCENDING)
6     total = starting_id.count()
7     last_id = starting_id[int(offset)][ _id ]
```

```
8
9       collections = []
10      for collectionTMP in collectionsTmp:
11          collections.append(Collection(collectionTMP[ id ], collectionTMP[ title ],
12          collectionTMP[ description ]))
13      return collections, total
```

Observa-se no Listing 4.4 que a linha 1 contém a função "getAllCollectionFromAPI", responsável por trazer todas as coleções armazenadas no MongoBD, usando como argumentos a url do banco, de qual índice serão extraídas as coleções e uma variável que verifica a quantidade de coleções. Nas linhas 5, 6 e 7 as coleções são extraídas do banco. Nas linhas 9, 10, 11 e 12 as coleções são organizados dentro de uma variável. Na linha 13 essa variável é retornada a quem chamar essa função.

A aplicação possibilita o manuseio de coleções específicas. A Figura 4.6 apresenta uma tela com a manipulação de coleções.

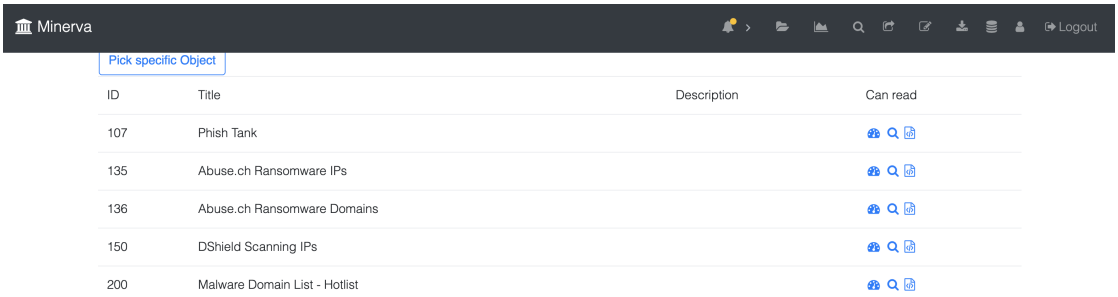

Figura 4.6: Apresentação de coleções no Minverva

4.1.4.2 objetos

Os objetos são registros que devem ser agrupados dentro de uma coleção. Os objetos podem ser inseridos manualmente dentro de uma coleção, ou podem ser coletados junto a coleção de um *feed* especifico. O Listing 4.5 é mostrado a função que realiza a coleta de objetos armazenados dentro dos índices.

Listing 4.5: Função que busca objetos armazenados dentro de índices

```
1  def getAllObjectsFromCollection(urlDB, nameAPI, collectionID, offset):
2      mydb = selectDatabaseMongoDb(nameAPI, urlDB)
3
4      starting_id = mydb[ objects ].find({"_collection_id": collectionID})
5      .sort( _id , pymongo.ASCENDING)
6      total = starting_id.count()
7      last_id = starting_id[int(offset)][ _id ]
8      objectsTmp = mydb[ objects ].find({"_collection_id": collectionID,  _id :
```

```
9
10      objects = []
11      for object in objectsTmp:
12          object[ _id ] = str(object[ _id ])
13          objects.append(Object(object[ _id ],json.dumps(object)))
14      return objects, total
```

Observa-se no Listing 4.5 que a linha 1 contém a função "getAllObjectsFromCollection", responsável por trazer todas os objetos armazenados no MongoBD que pertencem a uma coleção passada como argumento, junto com a url do banco e uma variável que verifica a quantidade de objetos. Na linha 2 é localizado o nome da base dados onde encontram-se os objetos. Nas linhas 4, 5, 6, 7 e 8, os objetos são extraídos do banco. Nas linhas 10, 11, 12, 13, os objetos são organizadas dentro de uma variável e na linha 14 essa variável é retornada a quem chamar essa função.

A Figura 4.7 apresenta uma tela padrão de listagem de objetos no Minerva.

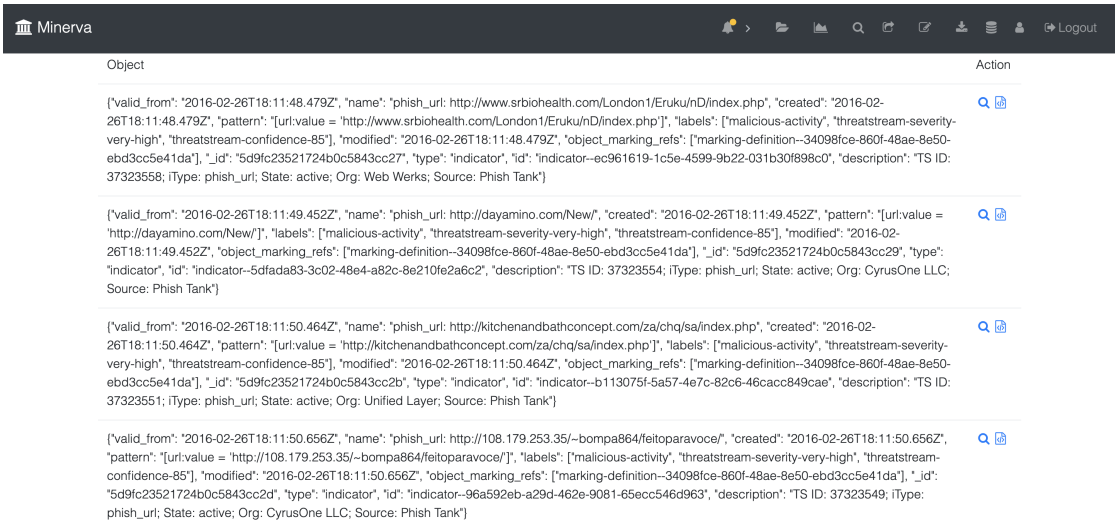

Figura 4.7: Listagem de objetos

4.1.5 Análise de dados

Para auxiliar o investigador na manipulação dos dados armazenados, foram desenvolvidos os seguintes mecanismos: localização e visualização detalhada de objeto específico, e mapa de vínculos.

4.1.5.1 Localização e visualização detalhada de objeto específico

O Minerva possui a funcionalidade voltada para a localização de um objeto específico em meio a massa de dados armazenada no MongoDB. Devido a grande quantidade de dados armazenados em ferramentas de CTI, essa funcionalidade permite economizar tempo na busca por registros

específicos. A Figura 4.8 apresenta a tela de inserção de ID, para buscar o registro requerido.

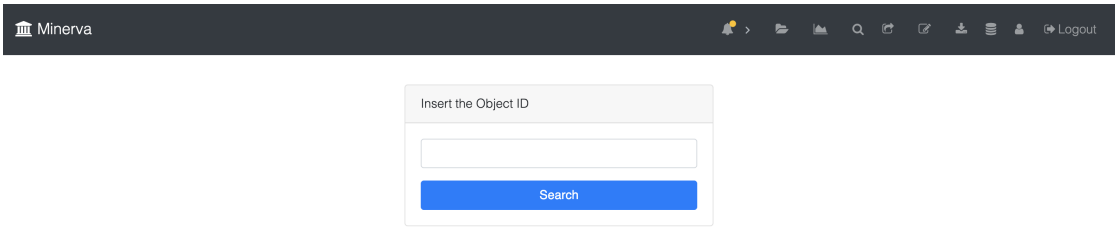

Figura 4.8: Busca por registro no MINERVA

Uma vez encontrado, o registro será apresentado de forma detalhada. Permitindo também a cópia desse objeto para outra coleção, ou apagá-lo. A Figura 4.9 ilustra a visibilidade de um registro no Minerva.

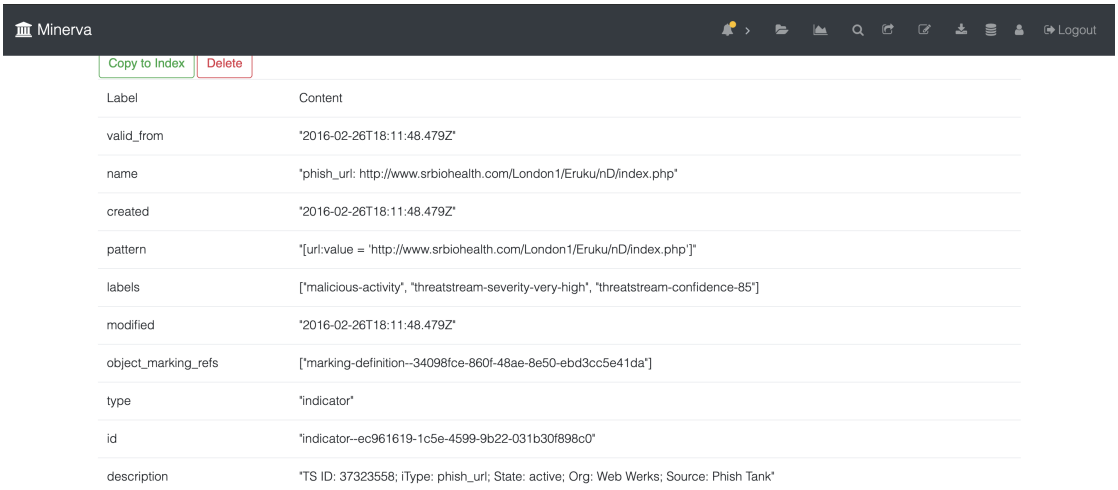

Figura 4.9: Detalhes de registro no MINERVA

4.1.5.2 Mapa de vínculos

Para facilitar a compreensão do relacionamento de objetos do tipo SDOs, mapeada pelo registro de objetos do tipo SRO, foi implementado um gráfico que proporciona essa visibilidade. O Mapa de vínculos, como é conhecido, traz uma visão completa de todos os objetos dentro da coleção, proporcionando uma visão holística de seus relacionamentos. A Figura 4.10 apresenta uma tela da aplicação contendo esse mapa projetado pelo Minerva.

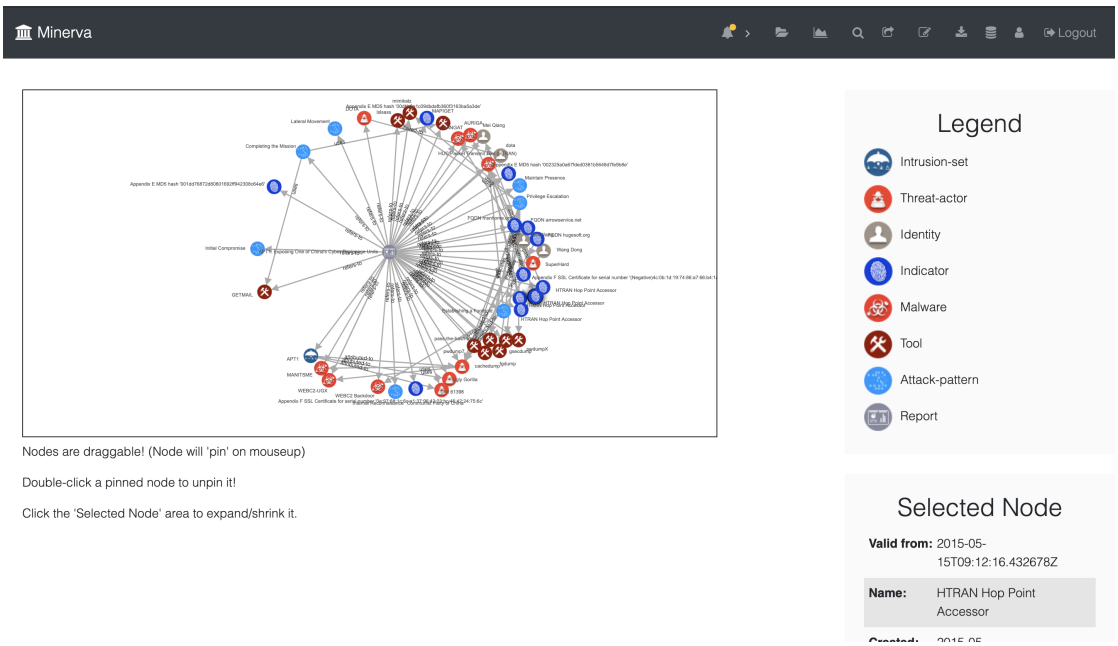

Figura 4.10: Mapa de vínculos projetado pelo Minerva

Sem essa visibilidade, seria necessário buscar uma forma de elaborar a percepção a partir de dados no formado JSON (que estrutura o Stix v2), conforme apresentado na Figura 4.11.

Figura 4.11: Dados estruturados no padrão STIX v2

4.1.6 Compartilhamento de dados

O compartilhamento de dados é uma das funcionalidades mais interessantes implementadas pelo Minerva, e o que o diferencia de outras soluções. Para isso, foi implementado um servidor web separado da infraestrutura que suporta a aplicação. O código que implementa a inicialização desse servidor e seus serviços é mostrado na Listing 4.6

Listing 4.6: Função principal, responsável por inicializar o servidor WEB

```
1  def main():
2      medallion_parser = _get_argparser()
3      medallion_args = medallion_parser.parse_args()
4      log.setLevel(medallion_args.log_level)
5
6      set_config(application_instance, confiUsers["users"])
7      init_backend(application_instance, configuration["backend"])
8      register_blueprints(application_instance)
9
10     application_instance.run(
11         host=medallion_args.host,
12         port=medallion_args.port,
13         debug=medallion_args.debug_mode
14     )
```

Observa-se no Listing 4.6 que nas linhas 2, 3 e 4 são coletadas as configurações do servidor. Nas linhas 6, 7 e 8 são verificadas as tabelas "Users"e "backend"do MongoDB, registrando-as no controlador de sessões. Nas linhas 10 a 14 o serviço é inicializado.

O servidor de compartilhamento foi programado para responder a requisições no formato TAXII. O código que implementa a função que constrói a chamada de informações gerais da API é mostrado no Listing 4.7.

Listing 4.7: Função que constrói resposta a requisição TAXII

```
1      @catch_mongodb_error
2      def server_discovery(self):
3          discovery_db = self.client["discovery_database"]
4          collection = discovery_db["discovery_information"]
5          pipeline = [{
6              "lookup": {
7                  "from": "api_root_info",
8                  "localField": "api_roots",
9
10                 "foreignField": "_name",
11                 "as": "roots"
12             }
13         }, {
14             "project": {
15                 "_id": 0,
16                 "title": 1,
17                 "description": 1,
18                 "contact": 1,
19                 "api_roots": "roots._url"
20             }
21         }]
22         info = list(collection.aggregate(pipeline))[0]
```

```
23          return info
```

Observa-se no Listing 4.7 que na linha 1 é feita a chamada de um método que vai garantir que, em caso de erro, a aplicação não pare de funcionar e uma mensagem de erro seja passada ao usuário. Na linha 2 é definida a função "server_discovery"que apresenta uma lista de todas as coleções armazenadas na base da dados. Na linha 3 são extraídos os dados da base de dados "discovery_database"que possui essa lista. Na linha 4 é coletado da base dados "discovery_information"que tem informações das coleções. Nas linhas 5 a 21 essas informações são estruturadas em formato JSON. Nas linhas 22 e 23 esses dados são retornados para o objeto que realiza a chamada da função.

A forma como a estrutura do Minerva foi planejada e implementada, permite a criação do vínculo de um usuário com a capacidade de extrair dados armazenados dentro de índices específicos, de fora da infraestrutura que armazena a ferramenta. A Figura 4.12 mostra a tela de cadastro de compartilhamento que permite a gestão dos índices vinculados.

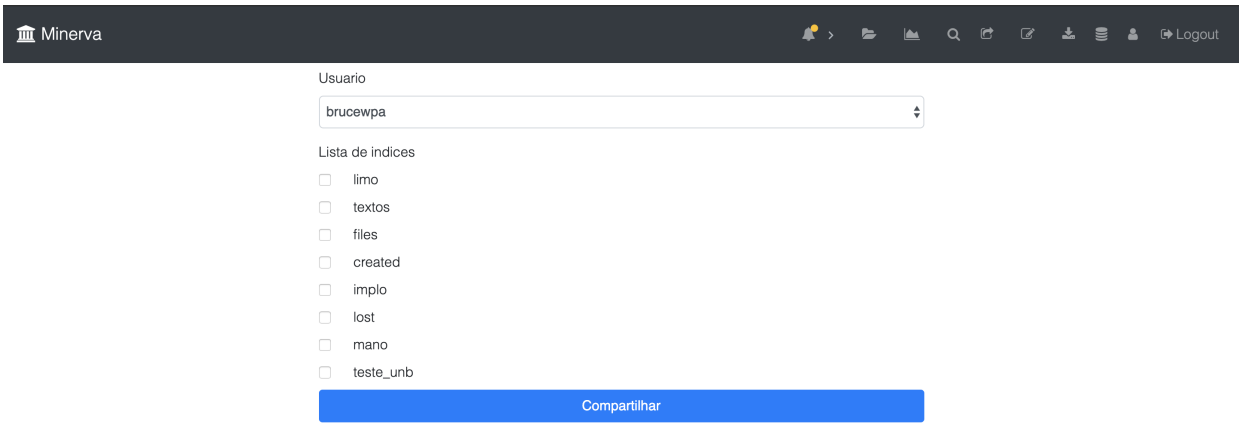

Figura 4.12: Cadastro de compartilhamento no Minerva

Uma vez com a permissão estabelecida, o usuário será capaz de acessar o conteúdo desse *feed* de qualquer ferramenta que realize requisições web no padrão TAXII (cliente), recebendo uma resposta em formato JSON, no padrão STIX v2. A Figura 4.13 apresenta um resultado dessa solicitação a partir de um navegador (*browser*) web.

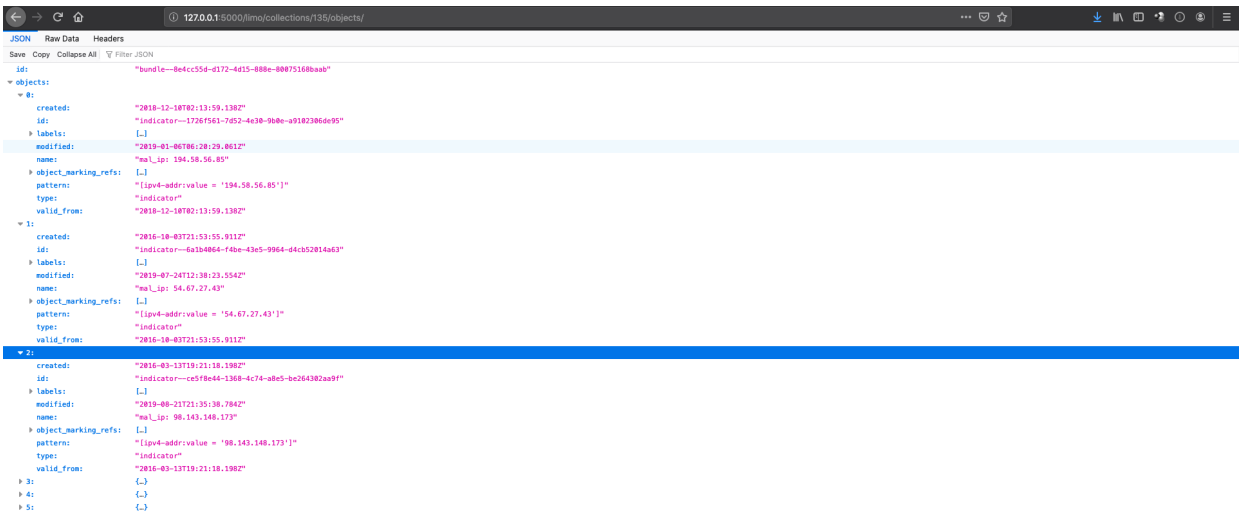

Figura 4.13: Resposta de solicitação ao servidor web

4.2 COMPARAÇÃO COM OUTRAS FERRAMENTAS

Neste trabalho foi feita uma comparação da ferramenta Minerva com outras ferramentas abertas e/ou gratuitas de CTI. Elas tiveram seus ambientes ideais (conforme documentação de cada uma) replicados, e foram manuseadas e entendidas. As ferramentas que passaram por esse procedimento foram:

- STAXX[1]
- YETI[2][3]
- MISP[4][5]
- OpenCTI[6][7]

Dentre as funcionalidades apresentadas, algumas excedem o usual, sendo implementadas apenas em ferramentas específicas. A comparação feita neste trabalho foca nessas funcionalidades, destacando quais são as ferramentas que as implementam.

[1] <https://www.anomali.com/community/staxx>
[2] <https://yeti-platform.github.io/>
[3] <https://github.com/yeti-platform/yeti>
[4] <https://www.misp-project.org/>
[5] <https://github.com/MISP/MISP>
[6] <https://www.opencti.io/en/>
[7] <https://github.com/OpenCTI-Platform/opencti>

4.2.1 Inserção de dados

A funcionalidade de inserção de dados proveniente de servidores TAXII, necessitam de um cliente que utiliza o padrão TAXII. Essa funcionalidade é implementada pelo STAXX e o pelo MISP, porém não são pelo YETI o OpenCTI. A Figura 4.14 apresenta a inserção de dados a partir do STAXX.

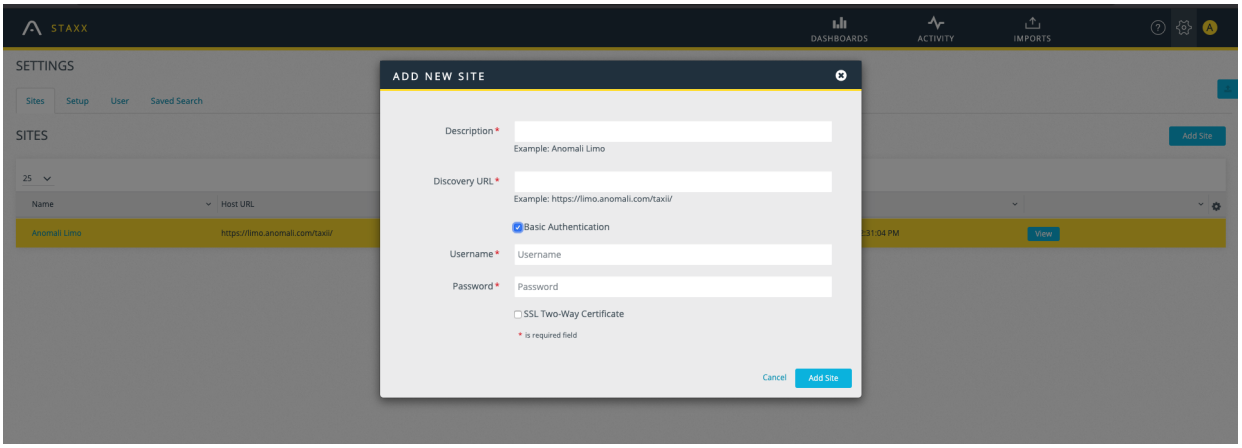

Figura 4.14: Inserção de dados provenientes de servidor TAXII pelo Minerva

4.2.2 Inserção de objeto SRO ou SDO

Essa funcionalidade permite ao usuário registrar um objeto diretamente na coleção. A única ferramenta que permite essa inserção é a YETI, porém com objetos SRO's e SDO's compartíveis exclusivamente com a versão 1 do STIX. O Minerva também possibilita essa inserção, porém se diferencia, pois trabalha diretamente com a versão 2 do STIX. Que é mais completa, pois possui um número maior de objetos. A Figura 4.15 apresenta uma tela de inserção de registro SDO de "malware", dentro do YETI.

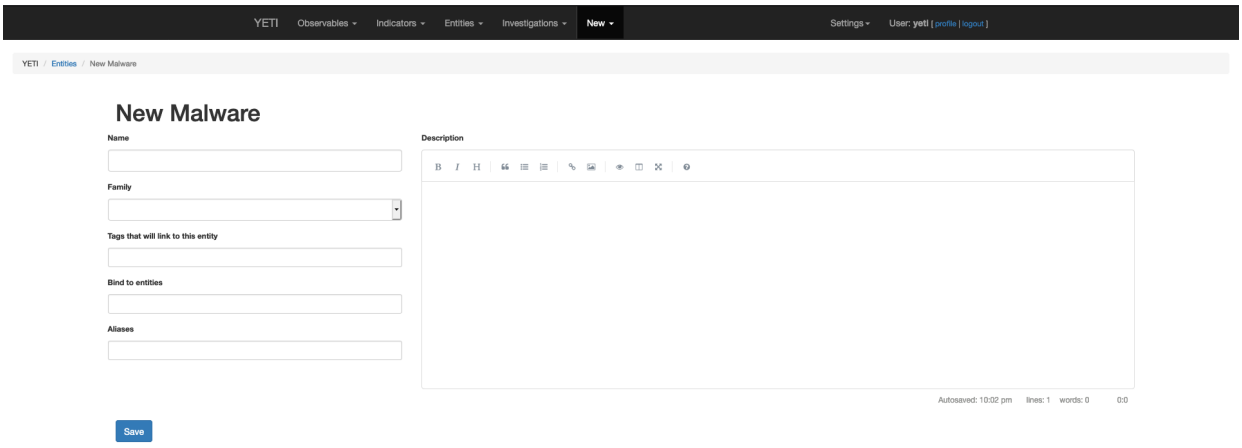

Figura 4.15: Cadastro de SDO de Malware no YETI

4.2.3 Mapa de vínculos

O mapa de vínculos provê uma visibilidade holística sobre os objetos armazenados, sendo considerada pelo Mitre uma das melhores maneiras de analisar dados de CTI. Do jeito que foi implementado, somente o YETI possui uma funcionalidade semelhante. Porém, como o YETI foi concebido no padrão STIX v1, tendo visibilidade gerada é mais limitada aos objetos que suporta. A visibilidade gerada pelo mapa de vínculo do Minerva apresenta mais objetos, pois trabalha na versão V2. A Figura 4.16 apresenta uma tela do mapa de vínculos implementada pelo YETI.

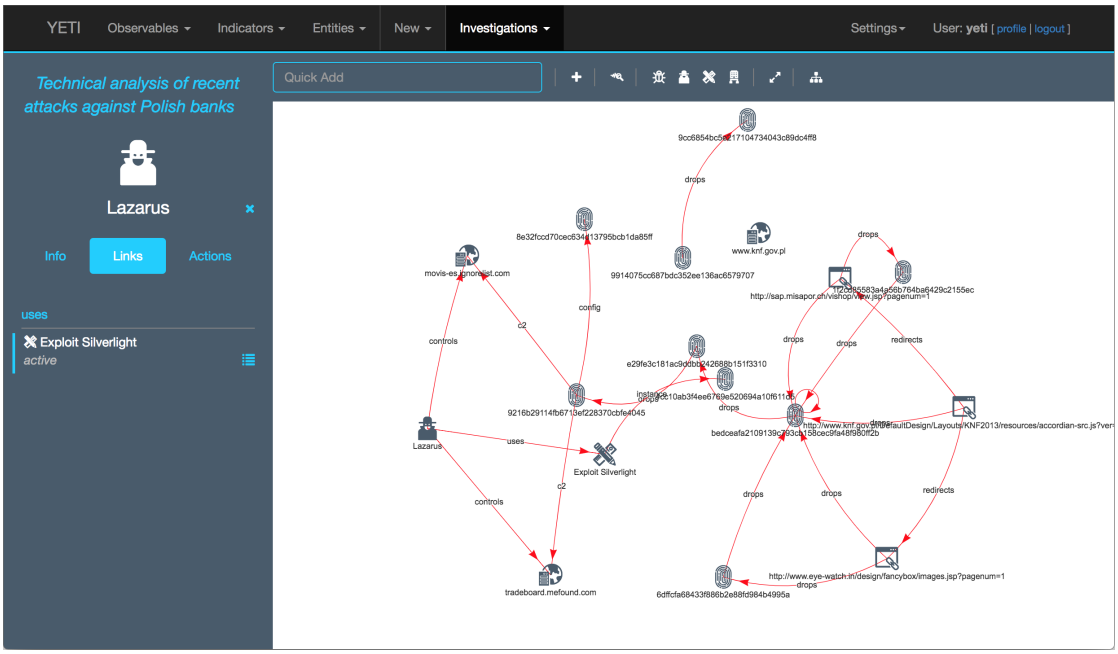

Figura 4.16: Mapa de vínculos no YETI

4.2.4 Compartilhamento

Das ferramentas analisadas, a única que possui a funcionalidade de compartilhamento de dados foi a MISP. Ela disponibiliza o acesso a todos os dados armazenados a usuários que estão no grupo que possui permissão. Na MISP não existe maneira de filtrar quais usuários acessam quais dados. O Minerva, por sua vez, teve sua estrutura construída para permitir o gerenciamento granular dos usuários e dos dados compartilhados. A Figura 4.17 apresenta a configuração dessa funcionalidade dentro do MISP.

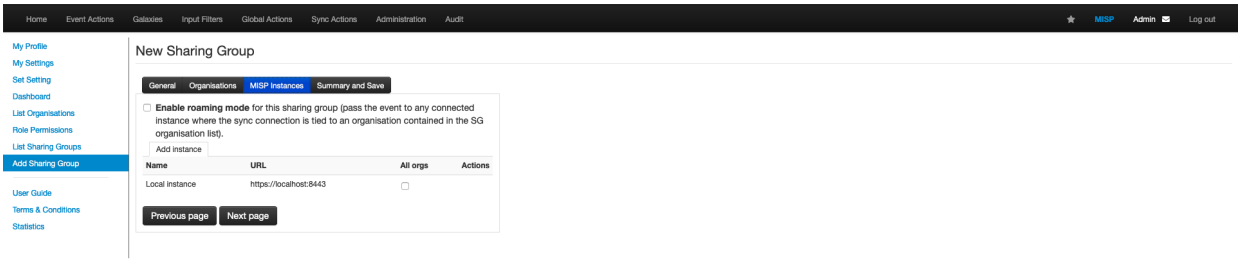

Figura 4.17: configuração de compartilhamento de dados no MISP

5 CONCLUSÃO

A utilização da Internet como meio para a construção e disseminação de táticas, técnicas e procedimento, utilizados para causar danos a ativos pertencentes a pessoas, organizações e até nações, é visto com frequência. Devido a ser uma fonte aberta e abrangente, seu uso cominou na diminuição significava no tempo necessário para a descoberta de uma brecha de segurança, até a tentativa de explorá-la. Os ativos de segurança convencionais não têm sido eficazes contra ataques avançados, que utilizam várias técnicas, e se modificam mais rápido que o tempo necessário para sua análise.

Ferramentas de CTI foram criadas para atuar nessa debilidade, utilizando as mesmas fontes que usuários mal-intencionados (a internet), porém buscando gerar inteligência operacional e estratégica, para as organizações apoiarem seus planejamentos contra ameaças.

Inteligência de ameaças cibernéticas é um assunto majoritariamente discutido no âmbito industrial, existindo já várias ferramentas que atuam nesse contexto, trabalhando para a criação e compartilhamento de inteligência para suas comunidades. No âmbito acadêmico, a maioria dos trabalhos detalham funcionalidades existentes em ferramentas, propondo outras abordagens ou descrevendo o por quê da sua necessidade.

Buscando proporcionar uma visão holística das funcionalidades e características encontradas nas ferramentas atuantes no âmbito de CTI, juntamente com os artigos acadêmicos que fundamentem esse tema, esta dissertação apresentou uma proposta de modelo de referencia voltada para o desenvolvimento de aplicações de CTI. Sua estrutura foi dividida em 6 camadas, inspirada na análise de diversas ferramentas e documentos, com o intuito de promover entendimento sobre seu escopo, esclarecer sobre o ciclo de vida dos dados para a produção de valor, auxiliar o entendimento sobre na real necessidade de seu uso e servir como fundamento para a auditoria e planejamento de ferramentas.

Visando comprovar a real eficácia do modelo proposto, uma ferramenta de inteligência de ameaças cibernéticas demonstrativa foi desenvolvida. Todas as suas funcionalidades e seu comportamento, foram prospectados baseados no modelo. Seus resultados, ainda que limitados, foram bem expressivos, possibilitando o compartilhamento de inteligência, utilizando padrão aberto, e de forma gerenciável granularmente, funcionalidade não encontrada em nenhuma outra ferramenta do tipo, assim agregando ao campo de conhecimento.

A ferramenta foi planejada para trabalhar em 5 das 6 camadas propostas pelo modelo. Pontos estratégicos do código utilizado em seu desenvolvimento foram descritos nesse documento, juntamente com a estrutura em que esse desenvolvimento foi realizado, e os casos de uso ofertados pela ferramenta.

A questão do compartilhamento de dados, tratado na construção da ferramenta, é uma maneira de se eliminar os *gaps* de conhecimento entre os usuários mal-intencionados e aqueles que buscam

proteger seus recursos. As iniciativas que implementam essas funcionalidades, a fazem apenas para funcionar dentro da sua comunidade. Existe a necessidade da idealização de iniciativas que busquem a criação de métodos para o compartilhamento de dados, de forma aberta e colaborativa. Este trabalho apresentou uma resposta a essa carência.

O trabalho proporcionou um modelo base para o planejamento e desenvolvimento de aplicações de inteligência de ameaças. E utilizando esse modelo, uma ferramenta de CTI que atua com forte foco em compartilhamento de dados foi criada. Até então, esse tipo de funcionalidade estava vedada somente a plataformas e formatos específicos, agora é possível compartilhar de forma granularmente gerenciável, com o uso de padrão aberto.

Os próximos passos dessa pesquisa, consiste na implementação da 6º camada do modelo de referência proposto e que não foi contemplada nessa versão demonstrativa, a camada de Geração, que consiste na inserção de dados provenientes da infraestrutura que suporta a aplicação.

A implementação dessa camada faz parte da estratégia que visa a automação dos processos necessários para a idealização das funcionalidades propostas pela ferramenta, juntamente com a sintetização de ferramentas abertas, com foco específico no mapeamento de ameaças e vulnerabilidades, buscando consolidar o máximo de *feeds* disponíveis.

O objetivo é automatizar a varredura de vulnerabilidades, com base contextualizada no tráfego de dados e *logs* de eventos, proporcionados pela infraestrutura vigente. Utilizando ferramentas abertas para esse fim, e consultando *feeds* externos sobre os dados obtidos. Visando gerar conhecimento de inteligência de forma automatizada, e baseada em evidências. Presando pela visibilidade e manipulação dos dados armazenados, para melhor atender o investigador.

Referências Bibliográficas

1 POUSHTER, J. et al. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*, Washington, DC, USA:, v. 22, p. 1–44, 2016.

2 BREWER, R. Advanced persistent threats: minimising the damage. *Network security*, Elsevier, v. 2014, n. 4, p. 5–9, 2014.

3 SAMTANI, S.; CHINN, K.; LARSON, C.; CHEN, H. Azsecure hacker assets portal: Cyber threat intelligence and malware analysis. In: IEEE. *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. [S.l.], 2016. p. 19–24.

4 GRAU, D.; KENNEDY, C. Tim lecture series–the business of cybersecurity. *Technology Innovation Management Review*, v. 4, n. 4, 2014.

5 CAVELTY, M. D. Cyber-security. *The routledge handbook of new security studies*, Routledge New York, NY, p. 154–162, 2010.

6 CHERTOFF, M.; SIMON, T. The impact of the dark web on internet governance and cyber security. 2015.

7 MACAULAY, T. *System and method for generating and refining cyber threat intelligence data*. [S.l.]: Google Patents, 2015. US Patent 9,118,702.

8 LIAO, X.; YUAN, K.; WANG, X.; LI, Z.; XING, L.; BEYAH, R. Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. [S.l.: s.n.], 2016. p. 755–766.

9 MALIK, J. Threat intelligence sharing: The only way to combat our growing skills gap. *Information Security Magazine.(May 2016)*, 2016.

10 BROMILEY, M. Threat intelligence: What it is, and how to use it effectively. *SANS Institute InfoSec Reading Room*, v. 15, 2016.

11 SAUERWEIN, C.; SILLABER, C.; MUSSMANN, A.; BREU, R. Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. 2017.

12 ABU, M. S.; SELAMAT, S. R.; ARIFFIN, A.; YUSOF, R. Cyber threat intelligence–issue and challenges. *Indonesian Journal of Electrical Engineering and Computer Science*, v. 10, n. 1, p. 371–379, 2018.

13 MILLER, D.; SLATER, D. *Internet*. [S.l.]: Berg Publishers, 2000.

14 MICHELSON, A.; ROTHENBERG, J. Scholarly communication and information technology: exploring the impact of changes in the research process on archives. *The American Archivist*, JSTOR, v. 55, n. 2, p. 236–315, 1992.

15 DEIBERT, R. Growing dark side of cyberspace (... and what to do about it). *Penn St. JL & Int'l Aff.*, HeinOnline, v. 1, p. xiii, 2012.

16 WATNEY, M. M. The way forward in addressing cybercrime regulation on a global level. *Journal of Internet Technology and Secured Transactions (JITST)*, v. 1, p. 62–67, 2012.

17 NUNES, E.; DIAB, A.; GUNN, A.; MARIN, E.; MISHRA, V.; PALIATH, V.; ROBERTSON, J.; SHAKARIAN, J.; THART, A.; SHAKARIAN, P. Darknet and deepnet mining for proactive cybersecurity threat intelligence. In: IEEE. *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. [S.l.], 2016. p. 7–12.

18 BLUELIV. *Chasing cybercrime: network insights of Dyre and Dridex Trojan bankers.* [S.l.], 2014.

19 LITTLEJOHN, A. *The analysis of language teaching materials: Inside the Trojan Horse*. [S.l.]: na, 1998.

20 CHEN, P.; DESMET, L.; HUYGENS, C. A study on advanced persistent threats. In: SPRINGER. *IFIP International Conference on Communications and Multimedia Security*. [S.l.], 2014. p. 63–72.

21 RADACK, S. *Managing information security risk: organization, mission and information system view*. [S.l.], 2011.

22 TECHNOLOGIES, P. *hack at all costs - putting a price on APT attacks*. [S.l.], 2019.

23 BU, Z.; LIN, Y. *Zero-day discovery system*. [S.l.]: Google Patents, 2018. US Patent 10,133,863.

24 TADDEO, M. Deterrence by norms to stop interstate cyber attacks. *Minds and Machines*, Springer, v. 27, n. 3, p. 387–392, 2017.

25 MILES, C.; LAKHOTIA, A.; LEDOUX, C.; NEWSOM, A.; NOTANI, V. Virusbattle: State-of-the-art malware analysis for better cyber threat intelligence. In: IEEE. *2014 7th International Symposium on Resilient Control Systems (ISRCS)*. [S.l.], 2014. p. 1–6.

26 MAYBURY, M. T. Knowledge management at the mitre corporation. *MITRE Corporation, Bedford, MA, http://www. mitre. org/work/tech_papers/tech_papers_02/maybury_knowledge/maybury_km. pdf*, Citeseer, 2002.

27 CHO, S.; HAN, I.; JEONG, H.; KIM, J.; KOO, S.; OH, H.; PARK, M. Cyber kill chain based threat taxonomy and its application on cyber common operational picture. In: IEEE. *2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*. [S.l.], 2018. p. 1–8.

28 YADAV, T.; RAO, A. M. Technical aspects of cyber kill chain. In: SPRINGER. *International Symposium on Security in Computing and Communication*. [S.l.], 2015. p. 438–452.

29 MCMILLAN, R. *Definition: Threat Intelligence*. [S.l.], 2013.

30 AHLBERG, C. *The threat intelligence book - moving toward a security intelligence program*. [S.l.]: Cyberedge press, 2019.

31 HUTCHINS, E. M.; CLOPPERT, M. J.; AMIN, R. M. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, v. 1, n. 1, p. 80, 2011.

32 BHATT, S.; MANADHATA, P. K.; ZOMLOT, L. The operational role of security information and event management systems. *IEEE security & Privacy*, IEEE, v. 12, n. 5, p. 35–41, 2014.

33 WRIGHT, W.; SCHROH, D.; PROULX, P.; SKABURSKIS, A.; CORT, B. The sandbox for analysis: concepts and methods. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. [S.l.: s.n.], 2006. p. 801–810.

34 WANG, G.; HUO, Y.; MA, Z. M. Research on university's cyber threat intelligence sharing platform based on new types of stix and taxii standards. *Journal of Information Security*, Scientific Research Publishing, v. 10, n. 4, p. 263–277, 2019.

35 DANYLIW, R.; MEIJER, J.; DEMCHENKO, Y. et al. The incident object description exchange format. *IETF Request For Comments*, v. 5070, 2007.

36 BARNUM, S. Standardizing cyber threat intelligence information with the structured threat information expression (stix). *Mitre Corporation*, v. 11, p. 1–22, 2012.

37 CONNOLLY, J.; DAVIDSON, M.; SCHMIDT, C. The trusted automated exchange of indicator information (taxii). *The MITRE Corporation*, p. 1–20, 2014.

38 ASGARLI, E.; BURGER, E. Semantic ontologies for cyber threat sharing standards. In: IEEE. *2016 IEEE Symposium on Technologies for Homeland Security (HST)*. [S.l.], 2016. p. 1–6.

39 ZHAO, W.; WHITE, G. A collaborative information sharing framework for community cyber security. In: IEEE. *2012 IEEE Conference on Technologies for Homeland Security (HST)*. [S.l.], 2012. p. 457–462.

40 OHTA, T.; TAKENAKA, M.; KATOU, M.; MASUOKA, R.; KAYAMA, K.; FUKUSHIMA, N.; IMAI, H. Cybersecurity solutions for major international events. *FUJITSU SCIENTIFIC & TECHNICAL JOURNAL*, FUJITSU LTD 1015 KAMIKODANAKU NAKAHARA-KU, KAWASAKI, 211, JAPAN, v. 54, n. 4, p. 57–65, 2018.

41 ROSS, A. D.; MORGAN, D. M. *System security event notification aggregation and non-repudiation*. [S.l.]: Google Patents, 2009. US Patent 7,571,474.

42 PINKERTON, S. A federated model for cyber security. In: *Cyberspace Research Workshop, Shreveport, LA (November 2007)*. [S.l.: s.n.], 2007.

43 OFFICER, V. Security for business innovation council an industry. Citeseer, 2010.

44 GOVETT, I. R. *Client/server architecture supporting concurrent servers within a server with a transaction manager providing server/connection decoupling*. [S.l.]: Google Patents, 1998. US Patent 5,761,507.

45 MAGEE, J. C.; ANDREWS, A. M.; NICHOLSON, M. W.; JAMES, J. L.; LI, H. C.; STEVENSON, C. L.; LATHROP, J. *Collective threat intelligence gathering system*. [S.l.]: Google Patents, 2014. US Patent 8,813,228.

46 WAGNER, C.; DULAUNOY, A.; WAGENER, G.; IKLODY, A. Misp: The design and implementation of a collaborative threat intelligence sharing platform. In: *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*. [S.l.: s.n.], 2016. p. 49–56.

47 BROWN, S.; GOMMERS, J.; SERRANO, O. From cyber security information sharing to threat management. In: *Proceedings of the 2nd ACM workshop on information sharing and collaborative security*. [S.l.: s.n.], 2015. p. 43–49.

48 MOHAISEN, A.; AL-IBRAHIM, O.; KAMHOUA, C.; KWIAT, K.; NJILLA, L. Rethinking information sharing for threat intelligence. In: *Proceedings of the fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies*. [S.l.: s.n.], 2017. p. 1–7.

49 TOUNSI, W.; RAIS, H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & security*, Elsevier, v. 72, p. 212–233, 2018.

50 NACHENBERG, C.; WILHELM, J. *Detecting polymorphic threats*. [S.l.]: Google Patents, 2010. US Patent 7,739,740.

51 NOVOSELOVA, O.; ROMANOV, A.; ROMANOVA, L. Communicative construct of the composite threat-performatives. *Procedia-Social and Behavioral Sciences*, Elsevier, v. 206, p. 71–75, 2015.

52 ROBERTSON, J.; DIAB, A.; MARIN, E.; NUNES, E.; PALIATH, V.; SHAKARIAN, J.; SHAKARIAN, P. *Darkweb cyber threat intelligence mining*. [S.l.]: Cambridge University Press, 2017.

53 SINNEMA, R. *Risk-adaptive access control of an application action based on threat detection data*. [S.l.]: Google Patents, 2018. US Patent 9,992,213.

54 WHEELUS, C.; BOU-HARB, E.; ZHU, X. Towards a big data architecture for facilitating cyber threat intelligence. In: IEEE. *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. [S.l.], 2016. p. 1–5.

55 FRANKLIN, L.; PIRRUNG, M.; BLAHA, L.; DOWLING, M.; FENG, M. Toward a visualization-supported workflow for cyber alert management using threat models and human-centered design. In: IEEE. *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)*. [S.l.], 2017. p. 1–8.