INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Aquisição e Modelação de *Threat Intelligence* para Desenvolver um Sistema de Reputação

Mário João Amaro da Costa

Mestrado em Informática e Gestão

Orientador:
Prof. Doutor João Carlos Silva, Professor Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Coorientadora:
Prof.ª Doutora Maria Pinto-Albuquerque, Professor Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Novembro, 2022

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Aquisição e Modelação de *Threat Intelligence* para Desenvolver um Sistema de Reputação

Mário João Amaro da Costa

Mestrado em Informática e Gestão

Orientador:
Prof. Doutor João Carlos Silva, Professor Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Coorientadora:
Prof.ª Doutora Maria Pinto-Albuquerque, Professora Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Novembro, 2022

Dedicado ao capítulo mais bonito da minha vida, a Família.

Agradecimentos

À minha namorada e companheira, Sara Gonçalves, pelo seu apoio incondicional para me ajudar a alcançar mais uma etapa da minha vida. Muito do que sou hoje é graças a ti. Obrigado por tornares tudo tão mais fácil.

Ao meu filho, que sem compreender devido à sua tenra idade, permitiu partilhar alguns momentos da sua atenção com a presente dissertação. Obrigado por todos os abraços e beijinhos para me manteres motivado.

À minha filha recém-nascida, a Olivia, pelo amor e motivação que veio acrescer.

Aos meus pais e à minha avó por todos os esforços que fizeram para me proporcionarem um futuro melhor. Um especial agradecimento ao meu avô, que apesar de não estar fisicamente entre nós, certamente estará orgulhoso por mais uma etapa da minha vida.

Ao Rodrigo Valente pela camaradagem e amizade, pois temos sido, literalmente, irmãos de armas. A nossa entreajuda e competição saudável tem sido fundamental para alcançarmos objetivos pessoais e coletivos. Só assim faz sentido o nosso mantra, "missão dada, é missão cumprida". Espero mantermos esta camaradagem durante longos anos, e que o futuro seja igualmente risonho.

Ao professor João Carlos Silva e à professora Maria Pinto Albuquerque por acreditarem neste projeto e aceitarem contribuir com todo o seu conhecimento para um trabalho melhor. Ambos foram essenciais para aprimorar este trabalho e aumentar o rigor científico do mesmo.

À equipa de cibersegurança da OutSystems por proporcionar a ideia deste projeto.

Ao Luís Paulino que é o rosto deste projeto na OutSystems. O Luís foi incansável durante todo este trabalho, esteve sempre disponível para tirar dúvidas e para dar novas ideias quando nos deparávamos com problemas, foi graças a ele que este projeto teve sucesso. Um sincero obrigado.

Ao Igor Antunes (OutSysems) pelo tempo disponibilizado para transmitir conhecimentos técnicos e para solucionar problemas desse cariz.

Ao ISTAR-IUL – Centro de Investigação em Ciências da Informação, Tecnologia e Arquitetura por ter proporcionado condições para apresentar este projeto em Aveiro, no âmbito do evento *INNCYBER Innovation Hub – 3rd Edition*.

Resumo

A internet é a tecnologia crucial da Era da Informação, pois permite melhorar o desempenho das organizações e agilizar processos de negócio. A pandemia que marcou a segunda década do século XXI, a COVID-19, veio reforçar esta situação, pois fez com que o teletrabalho se tornasse uma realidade na generalidade das organizações, resultando num crescimento exponencial dos dispositivos conectados às redes das organizações. Consequentemente, os dispositivos vulneráveis a ataques, bem como os pontos de acesso à rede aumentaram, como tal a segurança da informação, das infraestruturas digitais e a forma como são armazenados os dados, têm gerado uma preocupação crescente no seio das organizações.

Paralelamente, a *threat intelligence* aplicada no âmbito da cibersegurança é preponderante, pois permite partilhar dados sobre indicadores de compromisso com o objetivo de mitigar ameaças, bem como minimizar o impacto das ameaças do dia zero nos sistemas de informação.

O presente trabalho visa o desenvolvimento de um modelo preciso e robusto para calcular a reputação de ameaças, tendo como base a *threat intelligence*. Desta forma, foi desenvolvido um conector compatível com a plataforma OpenCTI, utilizada para recolher e partilhar informações sobre as ameaças. Este conector permite recolher dados de plataformas externas e, através de um algoritmo, avaliar o nível de ameaça (*ThreatScore*) do indicador de compromisso, bem como o nível de confiança (*TrustRating*) da pontuação atribuída. A *framework* desenvolvida é de prevenção de ameaças, ou seja, é um mecanismo complementar às defesas da organização para a tomada de decisão.

Palavras-chave: Ameaças, Cibersegurança, Reputação, *Score, Threat Intelligence*, Indicadores de Compromisso.

Abstract

Internet is the crucial technology of the information age. It improves company's performance and speeds up the business process. The pandemic situation that marked the second decade of the 21st century, COVID-19, reinforced this situation, many public and private organizations implemented teleworking, resulting in an exponential growth of devices connected to organizations networks. Therefore, devices vulnerable to attacks, as well as network access points, have increased, this generated a growing concern within organizations, about the security of information, digital infrastructures and the way in which data are stored.

At the same time, threat intelligence applied to the cybersecurity is beginning to be predominant, as it allows sharing data about indicators of compromise (IoC) with the aim of mitigating threat risks, as well as minimizing the impact of zero-day vulnerability to steal vital and sensitive data from the companies.

In the present work, we focus on developing a lightweight and accurate model to calculate a reputation score, based in the acquisition of threat intelligence. In this way, a compatible connector was developed for the OpenCTI platform, this platform is used to collect and share information about threats. The developed connector allows collecting data from external platforms and using an algorithm to calculate the threat level (*ThreatScore*) of the indicator of compromise analyzed, as well as the confidence level (*TrustRating*) of the assigned score. This framework is designed to complement, not to replace, cybersecurity program and risk management processes, providing credible information for decision making.

Keywords: Cybersecurity, Reputation, Score, Threats, Threat Intelligence, Indicators of Compromise.

Índice

Índice de Figuras

Índice de Quadros

Acrónimos

ENISA	Agência da União Europeia para Cibersegurança
AIPRA	*Automated IP Reputation Analyzer Tool*
ANS	*Autonomous System Number*
ANSSI	*Agence Nationale de la Sécurité des Systèmes d'Information*
API	Interface de Programação de Aplicação
APT	*Advanced Persistent Threat*
AWS	Amazon Web Services
BGP	*Border Gateway Protocol*
BPMN	*Business Process Model and Notation*
CERT-EU	*Computer Emergency Response Team of the European Union*
CTI	*Cyber threat intelligence*
DAbR	*Dynamic Attribute based Reputation*
DDOS	*Distributed Denial of Service*
DHCP	*Dynamic Host Configuration Protocol*
DLS	*Data Loss Protection*
DNS	*Domain Name System*
DSR	*Design Science Research*
ENISA	*European Union Agency For Cybersecurity*
Exploit	Códigos de *software* que permitem ganhar vantagem a partir de uma falha
HTTPS	*Hyper Text Transfer Portocol Secure*
HUMINT	*Human Intelligence*
IDS	*Intrusion Detection System*
IoA	Indicadores de ataque (*Indicators of Attack)*
IoC	Indicadores de Compromisso (*Indicators of Compromisse)*
IODEF	*Incident Object Description Exchange Format*
IP	Internet Protocol
IPS	*Intrusion Prevention System*
ISP	*Internet Service Provider*
ISP	Internet Service Provider
JSON	*JavaScript Object Notation*
log	*Ficheiros de registo*

MISP	*Malware Information Sharing Platform*
MuSeR	*Multi-observable session reputation* MuSeR
NPM	Métodos de Proteção das Redes (*Network Porection Methods)*
NPT	*Network Protection Tools*
OASIS	*Organization for the Advancement of Structured Informatiuon Standards*
OpenCTI	*Open Cyber Threat Intelligence*
OSINT	*Open Source Intelligence*
OTRS	*Open Source Ticket Request System*
P2P	*Peer-to-Peer*
Playbooks	Fragmentos de código ou *scripts*
PoP	*Pyramid of Pain*
PRI	Processos de Resposta a Incidentes
RAD	Plataforma de Desenvolvimento Rápido (*Rapid Application Delivery*)
SCOMINT	*Social Media Intelligence*
SEM	*Security Event Management*
SI	Sistemas de Informação
SIEM	*Security Incident Event Management*
SIM	*Security Information Management*
SOAR	*Security Orchestration, Automation and Response*
SOC	Centro de Operações de Segurança *(Security Operations Center)*
Software malicioso	*Malware*
STIX	*Structured Threat Information Expression*
TAXII	*Trusted Automated eXchange of Indicator Information*
TI	Tecnologias de Informação
TrustRating	Pontuação de Confiança
TS	*ThreatScore*
TTP	Táticas, Técnicas e Procedimentos
URL	*Uniform Resource Locator*
UUID	*Universally Unique Identifier*
VPN	*Virtual Private Network*
XML	*Extensive Markup Language*
Zettabyte	*$1 x10^{21}$ bytes*

CAPÍTULO 1

Introdução

Durante as duas últimas décadas, a internet tem desempenhado um papel preponderante na sociedade, na economia e nas infraestruturas, de tal forma que estas tornaram-se altamente dependentes desta tecnologia e dos computadores [1], [2]. Consequentemente, este avanço tecnológico, notório, torna-se num marco histórico de referência, pois permitiu criar redes de comunicação, interações e transações que se refletem no movimento de milhares de milhões de euros anualmente na economia global. Todavia, para que tal aconteça é necessário as infraestruturas e sistemas vitais estarem conectados ao ciberespaço, ou serem controlados neste, permitindo o fluxo de informações importantes e sensíveis [2], [3]. As necessidades anteriores aliadas à facilidade com que qualquer indivíduo tem acesso à internet para navegar no ciberespaço, a possibilidade de manter o anonimato nesta rede, o baixo custo e risco envolvidos em ações de caráter duvidoso, têm aliciado atores mal-intencionados, incluindo governos, grupos organizados e terroristas a desenvolverem atividades ilícitas no ciberespaço. Exemplos evidentes dessas situações são o crescimento da guerra cibernética, o crime cibernético, o terrorismo cibernético e a espionagem cibernética [1], [2].

De acordo com o mais recente estudo publicado pela empresa IBM[4], realizado a nível mundial, os custos médios devido a violação de dados bateram um recorde desde que existe este registo há 17 anos, pois ultrapassou os 4.20 milhões de dólares em 2021. Tendo em conta os dados do mesmo estudo, as causas mais frequentes para a violação de dados são credenciais comprometidas, ataques de *phishing* e configuração incorreta da *cloud,* com 20%, 17% e 15 % de frequências, respetivamente (Figura 1.1).

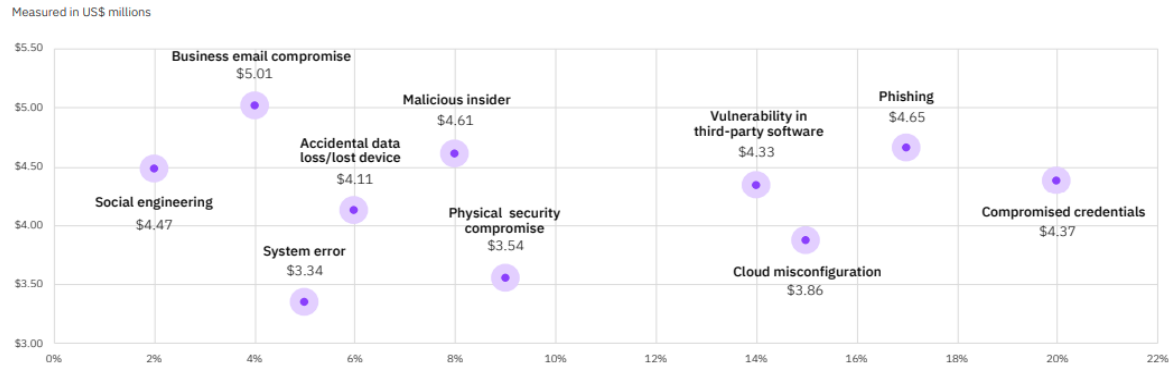

Figura 1.1 - Custos médios totais em milhões de dólares e frequência dos ataques responsáveis pela fuga de dados. Fonte: IBM Report "Cost of a Data Breach Report 2021"[4].

Em Portugal, a tendência crescente de ataques está concordante com as estatísticas a nível global, ou seja, o número de ataques tem vindo a crescer e as causas são idênticas às mencionadas a nível global, destacando-se o *phishing* e o *malware* (Figura 1.2) [5]. O compromisso de dados através da má configuração da *cloud* não consta nos lugares cimeiros a nível nacional, presumivelmente devido a muitas empresas nacionais não terem efetuado o processo de migração para a *cloud*.

2019				2020*					Ordenação	
RK	Tipo	Nº	%	RK	Tipo	Nº	% C/V	% S/V	Tendência absoluta %	Lugar RK
1º	*Phishing/smishing*	236	31	1º	*Phishing/smishing*	613	43	46	+ 160	=
2º	Infeção (*malware*)	123**	16	2º	Sistema infetado (*malware*)	169	12	13	+ 37	=
3º	Compromisso de Conta	95	13	3º	Distribuição de *malware*	119	8	9	+ 116	+
4º	Exp. de vuln. (intrusão)	58	8	4º	Compromisso de conta não privilegiada	111	8	8	N/A	N/A
5º	Distribuição (*malware*)	55	7	5º	Acesso não autorizado	58	4	4	+ 867	+
6º	Tentativa de *login*	30	4	6º	Compromisso de aplicação	55	4	4	N/A	N/A
7º	*Scan*	28	4	7º	Sistema vulnerável (vulnerabilidade)	41	3	N/A	N/A	N/A
8º	DoS/DDoS	27	4	8º	Utilização Ilegítima de nome de terceiros	32	2	2	+ 68	+
9º	Utilização ilegítima de nome de terceiros	19	3	9º	Indeterminado (outro)	28	2	2	+ 65	+
10º	Exp. de vuln. (tentativa de intrusão)	18	2	10º	Tentativa de *login*	26	2	2	- 13	-

Figura 1.2 - Top 10 de Incidentes registados por tipo em 2019 e 2020 em Portugal. Fonte: Relatório Ministério Público [5].

Exemplos recentes de ataques mediáticos, pelas suas dimensões e consequências, são os que foram realizados contra produtos das empresas *SolarWinds* e *Microsoft* [6], [7], que causaram enormes impactos económicos e sociais nos seus clientes. Entre as vítimas do ataque à *SolarWinds* em 2020, destacam-se agências governamentais americanas (*Homeland Security* e o departamento de justiça), companhias como a *Microsoft, Intel* e *Cisco*.

No início do ano de 2021, ocorreu o ataque à empresa tecnológica *Microsoft*, mais concretamente ao seu serviço de *email Exchange Server* [6], [7]. Neste ataque, mais de 10 grupos categorizados como ameaças avançadas persistentes (*Advanced Persistent Threat – APT*) exploraram as vulnerabilidades do serviço, entre os quais alguns estariam, alegadamente, ligados ao governo chinês. Como consequência, mais de 250 000 vítimas foram afetadas, destas, 23% eram instituições políticas e militares, 14% eram agências financeiras e bancárias [7], [8].

A monetização das informações e vulnerabilidades das vítimas exploradas pelos diferentes tipos de ataques apresentados, resultam em proveitos financeiros ilícitos que promovem de forma persistente este tipo de atividades, e que se traduz numa notável ascensão da economia do cibercrime. Além do mais, nenhum sistema é 100% seguro, como tal, a única forma de evitar ou diminuir a suscetibilidade das organizações a este tipo de ações é investir na cibersegurança. Esta é uma área crucial para qualquer organização que utilize sistemas de informação e que ambicione ter a confiança dos seus clientes.

1.1. Problemáticas de Investigação

A batalha entre os responsáveis pela cibersegurança das organizações e os cibercriminosos, tem conhecido novos capítulos em termos de avanços tecnológicos. Aliado a esta situação, tem-se verificado táticas e técnicas cada vez mais complexas e difíceis de detetar. Consequentemente, esta parece ser uma batalha sem fim à vista [9].

De forma a prevenir as ciberameaças, os especialistas em cibersegurança necessitam de melhorar e adaptar as suas defesas perante a volatilidade das mesmas. Para melhorarem estes aspetos, necessitam de obter informações fidedignas sobre os indicadores de compromisso (IoC) associados às ameaças, bem como as motivações e estratégias utilizadas pelos atacantes. Como tal, a *Cyber Threat Intelligence* (CTI) tem-se revelado uma excelente adição para as organizações, mais especificamente para os centros de operações de segurança, pois permite obter conhecimento técnico aprofundado sobre as ameaças e vulnerabilidades existentes. Além do mais, a correlação de *threat intelligence* interna das próprias organizações, com a *threat intelligence* recolhida de fontes externas, permite a produção de um maior conhecimento sobre as ameaças. Consequentemente, esta deverá ser considerada uma área com um papel vital nas organizações [10], [11].

Infelizmente, diversos fatores relacionados com o excesso de volume de dados, incerteza sobre as fontes dos dados, formatos erróneos, ausência de partilha de informações sobre as ameaças em tempo útil e inconsistência dos dados, tornam complexa a implementação de CTI nas organizações e evidencia a importância de existir sistemas autónomos que suportem a análise de CTI [10]. Recentemente, as organizações têm apostado em novas abordagens, complementares aos mecanismos de defesa das organizações, no sentido de integrarem dados obtidos através de *threat intelligence* e aplicarem esse conhecimento na construção de sistemas de defesa à base de *machine learning*, bem como em sistemas de reputação de ameaças [9], [12]–[15].

A implementação de sistemas de reputação nas organizações poderá ser uma mais-valia na proteção das redes das mesmas. Contudo, existem algumas limitações que têm de ser ultrapassadas, designadamente: a deteção de ameaças do dia zero, taxa de falsos positivos elevada, baixa precisão na deteção de ameaças, ausência de dados que suportem a reputação atribuída à ameaça e sistemas que recorrem a plataformas pouco precisas para recolher dados sobre ameaças [9], [16].

1.2. Objetivos

Nos dias de hoje, estar conectado a uma ou mais ferramentas de *threat intelligence* é a chave para realizar uma monitorização proativa de sistemas de informação (SI) expostos publicamente. O grande desafio prende-se com a capacidade de as organizações conseguirem deduzir, a partir da troca de informações, quem são os atores responsáveis pelas ameaças e os vetores de ataque utilizados para propagar as mesmas. Portanto, é fundamental para qualquer organização ser capaz de filtrar as informações úteis, bem como determinar se existe alguma ação de caráter preventivo que seja de implementação imediata para a organização e, a partir da qual, deva ser lançado um aviso aos *stakeholders* sobre riscos iminentes que coloquem em causa a cadeia de valor. Por fim, e não menos importante, é necessário correlacionar os dados obtidos dos sistemas de informação sobre o comportamento corrente do sistema com as atividades suspeitas recolhidas a partir de plataformas *Open Source Intelligence* (OSINT) sobre cibersegurança e tomar as ações necessárias.

Este trabalho foi desenvolvido em colaboração com a organização OutSystems, desta forma tem como objetivo geral o desenho e implementação de uma ferramenta para aquisição de dados sobre potenciais ameaças acionável para ajudar a proteger as organizações. Por outras palavras, é desenvolver uma ferramenta que possa ser útil e eficiente na recolha de dados referentes a potenciais ameaças à cibersegurança da organização, e providenciar formas de avaliar os riscos associados a essas potenciais ameaças.

1.2.1. Objetivos Específicos

Os objetivos específicos foram estabelecidos de acordo com as necessidades da organização, como tal foram definidos os seguintes:

1. Implementar e providenciar uma plataforma, que permita centralizar informações sobre potenciais ameaças à cibersegurança, incluindo indicadores de compromisso específicos, com a finalidade de ser utilizada na aquisição, produção e partilha de informação sobre ameaças;

2. Apresentar os dados sobre potenciais ameaças à cibersegurança normalizados, de forma a cumprir o formato utilizado para partilha de dados através dos protocolos STIX/TAXII;

3. Implementar um modelo de avaliação da ameaça com base nos dados da organização e nos dados recolhidos através de *Open Source Intelligence* (OSINT) sobre cibersegurança;

4. Permitir que o modelo de avaliação de risco seja dinâmico, de forma a fornecer resultados com base na evolução da ameaça.

1.3. Organização

A OutSytems é uma empresa nacional, fundada no ano de 2001, líder de mercado em Portugal e com forte presença internacional nas plataformas *low code* para a transformação digital. É detentora da plataforma de desenvolvimento rápido (*Rapid Application Delivery*) que permite acelerar o processo de criação de aplicações móveis e *web* [17].

O crescimento da empresa evidenciou a necessidade de a empresa se reforçar com um departamento próprio de segurança de redes e, para tal foi criada uma equipa de segurança em 2016. Atualmente, este departamento está dividido em 4 subgrupos: *risk and compliance*; *security automation*; *application security; security operations and incident response*. Este último departamento é responsável pela segurança de todos os *stakeholders*, desde os clientes aos elementos corporativos.

O âmbito deste trabalho está enquadrado no subgrupo de *security operations and incident response*, no entanto todo o trabalho estará estreitamente relacionado com os restantes grupos da segurança de redes.

1.4. Desenho de Investigação Científica

A metodologia do Desenho de Investigação Científica (*Design Science Research* – DSR) na área dos sistemas de informação permite produzir novo conhecimento através da construção e avaliação de diferentes artefactos, tais como *software*, processos, modelos, entre outros [18], [19].

Na presente dissertação, serão seguidos os princípios do DSR, em que a partir dos problemas identificados e com o conhecimento científico existente, será apresentado um artefacto que consiste num modelo de reputação de indicadores de compromisso. Por conseguinte, foi utilizado o *feedback* dos analistas da OutSystems como processo iterativo (metodologia ágil), para desenvolver o desenho tecnológico alinhado com os objetivos organizacionais. Esta interação constante, permitiu desenvolver um trabalho mais completo e que permite atender a requisitos de muitas das organizações ao nível da cibersegurança.

O modelo desenvolvido, permite ser implementado e testado de forma rigorosa para resolver problemas reais da indústria. Portanto, a validação do mesmo com recurso a uma matriz de confusão, permitirá extrair métricas para avaliar o desempenho e permitirá corroborar que o desenho prático deste trabalho é adequado para o desenvolvimento de modelos de reputação. Adicionalmente, após a implementação do conector responsável pelo modelo de reputação na plataforma de *threat intelligence* escolhida para o efeito, será realizado um estudo empírico com recurso aos dados analisados pelo conector e que serão importantes para comprovar a sua utilidade, demonstrar as suas características e benefícios na avaliação de ameaças [18].

A Figura 1.3 exemplifica as interações entre o DSR, a base de conhecimento e o ambiente. O valor do DSR neste projeto, reside na forma como o modelo foi desenvolvido, ou seja, nas atividades de construção e avaliação realizadas de forma iterativa, que permitiram acrescentar valor ao trabalho.

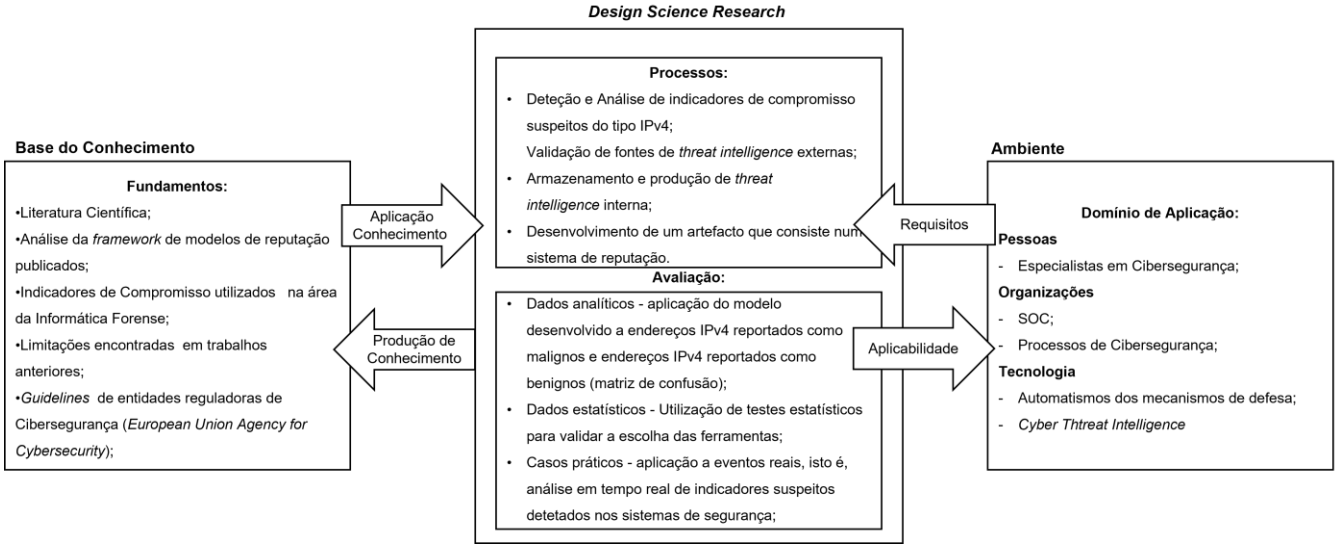

Figura 1.3 – Ciclo do Design Science Research aplicado ao trabalho com base nas recomendações de Hevner e colegas (2004) [19]. *Fonte: elaboração do autor.*

Para validar o contributo da presente dissertação para a área de conhecimento em cibersegurança e a forma como poderá ser utilizada para mitigar limitações existentes, foram utilizadas as sete linhas orientadoras publicadas por Hevner e colegas em 2004 [19]. Portanto, o Quadro 1.1 foi construído com base nessa publicação científica, e tem como objetivo validar o DSR aplicado.

Quadro 1.1 Validação do contributo do Design Science Research para a área de conhecimento em cibersegurança, de acordo com as linhas orientadoras de Hevner e colegas (2004) [19].*Fonte: elaboração do autor*

Guideline	**Descrição**	**Aplicação do DSR ao trabalho**
Guideline 1: Desenho de um artefacto	*DSR* deve produzir um artefacto viável na forma de um modelo ou instanciação.	O projeto apresentado resultará num artefacto viável, neste caso num modelo de reputação de IoC.
Guideline 2: Relevância do Problema	O objetivo do DSR é desenvolver uma solução tecnológica importante e relevante para o problema de negócio.	O modelo apresentado foi formulado para atender às atuais limitações nos SOC de muitas organizações, tendo como organização parceira a OutSystems.
Guideline 3: Avaliação do Desenho	A utilidade, qualidade e eficácia do artefacto deve ser rigorosamente demonstrada através de métodos bem executados.	O modelo é avaliado quanto à sua viabilidade na integração com outras ferramentas, quanto à sua utilidade e quanto ao seu rigor através de métricas extraídas de testes (por exemplo Matriz de Confusão) .
Guideline 4: Contribuições da Investigação	O DSR deverá providenciar contribuições claras e verificáveis na área do desenvolvimento do artefacto e /ou na metodologia desenvolvida.	A construção do modelo contribui para o desenvolvimento de futuros modelos, através das técnicas utilizadas. Além do mais, a *framework* desenvolvida para a construção do conector que integra a plataforma de *threat intelligence,* é uma mais-valia para organizações que pretendam implementar sistemas idênticos.
Guideline 5: Rigor Científico	DSR depende da aplicação de métodos rigorosos, tanto na construção como na avaliação do artefacto.	Foram aplicadas metodologias científicas e tecnológicas para aumentar o rigor do modelo. Adicionalmente, foram seguidas *guidelines* da literatura científica e agências de cibersegurança para avaliar o modelo.
Guideline 6: Desenho como um Processo de Procura	A procura por um artefacto eficaz, requer a utilização de meios disponíveis para alcançar os fins desejados. Para tal, nunca deverá infringir a legislação aplicável existente.	Foram utilizadas fontes científicas para orientar o modelo desenvolvido, bem como o *feedback* de analistas do SOC da OutSytems. O presente trabalho cumpre todas as normas em vigor e foi supervisionado pela organização parceira.

Guideline 7: Comunicação da Investigação	DSR deverá ser apresentado de forma eficaz, quanto ao caráter tecnológico, como para o público-alvo.	A investigação desenvolvida será publicada sob a forma de dissertação. O presente trabalho foi apresentado num evento de cibersegurança (*INNCYBER Innovation Hub – 3rd Edition*), e foi alcançado o 3º lugar no evento em questão.

1.4.1. Identificação de Questões Teórico-Científicas

O presente trabalho contribuirá com conhecimento científico no âmbito da cibersegurança, como fora descrito anteriormente. Desta forma, a questão fundamental que orientará o desenvolvimento teórico-prático deste trabalho, tendo em conta o objetivo principal, é: Como modelar a aquisição de dados de *threat intelligence* para integrar de forma autónoma com outras ferramentas de segurança para melhorar a proteção das organizações?

Segundo o artigo científico publicado por Knauss[35], as dissertações de mestrado são importantes para desenvolver uma forte componente sobre métodos de investigação empíricos. Como tal, para desenvolver um trabalho de investigação rigoroso, mesmo que aplicado à indústria, deverão ser definidas questões de investigação direcionadas ao problema, à solução e à avaliação da solução [35]. Portanto, as questões definidas para a metodologia de investigação foram as seguintes:

- **RQ1 (problema)**: Quais os problemas associados à gestão e integração de *threat intelligence* nas organizações?
- **RQ2.1 (solução)**: Quais as potenciais soluções para a gestão e integração de *threat intelligence* nas organizações?
- **RQ2.2 (solução):** De que forma estas soluções podem ser implementadas?
- **RQ2.3 (solução):** De que forma a(s) solução desenvolvida(s) ajudam a proteger a organização?
- **RQ3.1 (avaliação):** Até que ponto os problemas típicos associados à gestão e integração de *threat intelligence* nas organizações podem ser resolvidos?

1.5. Estrutura da Dissertação

Com vista a apresentar soluções para os problemas afetos à cibersegurança das organizações e alcançar os objetivos apresentados, a presente dissertação está organizada em 5 capítulos. O presente capítulo corresponde à introdução, é realizado um enquadramento do tema, são descritas as problemáticas associadas, são definidos os objetivos com vista à solução dos problemas e é apresentada a organização parceira para aplicar o modelo de reputação desenvolvido à indústria. Por fim, é descrito o desenho de investigação aplicado ao trabalho.

No capítulo 2 é realizada uma revisão da literatura, de forma a abordar os conceitos mais relevantes nesta área, bem como são descritas com elevado grau de especificidade as problemáticas associadas a esses conceitos. De uma forma geral, são descritas as principais ameaças à cibersegurança, os indicadores que permitem reconhecer que os sistemas de informação foram afetados por uma ou várias destas ameaças. Por conseguinte, é descrito o conceito de centro de operações de segurança, responsável por detetar e mitigar as ameaças à cibersegurança, bem como a tríade que compõe estes centros de operações de segurança, designadamente: as ferramentas existentes para integrar nestes centros, os recursos humanos que os compõem e a área de *cyber threat intelligence*.

No capítulo 3 é apresentada a arquitetura de integração proposta e metodologia de implementação, onde é apresentada uma visão geral das componentes que compõem a solução apresentada e a forma como estas se relacionam. Seguidamente, é descrito a metodologia de trabalho para implementar a solução, sendo que numa primeira instância são identificadas as questões teórico-científicas com vista a manter o rigor científico do trabalho desenvolvido e, numa segunda instância, são apresentadas as técnicas utilizadas para desenvolver o trabalho prático.

No capítulo 4, são apresentados e discutidos os resultados obtidos com o objetivo de analisar as vantagens e desvantagens do modelo apresentado. Além do mais, é neste capítulo que são descritos os parâmetros de validação do modelo e onde ocorre a validação do mesmo.

Por fim, no capítulo 5, é apresentada a conclusão da dissertação, através da súmula do trabalho desenvolvido e da contribuição da ferramenta para auxiliar na triagem de ciberameaças, e sugestões para trabalhos futuros com o intuito de melhorarem o modelo.

CAPÍTULO 2

Revisão da Literatura

2.1. Contexto

A partir de 1980 começou a existir uma preocupação crescente com a segurança da informação, uma vez que é nesta década que, após a Era Industrial, se dá início à Era Tecnológica. Esta Era é marcada por uma panóplia de inovações que permitem levar à globalização, entre as quais destaca-se o armazenamento digital. No início do desenvolvimento desta capacidade tecnológica, para aceder aos dados era necessário estar no mesmo local físico dos equipamentos responsáveis por armazená-los, ou seja, a ameaça era bastante reduzida pois era necessário aceder ao local com controlo de acesso físico [20]. Na década de 90 os equipamentos tecnológicos evoluíram de forma a terem capacidades de armazenamento digital superiores, passaram da ordem de grandeza dos *bytes* para os *terabytes*, e fisicamente passaram a ter tamanhos mais reduzidos, bem como, níveis de performances incomparáveis aos antecessores. Além destas mudanças, o paradigma também mudou, uma vez que os dados pessoais deixam de ser irrelevantes e passam a ter valor, por exemplo o número do cartão de crédito, números de contas bancárias, códigos de segurança, entre outros. Além do mais, os dados passam a estar acessíveis a partir de qualquer parte do mundo, levando a uma crescente preocupação em mantê-los seguros.

Em 2025 estima-se que existirão cerca de duas centenas de *zettabytes* de dados em todo o mundo, sendo que 1 *zettabyte* corresponde a 1 $x10^{21}$ *bytes*. Estes dados correspondem, por exemplo, a aplicações, vídeos, documentos, entre outros. O risco de acesso ilegítimo a estes dados também aumentou proporcionalmente, conduzindo à necessidade de desenvolver e aprimorar estratégias de segurança de redes [21].

2.2. Ameaças

A evolução da tecnologia foi acompanhada pela sofisticação das ameaças aos sistemas de informação, ou seja, estas ameaças tornaram-se mais versáteis, discretas e utilizam cada vez mais pontos de acesso para perpetuar os ataques e penetrar nos ambientes das Tecnologias de Informação (TI) [22]. De acordo com o Fórum Económico Mundial, os ciberataques e vulnerabilidades estão classificados entre os quatro riscos mais elevados que ameaçam a estabilidade global [23]. Esta situação é consequência da dependência que a maioria dos países têm em relação ao ciberespaço para comunicações e, até

mesmo, para controlo do mundo físico [2]. Portanto, a segurança de cada país é cada vez mais dependente da segurança do ciberespaço.

A maioria dos especialistas em cibersegurança afirmam que o *malware,* também conhecido como *software* malicioso[24], é a principal ferramenta para realizar atividades maliciosas no ciberespaço. *Malware* é definido como um software que permite a realização de um conjunto de ataques sem o conhecimento legítimo do proprietário, que visam comprometer um sistema em benefício do atacante [1], [3], [25]. Todavia, as ameaças não são apenas derivadas de *malwares*, são ações perpetuadas por uma grande variedade de atores capazes de causar danos às organizações governamentais e não governamentais, entre os quais destacamos os seguintes:

- Pessoas pertencentes às próprias organizações alvo, que devido a estarem insatisfeitas, e de uma forma consciente, pretendem causar danos ou roubar dados importantes. Além destas, existem também os indivíduos que de forma inconsciente devido a diversos motivos, como falta de sensibilização ou despreocupadas com a segurança, são vetores para terceiros mal-intencionados. Estes dois tipos de entidades acabam por ser as mais perigosas, uma vez que já se encontram dentro dos sistemas de defesa e têm acesso direto aos sistemas de informação [2], [26];
- *Hackers* - Os *hackers* podem ser categorizados quanto à sua motivação, isto é, existem os que têm motivações pessoais, e querem dar-se a conhecer ou expressarem-se a si próprios, e os que têm motivações políticas ou religiosas, e atacam sobretudo páginas de *internet* populares de um país/religião ou servidores de *e-mails* de uma dada organização. Estes últimos são chamados de *hacktivists* [2];
- Organizações que têm a finalidade de roubar dados ou chantagear os alvos com pedidos de resgate. Como já foi referido, os dados digitais têm um enorme valor e são considerados o "novo ouro", consequentemente são facilmente negociados na *dark web*[1] ou grupos de redes clandestinas *peer-to-peer* (P2P) (*darknets*) [27];
- Ameaça Persistente Avançada – são ameaças muitas vezes patrocinadas pelos estados, têm grandes recursos financeiros e técnicos que são utilizados para fins de espionagem militar ou comercial, permitindo assim que perdurem no tempo até se infiltrarem e assumirem o controlo dos SI alvo [2], [23].

[1] A *internet* está dividida em três subcomponentes: *Surface Web* onde estão disponíveis a maioria das páginas que os utilizadores estão habituados a navegar; *Deep Web* que corresponde a informações disponíveis em páginas escondidas e protegidas de forma a inibir o acesso fácil e não aprovado; *Dark Web* composta por páginas que requerem conhecimento e *software* especializado, por exemplo o *browser* Tor, para obter acesso [27].

Estes atores mal-intencionados recorrem a diferentes técnicas para espalharem *malware,* entre as quais se destacam *spam, phishing* e *downloads* da internet [1]. Após conseguirem obter acesso aos sistemas das vítimas, existem diferentes variantes de *malwares* utilizados para perpetrarem os ataques:

- Bomba lógica – é um código introduzido no *software* alvo, que é executado a partir do momento em que são cumpridas determinadas condições de forma a causar danos no sistema [28];
- *Botnet* – rede controlada remotamente, composta por máquinas infetadas, utilizada para distribuir *malware,* coordenar ataques e enviar *spam* para causar novas infeções [29], [30]. Estas máquinas controladas remotamente podem ser controladas para realizar um ataque *Distributed Denial of Service* (DDOS);
- Cavalos de troia – código malicioso escondido num *software* que aparenta ser fidedigno e útil para o utilizador, no entanto pode ser iniciado em determinado momento remotamente ou após serem reunidas algumas condições de forma a prejudicar o sistema alvo [28];
- Ferramentas abusivas – estão ao dispor de qualquer indivíduo na internet, isto é, não é necessário ter conhecimentos na área, qualquer indivíduo que tenha curiosidade em explorar vulnerabilidades dos sistemas e nas redes, pode fazê-lo com recurso a estas;
- Negação do serviço (*denial of service*) – nesta situação é impedido o acesso do utilizador ao sistema e vice-versa. Esta interrupção pode ser realizada através do envio de várias mensagens a partir de uma única fonte para o sistema da vítima, impedindo o fluxo normal de dados, de tal forma que poderá levar ao bloqueio do mesmo. Existe uma variante deste ataque, em que este é perpetuado a partir de múltiplas fontes, e não apenas de uma, ou seja, é um ataque distribuído e simultâneo (DDOS) [2], [28], [31];
- *Ransomware* – causa diferentes formas de ataques sofisticados com múltiplas mutações. Existem dois grandes grupos compostos por diferentes variantes, *Crypto* e *Locker,* que são responsáveis por utilizar um algoritmo de encriptação para codificar os ficheiros e utilizar técnicas de escalonamento de privilégios de várias aplicações de gestão para impedir o acesso do utilizador aos recursos, respetivamente. Os ataques têm como consequência a encriptação de ficheiros do sistema e bloqueio do mesmo. Este *malware* tem comportamentos polimórficos e metamórficos com ofuscação do próprio código, o que dificulta a sua deteção por parte da cibersegurança [24]. Atualmente, é dos ataques mais utilizados e que mais prejuízos causa em diferentes tipos de indústrias.

- *Sniffer* - *software* utilizado para procurar informações específicas nos pacotes de dados que passam pelo *router*;
- *Worm* – programa autónomo que ao ser iniciado infeta outros computadores na rede [28], [32];
- Vírus – contamina os arquivos do sistema ao introduzir uma cópia nesses ficheiros, uma vez carregados na memória infetam outros ficheiros aí presentes. Ao contrário dos *worms*, os vírus requerem a intervenção humana [28], [32].

Além dos diferentes ataques mencionados anteriormente, existe uma grande panóplia de outros ataques, cada um deles com taxonomia própria para classificar as suas ameaças [23]. A relação entre as diferentes formas de ataques e os atores está representado na Figura 2.1.

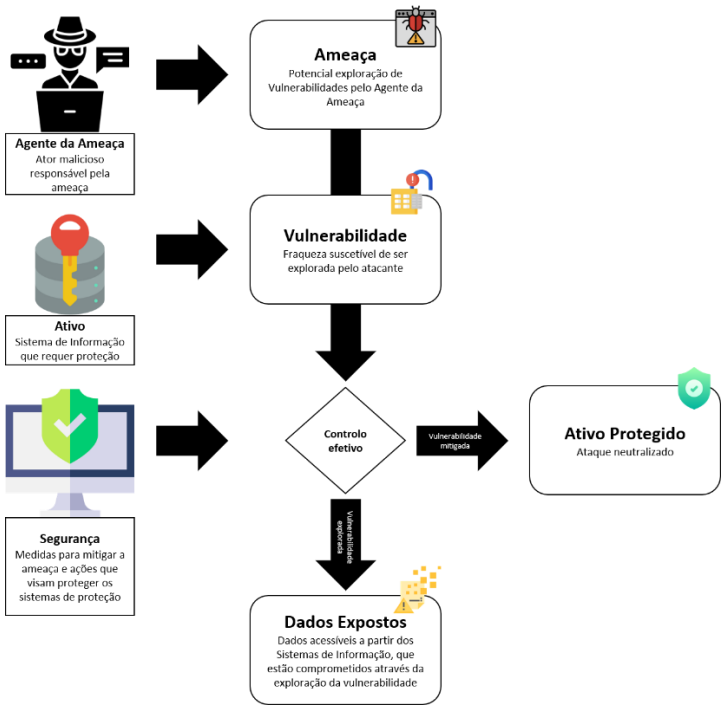

Figura 2.1 - Relação entre atacantes e ameaça. Fonte: elaboração do autor.

2.3. Indicadores de Compromisso

Indicadores de Compromisso (*Indicators of Compromisse* – IoC) são descritos como artefactos forenses que são indiciadores de um SI estar comprometido perante uma das ameaças descritas anteriormente [33]. IoC são caracterizados por ter diferentes tamanhos, formatos e são categorizados em dois grandes grupos, consoante o local de onde são extraídos: IoC presentes nas máquinas hospedeiras e IoC presentes na rede. Todavia, existe alguns que podem existir em ambos os grupos, como por exemplo *strings* e nomes de ficheiros [33], [34], [35].

Podemos destacar os seguintes IoC de ambos os grupos [35]:

Quadro 2.1 - Diferentes tipos de Indicadores de Compromisso e as suas origens. Fonte: Adaptado de [35].

IoC baseados em Máquinas Hospedeiras	**IoC baseados na rede**
Chaves de registo	Endereços IPv4
Nome de ficheiros	Endereços IPv6
Strings	*Strings*
Nome de processos executados na máquina	Nomes de domínios de páginas de internet
Valores *hash* de ficheiros. Por exemplo, *Message-Digest Algorithm 5* (MD5) *hashes* – identifica de forma inequívoca *malwares*, cada *malware* tem o seu próprio código *hash*	Certificados *hash* X509 – assinaturas e chaves públicas
Contas de utilizadores comprometidas	Protocolos de comunicação
Caminhos de diretorias	*Uniform Resource Locators* (URL's) – endereços da localização de ficheiros ou páginas na Internet;
Assinatura do vírus (*Virus Signatures)* – é uma porção única de dados ou *bits* de código que permite identificar um vírus	Nome de ficheiros

Estes indicadores são analisados nos sistemas operativos ou nas redes de internet e intranet, de forma a detetar ameaças e prevenir ataques futuros. Portanto, é de extrema importância as ferramentas de cibersegurança terem acesso a bases de dados atualizadas sobre estes indicadores, pois permitirá aumentar a eficácia da resposta a incidentes de segurança.

2.4. Security Operations Center (SOC)

O nível de sofisticação das ameaças, a constante procura, por parte dos atacantes, de vulnerabilidades e equipas de cibersegurança mal preparadas, implica que as organizações sejam forçadas a implementar e a desenvolver departamentos próprios para lidar com estas questões. É neste sentido que surge o conceito *Security Operations Center* (SOC), numa primeira instância utilizado pela empresa Cisco em 2005 [6], [22], [36].

Estes centros são uma unidade interna ou externa às organizações, que permitem ter uma visão holística da segurança informática destas. Têm como principal missão monitorizar, controlar, evitar e defender os sistemas de informação contra ameaças à cibersegurança [37]. Para tal, é necessário que os referidos centros estejam implementados de forma correta, que tenham recursos informáticos atualizados e avançados [38], capacitando-os de mitigar o elevado número de ataques a que as instituições são sujeitas. A única forma de garantir estas premissas é através de auditorias regulares a estes centros.

Como foi referido anteriormente, este termo foi mencionado pela primeira vez há quase duas décadas, porém só nos últimos anos é que começou a suscitar um interesse crescente e, por conseguinte, resultou em objeto de análise de diversos estudos científicos na área de Ciências da Computação. De acordo com o estudo realizado pela equipa de Vielberth em 2020 [38], a evolução dos artigos científicos relacionados com o tema SOC teve um aumento significativo desde 2015 e é expectável que se acentue nos próximos anos (Figura 2.2).

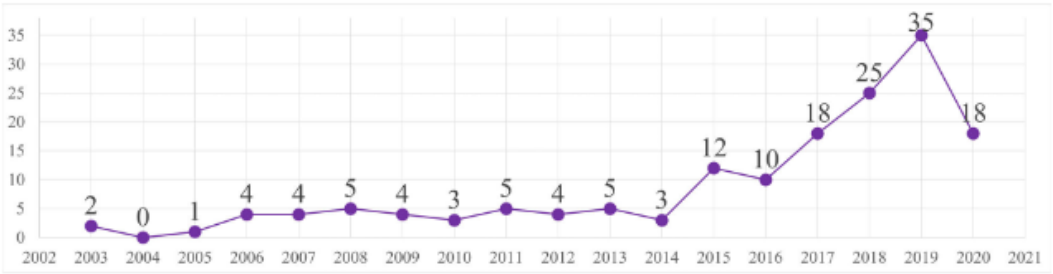

Figura 2.2 - Publicações relevantes sobre SOC até ao final do 1º semestre de 2020. Fonte: Retirado de [38].

Apesar deste tema ser cada vez mais debatido e objeto de estudo em diversas publicações, ainda não existe consenso quanto ao modelo mais aconselhável para a estrutura do SOC, isto é, quanto à metodologia, ferramentas e recursos que deverão ser implementados nestes centros. Esta lacuna reflete a forma como algumas organizações e investigadores ainda olham para estes centros, ou seja, muitos destes ainda fazem uma caracterização errónea ao considerá-los uma mera entidade responsável por monitorizar a rede.

Adicionalmente, existem outros fatores que são altamente negligenciados no desenvolvimento dos SOC e acabam por ter um impacto negativo, sendo eles [37]:

a) Soluções não escaláveis - incapacidade de monitorizar em larga escala os diferentes sistemas de informação e utilizadores conectados à rede da organização;
b) Dados não fiáveis – registos recolhidos dos sistemas de informação pouco fiáveis, aumentando a suscetibilidade às ameaças;

c) Recursos limitados – *software* com recursos mínimos de análise de eventos, com triagem de eventos básicos e sem capacidade de integrar *inputs* de outros *softwares*;
d) Atitude reativa – os recursos humanos disponíveis estão preocupados com tarefas rotineiras que os impedem de analisar todos os eventos em tempo real, resultando numa atitude reativa às ameaças ao invés de uma postura proativa na prevenção destas;
e) Ausência de automatismos – integração manual dos dados provenientes de diferentes fontes, em vez de existir *softwares* compatíveis para trocarem um grande volume de dados de forma automática.

Nos dias que correm, os erros apresentados, as divergências e estratégias mal definidas começam a ser exceções. A maioria das organizações e publicações científicas são unânimes quanto à utilidade do SOC, considerando-o como uma entidade organizacional de extrema relevância para evitar possíveis ataques e na produção de *threat intelligence* [38]. Como será abordado nos tópicos seguintes, existe cada vez mais tecnologia disponível para integrar nestes centros de forma a torná-los mais capacitados. No entanto, para aumentar a eficácia nas suas tarefas é crucial a partilha de informação entre os centros homólogos de cada organização, o que por vezes gera bastante relutância entre as instituições envolvidas, devido ao, suposto, receio de estarem a partilhar segredos de negócio e políticas da empresa [39].

2.4.1. Recursos Humanos

A natureza imprevisível do comportamento humano faz com que os recursos humanos sejam um elemento indispensável na análise das ameaças à cibersegurança [40]. Desta forma, as funções centrais do SOC devem ser desempenhadas por analistas, que através das ferramentas certas podem detetar e mitigar incidentes de segurança, impedindo o prejuízo de milhões de euros às organizações. A importância crescente dos recursos humanos na cibersegurança, tem resultado na necessidade de contratar mais profissionais desta área e, consequentemente, a falta de analistas SOC competentes e experientes tem dificultado a contratação destes especialistas [22], [41].

A melhor forma dos analistas estarem alertas para detetar ameaças, é serem treinados com frequência, como tal existem estudos que comprovam cientificamente a utilidade de plataformas tecnológicas que simulam ataques reais e permitem desenvolver competências críticas nestes indivíduos [22]. Além do mais, estes treinos permitem que sejam adquiridas outro tipo de competências, como por exemplo um maior domínio das ferramentas ao seu dispor, uma vez que estas simulações são feitas com recurso às suas ferramentas de trabalho.

Outra limitação, é o número excessivo de alertas e eventos gerados diariamente pelos sistemas de segurança, chegam a atingir as centenas de milhares, um número bastante superior ao que aos analistas conseguem investigar. Consequentemente, na maioria das vezes, só são investigados os eventos com classificação severa, o que pode resultar na negligência de um evento que possa resultar num incidente ou num ataque real [42].

Outro erro crasso que deve ser evitado pelas organizações, é olharem para os recursos humanos num sentido restrito no contexto de cibersegurança, isto é, consideram apenas importante a formação dos indivíduos que estão ligados à área da segurança de redes dentro da organização, e negligenciam os demais, cuja formação e sensibilização é crucial, pois estes são vetores para uma grande percentagem dos ataques que ocorrem às organizações [2].

2.4.2. Ferramentas Disponíveis

Ao longo destes últimos anos, tem ocorrido uma grande evolução e desenvolvimento de diferentes tecnologias no domínio da segurança de redes, como está descrito no estudo científico de Miloslavskaya publicado em 2021 [36]. Estas técnicas desenvolvidas podem dividir-se em dois grandes grupos: as ferramentas de proteção das redes (*Network Protection Tools* - NPT) e os métodos de proteção das redes (*Network Protection Methods* – NPM). Neste subcapítulo serão abordadas algumas das ferramentas mais relevantes de ambos os domínios.

2.4.2.1. *Security Incident Event Management (SIEM)*

O termo *Security Information and Event Management* (SIEM) foi descrito inicialmente pela empresa de investigação e consultoria "Gardner" em 2005, tendo esta ferramenta surgido para substituir dois sistemas distintos que existiam até à data, *Security Information Management* (SIM) e *Security Event Management* (SEM). O primeiro recolhia registos, arquivava e gerava reportes ao longo do tempo, ou seja, era orientado para um registo histórico. Por sua vez, o segundo fornecia relatórios em tempo real, recolhia registos e correlacionava-os, ou seja, era um sistema orientado ao imediato [43].

Desde então, os sistemas SIEM tornaram-se numa ferramenta indispensável no seio das organizações, mais especificamente nos SOC, sendo considerada a ferramenta fulcral destes centros [22] . Este mecanismo tem evoluído ao longo do tempo para versões mais complexas e completas, atualmente está na versão 2.1, sendo que tem os seguintes propósitos:

- Recolher e correlacionar dados de segurança, usualmente chamados de *logs* ou registos, a partir de dispositivos, *firewalls*, mecanismos *Intrusion Detection/Prevention System* (IDS/IPS), *Data Loss Protection* (DLS), *Domain Name System* (DNS), *Dynamic Host Configuration Protocol* (DHCP) e eventos de segurança de sistemas operativos *Windows/Linux* [42];

- Monitorizar atividade de utilizadores e aplicações [22];
- Gerar reportes automáticos de incidentes de segurança, bem como recomendações para os mesmos [43];
- Rastrear o ciclo de vida de um incidente de segurança, desde a sua origem até resultar num alerta e acionar mecanismos de resposta para resolução [43];

Durante a monitorização constante dos ficheiros de registo dos SI (*logs*), são recolhidos *terabytes* de dados e com diferentes formatos, o que enaltece a importância de automatizar ao máximo estas tarefas e a centralização destes registos nesta ferramenta. A partir destes *logs* de segurança, as equipas que trabalham nos SOC, definem regras e condições (*use cases*) com base na sua experiência e nas informações provenientes da *threat intelligence* [22], [23], [42]. Como tal, estas regras são flexíveis ao ponto de se adaptarem a novas circunstâncias e a ameaças dinâmicas [43]. Estas condições correlacionam diferentes indicadores dos registos, e caso surjam alertas de que os SI estão comprometidos, a ferramenta SIEM despoletará um evento suspeito que terá de ser analisado pelas equipas de forma a verificarem o risco, classificando-o como um verdadeiro positivo ou um falso positivo. Caso seja considerado um verdadeiro positivo, resultará na abertura de *ticket* de incidente, chamado de *Open Source Ticket Request System* (OTRS), que carece de uma análise mais minuciosa e poderá passar para o nível 2, que correspondente a um incidente de segurança severo, e para o qual é necessária uma investigação mais exaustiva pela equipa de resposta a incidentes.

Resumindo, a plataforma SIEM combina a informação sobre segurança proveniente de diferentes fontes, correlacionando-a com os registos recolhidos e armazenados num sistema centralizado, como é possível observar na Figura 2.3 [23].

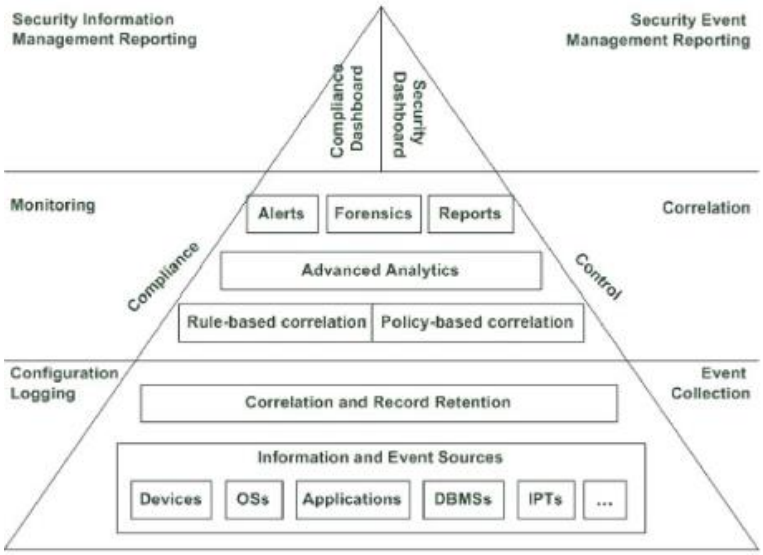

Figura 2.3 - Arquitetura de um sistema SIEM. Fonte: retirado de [43].

2.4.2.2. *Security Orchestration Automation and Response (SOAR)*

As plataformas *Security Orchestration, Automation and Response* (SOAR) correspondem à geração seguinte dos sistemas SIEM, uma vez que têm a capacidade de lançar alertas mediante atividades suspeitas, mas também de agir contra as mesmas [44]. Por conseguinte, atualmente, estas plataformas estão no centro das atenções pela sua projeção para interligar os recursos humanos que trabalham no SOC com os processos implementados e a tecnologia disponível. Por outras palavras, as principais funções de um sistema SOAR são: integração, orquestração e automação [45].

A integração baseia-se na capacidade de as plataformas SOAR suportarem a ligação a uma grande panóplia de outras ferramentas, sejam elas de código aberto, comerciais ou de propriedade intelectual da própria organização, e a partir destas gerar dados uniformizados para a interoperabilidade entre os recursos disponíveis.

A orquestração permite às organizações implementarem e operacionalizarem os Processos de Resposta a Incidentes (PRI), estes baseiam-se em fragmentos de código ou *scripts* conhecidos como *playbooks*. Estes *playbooks* resultam em atividades desenvolvidas de forma ordeira, em que o resultado da atividade anterior é o *input* da próxima atividade, e podem ser realizadas de forma automática ou com recurso à intervenção dos especialistas SOC. Porém, o objetivo é minimizar a intervenção humana e os processos repetitivos desenvolvidos por estes.

A automatização ou resposta depende das atividades que cada organização pretende automatizar através da orquestração de processos, não existe um consenso sobre quais deverão ser. As mais comuns, e utilizadas, são tarefas de validação, verificação de autorizações de acesso e redução de falsos alarmes (Figura 2.4).

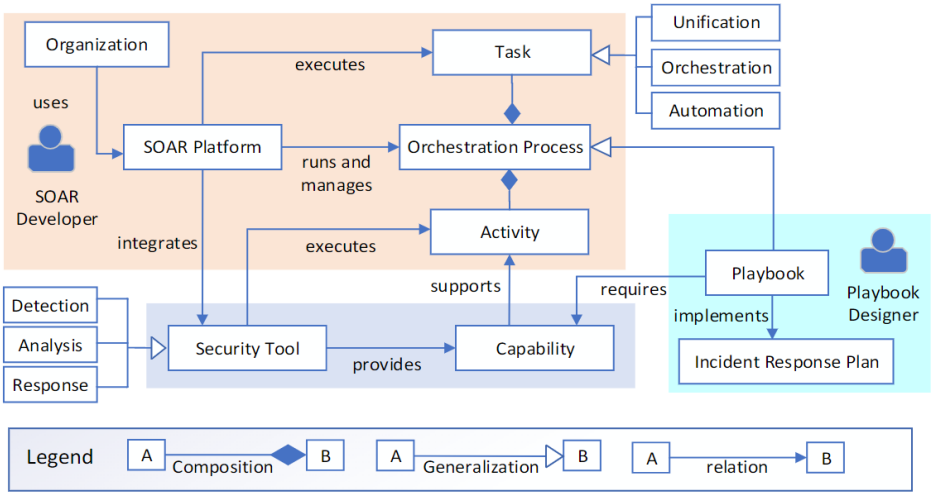

Figura 2.4 - Mapa de atividades do SOAR. Fonte: retirado de [45].

2.4.3. *Cyber Threat Intelligence (CTI)*

Os atacantes trocam constantemente informações sobre vulnerabilidades das organizações, táticas bem-sucedidas e códigos de *software* que permitem ganhar vantagem a partir de uma falha nos SI (*exploit*). Um exemplo flagrante de partilha de informação foi o ataque ocorrido em 2021, que já fora mencionado anteriormente, quando foram detetadas vulnerabilidades nos servidores da ferramenta *Microsoft Excehange,* e por sua vez foram exploradas por diversos grupos de atacantes. Paralelamente, como consequência desta e outras situações, surgiu a necessidade da partilha de informação sobre ameaças, conhecida como *threat intelligence* [6], [46]. Esta metodologia é constituída por ferramentas muito poderosas, usadas não só na área de cibersegurança, mas também noutras áreas, como por exemplo na área militar para contraterrorismo [47].

A CTI pode ser definida como aquisição e gestão de conhecimento contextual sobre ameaças através da análise de dados e informações sobre potenciais riscos à segurança de uma organização. Por outras palavras, este conhecimento é a ligação entre o conhecimento das equipas de segurança sobre as ameaças que afetam as organizações e as previsões de ataques [23], [48]. As principais fontes de dados para CTI são as seguintes: *Open Source Intelligence* (OSINT), *Social Media Intelligence* (SCOMINT), *Human Intelligence* (HUMINT) ou *dark web* [48]–[51].

Não existe um consenso quanto à categorização dos diferentes tipos de conhecimento sobre as ameaças. No entanto, alguns estudos mais recentes, dividem em 4 domínios relacionados com a forma a quem este é direcionado e quem o consome [52], sendo estes os seguintes (Figura 2.5):

- Estratégico – Informação de alto nível sobre mudanças nos riscos e ameaças que perduram no tempo, e que influenciam os gestores nas suas decisões estratégicas. O propósito é ajudar os decisores a perceberem as ameaças atuais e as que poderão enfrentar consoante a estratégia que adotarem, em termos de impacto financeiro causado pelos ataques e suscetibilidade às ameaças [52];
- Operacional – Informação fidedigna sobre ataques iminentes ou que já tenham ocorrido, em que os gestores dos departamentos responsáveis por mitigar as ameaças têm de tomar decisões no imediato ou fazer relatórios sobre a reconstrução do ataque [48]. Este tipo de informação é referente a um espaço temporal curto que abrange as fases pré, durante e pós ataque. Este tipo de inteligência é difícil, uma vez que não é possível as organizações privadas acederem a infraestruturas de grupos de atacantes, por exemplo para intercetar comunicações e obter um bom conhecimento operacional [48], [52];
- Tático – Detalhes sobre o *modus operandi* dos atacantes, ou seja, é informação que é recolhida ao longo do tempo sobre as táticas, técnicas e procedimentos (TTP's) utilizados pelos atacantes, e é relevante para alertar os administradores dos SI, bem como dos SOC, para estarem preparados para as táticas correntes;

- Técnico – Informação consumida pelos recursos humanos peritos em segurança, sobre indicadores técnicos das ameaças, por exemplo IoC específicos que sejam relevantes nas tarefas dos analistas dos SOC ou dos administradores dos equipamentos informáticos responsáveis pela configuração de *firewalls*, apesar destes poderem ser pervertidos pelos atacantes. Este tipo de conhecimento é o que as organizações consideram prioritário, uma vez é utilizado para acionar os mecanismos de defesa de forma automática e é facilmente quantificável em comparação com os restantes tipos de conhecimento. Por exemplo, se é detetado um endereço IP que já está reportado como malicioso, vai ocorrer uma ativação automática das ferramentas de defesa para bloquear esse endereço. Este tipo de *threat intelligence* é a mais partilhada devido à sua fácil padronização [52].

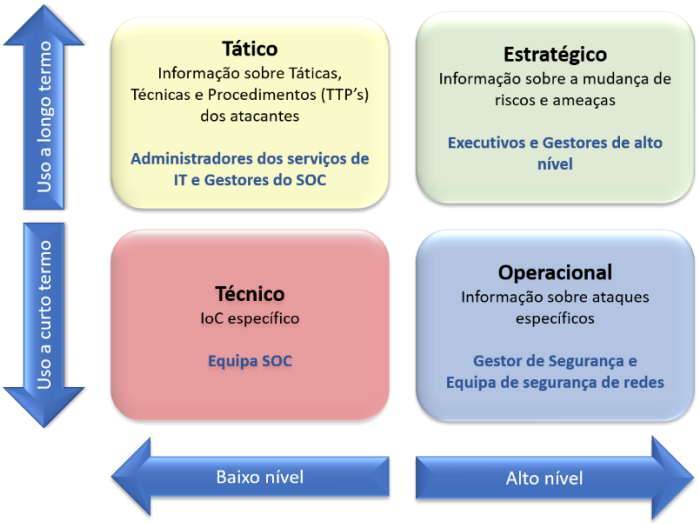

Figura 2.5 - Tipos de Cyber Threat Intelligence. Fonte: elaboração do autor.

Existe uma relação hierárquica entre os indicadores de compromisso do nível técnico e tático, que é representada pela pirâmide da dor (originalmente chamada de *Pyramid of Pain - PoP*), representada na Figura 2.6. A ideia desta pirâmide é indicar o nível de dificuldade técnica com que os atacantes têm de lidar para desenvolver um ataque sem falhas e que não deixe rasto sobre a sua entidade. Paralelamente, se analisarmos sob a perspetiva do analista que defende os SI, a pirâmide representa o tipo de informação que têm de explorar para extrair o máximo de conhecimento sobre a atividade do atacante. Os níveis mais baixos incluem os indicadores técnicos, tais como endereços IP, nomes de domínio, valores de *hash* e outros artefactos da rede. Estes indicadores têm valores precisos, todavia são facilmente modificados pelo invasor. Os níveis mais altos da pirâmide incluem indicadores táticos, orientados ao comportamento, tais como TTP, que são padrões normalmente utilizados pelos atacantes. O grande desafio é descodificar estes comportamentos e modelá-los de forma a compreender o método de ataque, para antever estas condutas [48], [53].

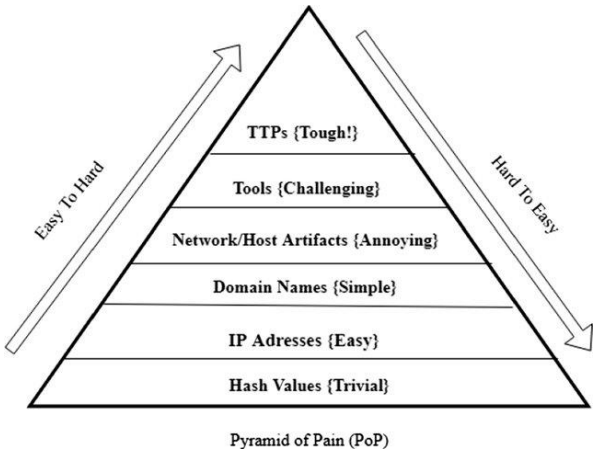

Figura 2.6 - Pirâmide da dor. Fonte: retirado de [53].

A criação de *threat intelligence* tem dois intervenientes, o consumidor e o produtor, que são responsáveis por procurar alertas sobre novas ameaças a partir de terceiros e pela partilha de conhecimento com bases em evidências, respetivamente. Como já foi referido anteriormente, todas as organizações devem desempenhar ambos os papéis, pois não deverá haver receio em partilhar este tipo de informação com as demais, só assim será possível acompanhar a evolução das ameaças [6]. De acordo com estudos recentes, cerca de 60 % das organizações já utilizam CTI e, cerca de 50% destas, têm equipas dedicadas para o efeito [54].

2.4.3.1. Problemas Associados

Gerar *threat intelligence* não é uma tarefa simples, pois carece de tempo e trabalho para colecionar os dados, rever, consolidar e integrá-los num formato compreensível e partilhável. O conhecimento produzido através das diferentes plataformas, envolvem indicadores de compromisso que permitem a deteção automática destes, de forma a serem consumidos pelas ferramentas presentes nos centros de operações de segurança. Todavia, a utilidade e fiabilidade destas informações para as ferramentas que absorvem estes dados têm algumas limitações:

- IoC têm confiabilidade e validade limitada – consoante a tipologia do IoC, a maioria destes são compostos por padrões fáceis de evitar pelos atacantes, como já foi referido. Portanto, as ferramentas que se baseiam apenas nestes dados conferem uma proteção temporária [55];
- Táticas, Técnicas e Procedimentos não detetáveis – as descrições que existem neste âmbito são muito abstratas e necessitam de indicadores mensuráveis para a deteção automática ser eficaz. Desta forma, os TTP normalmente apoiam as análises de ataques

ocorridos, ou seja, ajudam a realizar uma retrospetiva e a descrever o ciclo de vida do ataque, ao invés de fornecer mecanismos que sejam úteis em estratégias proativas [55].

Não obstante das problemáticas descritas anteriormente, o cerne da questão relativamente à *cyber threat intelligence* não é a inexistência de dados ou incapacidade recolher dados com recurso às ferramentas existentes, mas sim a grande variedade de dados que precisam de ser examinados para tornar todo o conhecimento de *threat intelligence* em passos operacionais [48]. Este dilema enfatiza a necessidade de existir uma maior automação com recurso ao *machine learning*, bem como a necessidade desta área evoluir através de abordagens direcionadas à análise dados. Desta forma, será possível extrair conhecimento de *threat intelligence* operacional para prever ataques, ao invés do que acontece atualmente, uma abundância de soluções de *threat intelligence* técnica, como consequência da maioria dos fornecedores destas tecnologias focarem-se, sobretudo, na deteção de IoC específicos [41], [52], [54].

Outro problema, é a grande quantidade de dados não uniforme e redundante, o que enaltece a necessidade de representar a informação num formato homogéneo, mas para tal é necessário haver um consenso sobre a representação e estrutura dos dados de *threat intelligence*, o que proporcionaria numa melhoraria da qualidade da informação e da automação das soluções analíticas. Estas situações refletem o investimento errado de muitas organizações neste sector, apesar dos mais de 20 mil milhões de dólares gastos anualmente pelas empresas em ferramentas de cibersegurança, este investimento é sobretudo em ferramentas tradicionais que são facilmente ultrapassadas pela nova geração de ataques [52]. Consequentemente, há que apostar mais na inovação e desenvolvimento de novas ferramentas, o que nem sempre é facilmente compreendido pelas empresas, uma vez que o retorno financeiro não é direto, mas evita grandes perdas económicas causadas por incidentes desta natureza, além do impacto na reputação da empresa que estes eventos causam [42].

Além destes problemas de cariz técnico e operacional, existem também problemas estratégicos e organizacionais das empresas que colocam entraves à extração de conhecimento, entre os quais se destacam:

- receio da desvantagem competitiva – as organizações receiam partilhar informações sobre ataques sofridos, uma vez que pode resultar numa desvantagem competitiva causada pela utilização destas informações pelas empresas rivais para criar uma publicidade negativa [42], [52];
- leis e problemas de privacidade – este fator deixa as organizações relutantes, pois não têm a certeza de qual o tipo de informação que podem tornar pública, sob a ameaça de sofrerem consequências jurídicas em relação á proteção de dados. As leis do país onde uma organização desenvolve a sua atividade, podem ser diferentes no país das

organizações com quem partilha os dados; ou mesmo dentro da mesma organização tendo esta equipas espalhadas em diferentes países que se regem por legislações diferentes [52];

- critérios de qualidade – algum do conhecimento partilhado não tem qualidade suficiente quanto à precisão, momento de partilha e clareza. Ou seja, por vezes o conhecimento partilhado é referente a ameaças que já não se verificam ou sobre as quais já existe mais conhecimento, e os dados partilhados não são úteis para a tomada de decisões;
- custos adicionais – a troca de informações sobre ameaças em tempo real tem custos, consequentemente a maioria das companhias consideram estes custos desnecessários, e que só fazem sentido nas grandes organizações. Esta situação é falaciosa, uma vez que quantas mais organizações contribuírem nesta partilha, mais rápido será a preparação para um eventual ataque;
- mecanismos legais ineficientes – por vezes as organizações não fazem acusações às entidades judiciais, porque acreditam que é uma perda de tempo dos recursos que dispõem para relatar estes incidentes a essas entidades. No entanto, por vezes pode ser benéfico para abrir um processo de investigação e saber mais sobre a ameaça e, eventualmente, punir os culpados legalmente de forma a evitar-se novos ataques.

2.5. Trabalhos de Investigação Semelhantes

As organizações e empresas tentam proteger-se dos ataques do dia zero, recorrendo a abordagens de análise do comportamento de diversos IoC na rede da organização, bem como na deteção de anomalias. O Anexo B ilustra diversas abordagens utilizadas para classificar as ameaças e desenvolver sistemas de reputação, tendo como base diferentes indicadores de compromisso examinados para devolver o nível de ameaça. Alguns destes estudos propõe a utilização de técnicas de *machine learning* para classificação autónoma de ameaças [9], [56].

CAPÍTULO 3

Arquitetura e Metodologia

3.1. Arquitetura de Integração

Existem dezenas de ferramentas e plataformas de *threat intelligence*, muitas destas desenvolvidas através da colaboração entre programadores e analistas informáticos sem qualquer teor comercial, e outras desenvolvidas por empresas para serem comercializadas como soluções empresariais [52], [54]. Devido às diferentes especificidades que cada plataforma visa compreender, não há uma plataforma que seja mais completa que as restantes para atender todos os processos de CTI. Todavia, a *Open Cyber Threat Intelligence* (OpenCTI) é uma das que se destaca em termos de abordagem holística na combinação de dados provenientes de diferentes fontes, bem como permite uma grande variedade de *queries* aplicáveis a diversos cenários [16], [57], [58].

Desta forma, esta será a ferramenta escolhida para o desenvolvimento do presente trabalho, não só devido às características apresentadas anteriormente, mas sobretudo pela decisão estratégica da empresa OutSytems tendo em conta as ferramentas que tem disponíveis no seu SOC. Além do mais, esta ferramenta permite ter um total controlo sobre os seus dados, ou seja, não são partilhados com terceiros através da plataforma, a não ser que a OutSystems o permita.

O presente trabalho visa implementar a plataforma OpenCTI nos servidores da organização e culminará com a criação de um conector do tipo *internal enrichment,* que será integrado na mesma e permitirá atribuir *scores* aos IoC de acordo com o seu nível de ameaça. Numa primeira instância, os IoC analisados serão do tipo endereços IPv4 que geraram eventos suspeitos na rede da organização e, por sua vez, carecem de análise pelo analista do SOC. Todavia, esta *framework* poderá ser estendida a outros tipos de IoC. Doravante, este conector será designado de "*OsThreatEnrichment".*

A Figura 3.1 representa a integração das diferentes ferramentas, como é possível observar a OpenCTI é o elemento central que permite a recolha, produção e armazenamento de *threat intelligence* sobre indicadores de compromisso de interesse para a organização. No âmbito deste trabalho, o processo de avaliação de risco quantitativo dos IoC através de um *score*, será despoletado sempre que os mecanismos de defesa detetarem um evento suspeito nas redes da organização e, consequentemente, gerarem um alerta para os analistas no SOC (esta fase está representada a vermelho na Figura 3.1). Por sua vez, os indicadores de compromisso associados a esse alerta serão submetidos para análise na plataforma OpenCTI, através de um automatismo ou manualmente pelo analista. Na fase seguinte, o conector "*OsThreatEnrichment*" será responsável por obter informações em plataformas externas de *threat intelligence*, denominadas *Feeds* na Figura 3.1, e correlacionar com informação interna da organização, caso este IoC seja reincidente (esta fase esta representada a amarelo na Figura 3.1). Por fim, após a recolha de dados externos e internos, o conector correrá o algoritmo que permitirá calcular a pontuação da ameaça (*ThreatScore*), bem como a pontuação de confiança (*TrsutRating*) associada a essa pontuação.

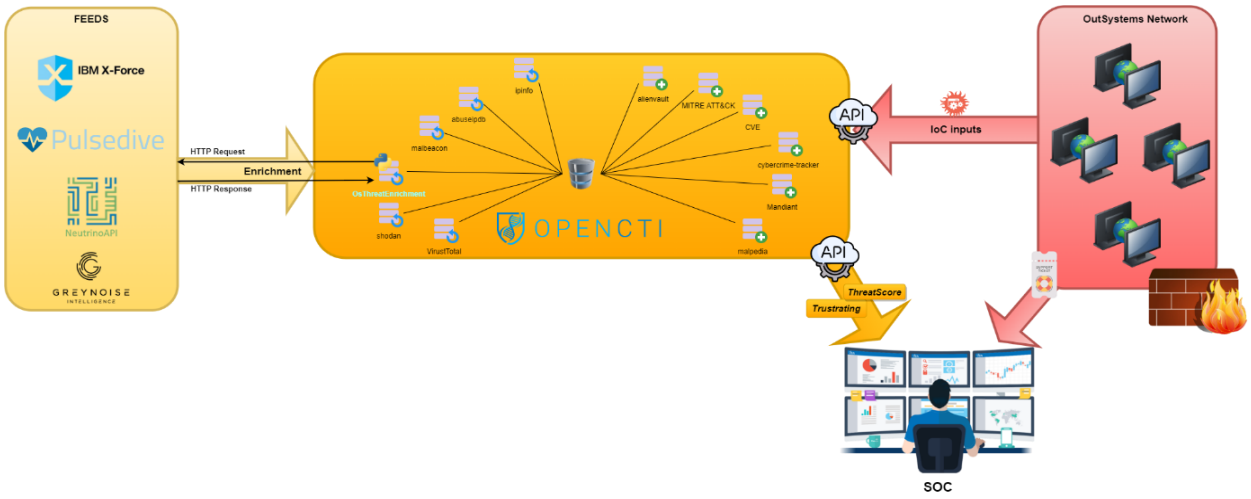

Figura 3.1- Diagrama geral da arquitetura de integração das diferentes ferramentas com a plataforma OpenCTI. Fonte: elaboração do autor.

3.1.1. Características e Funcionamento da Plataforma OpenCTI

Esta é uma plataforma *Open Source Intelligence* (OSINT) inicialmente desenvolvida pela *Agence Nationale de la Sécurité des Systèmes d'Information* (ANSSI) do governo francês, em colaboração com *Computer Emergency Response Team of the European Union* (CERT-EU) [59]. Posteriormente, foi disponibilizado o código fonte [60] de forma a existir contribuições de diferentes voluntários com o objetivo de ficar mais completa a nível da partilha de informação.

Os dados desta plataforma são estruturados de acordo com a linguagem STIX e podem ser integrados com outras ferramentas. O principal objetivo desta plataforma é permitir aos utilizadores correlacionar informações técnicas e não técnicas a partir de outras ferramentas, ou através dos seus próprios conjuntos de dados relativos a *threat intelligence*. Como resultado desta correlação, novos atributos podem ser inferidos em tempo real de forma a extrair conhecimento sobre as ameaças.

Na Figura 3.2 está representada a arquitetura da ferramenta, como se pode verificar a API é a componente central, construída em *NodeJS* e recorre à implementação da linguagem *GraphQL.* É com recurso a esta linguagem que os clientes podem interagir com a base de dados e com o sistema de mensagens (*broker*). Os trabalhadores (*workers*) são processos autónomos e assíncronos de *queries*, desenvolvidos na linguagem *python,* que consomem mensagens do *broker RabbitMQ.* Podem ser lançados o número de *workers* que forem pertinentes para aumentar o desempenho da ferramenta, no entanto haverá um limite a partir do qual o desempenho será limitado pela taxa de transferência da base de dados (*ElasticSearch*) [61].

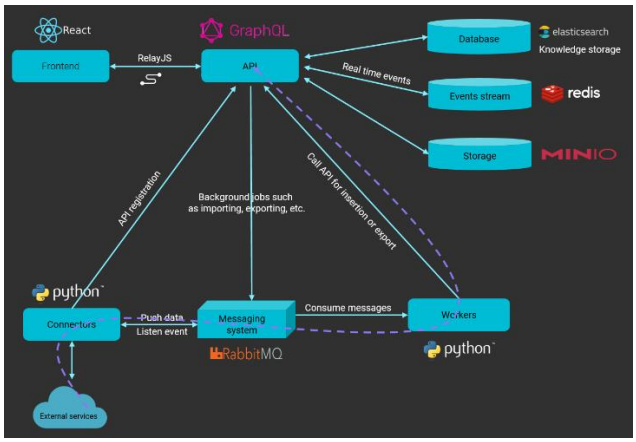

Figura 3.2 - Arquitetura da ferramenta OpenCTI. Fonte: retirado de [61].

Em termos do universo de dados de *threat intelligence*, a OpenCTI tem sido enriquecida com novas funcionalidades que permitem obter dados com regularidade, portanto não está estabelecido um limite de dados a recolher e a serem trabalhados pela mesma. Além do mais, quanto maior o número de dados, maior a qualidade destes para estabelecer correlações e, numa fase posterior, serem utilizados nas regras de inferência. Adicionalmente, sendo esta ferramenta pública, não necessita de subscrição para ter acesso a todas as funcionalidades e permite a integração de outras plataformas através de conectores, que se podem dividir em dois grandes grupos: *internal enrichment* e *external import*. O primeiro grupo tem como objetivo enriquecer os dados presentes na plataforma com informação adicional e contextual, enquanto os segundos permitem importar novos IoC reportados noutras plataformas.

3.2. Implementação

3.2.1. Configuração da Plataforma

Esta ferramenta foi implementada num Docker, instanciado numa máquina virtual e a correr nos servidores da Amazon Web Services (AWS) sob o domínio da OutSystems. Contudo, para desenvolver o código, testar e implementar o conector, o ambiente da plataforma OpenCTI foi replicado numa máquina local antes de ser exportado para os servidores da AWS.

A imagem referente à plataforma OpenCTI pode ser instalada no Docker através da linha de comandos, sendo esta a forma convencional de correr instâncias de imagens neste *software*. Todavia, para configuração das variáveis de ambiente e configuração das dependências do *container*, recorreu-se à utilização do Portainer. Por sua vez, este é um *container* que pode ser instanciado através de uma imagem no Docker, o que o torna numa página *front-end* de gestão. Consequentemente, permite gerir os restantes *containers* de forma mais intuitiva devido à sua interface *user-friendly.* Após a instalação, este pode ser acedido localmente no *browser* através da porta 943, ou seja, através do endereço http://localhost:943/ (Figura 3.3). Por último, procedeu-se à instalação da plataforma OpenCTI na máquina virtual.

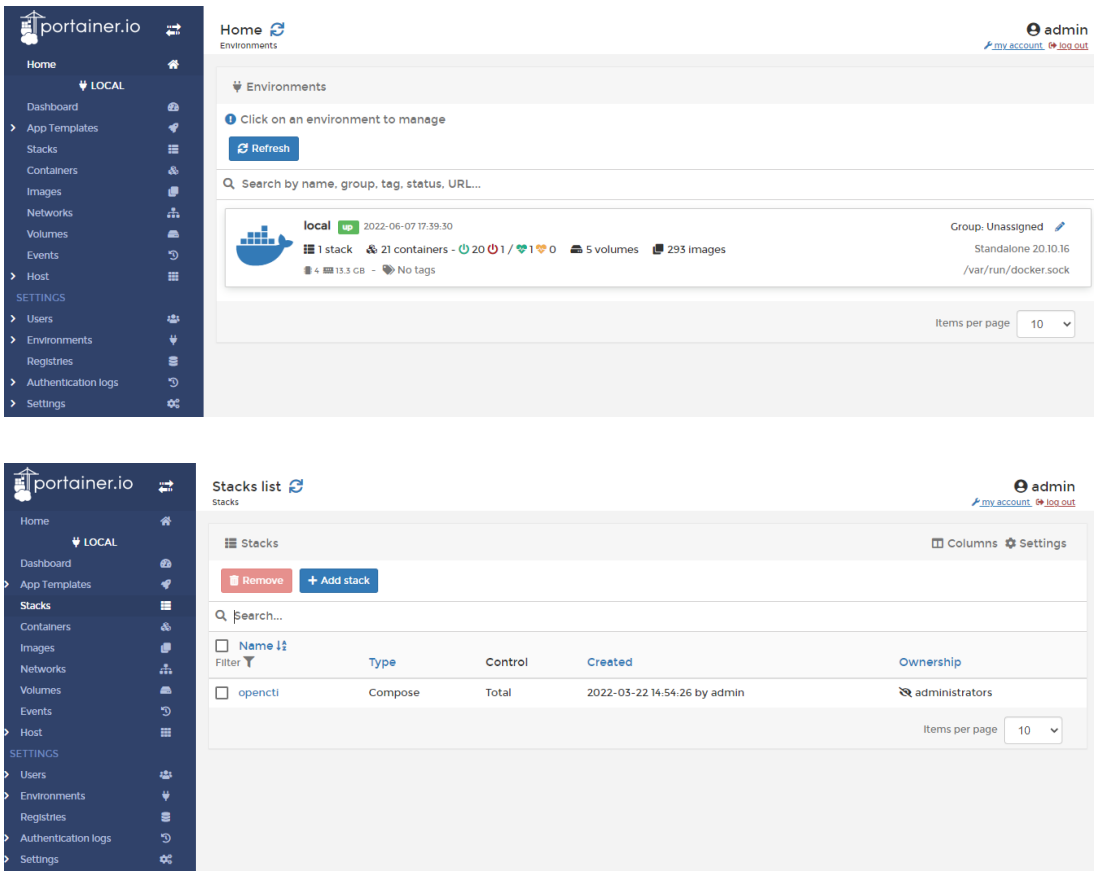

Figura 3.3 - Interface do Portainer. Fonte: elaboração do autor.

Numa primeira instância foi efetuada a transferência do ficheiro "docker-compose.yml" da OpenCTI[2].Em seguida, na interface do Portainer foi adicionado um novo *stack,* cujo nome foi o mesmo da plataforma, ou seja, "OpenCTI". Antes de ser feito o *deploy* do *stack* com o ficheiro transferido anteriormente, foi necessário configurar as variáveis de ambiente num ficheiro específico para o efeito (".env"). Um exemplo deste ficheiro pode ser encontrado no GitHub[3] [60].

Algumas das variáveis de ambiente requerem *tokens* de identificação universal *Universal Unique Identifier* (UUID), como tal foram gerados numa página de internet específica para o efeito[4]. Por fim, foi feito o *deploy* do *stack* e a plataforma OpenCTI ficou disponível para ser acedida na máquina local através do endereço http://localhost:80/. Numa fase posterior, foram conectadas outras ferramentas que são incorporadas como conectores *internal enrichment* e *external import*, consoante a sua finalidade e conforme descrito previamente. Estas ferramentas estão documentadas na página do código da OpenCTI [60].

Com a finalidade de desenvolver, testar e validar o código do conector compatível com a plataforma, foi utilizada a extensão *Containers* integrada no Microsoft Visual Studio Remote. Esta abordagem permite, em simultâneo, implementar um *container* com o ficheiro *python* em desenvolvimento e interagir com a plataforma OpenCTI instanciada na máquina local. Após o desenvolvimento do conector, este foi clonado para a máquina virtual a correr nos servidores da AWS.

3.2.2. Validação das Fontes Externas

Para tornar o trabalho mais robusto na inferência do nível das ameaças, recorreu-se a ferramentas externas que têm planos grátis limitados e planos pagos com poucas ou nenhumas limitações. Estas ferramentas são importantes para consultar a informação externa sobre os indicadores de compromisso. Todavia, variam quanto ao tipo de informação disponibilizada e quanto à fiabilidade, como tal para determinar quais as mais adequadas no âmbito deste estudo, as mesmas foram analisadas e testadas. O estudo foi iniciado com as seguintes ferramentas: Greynoise; IBM X-Force Exchange; MyIP; Neutrino; Pulsedive; Polyswarm e RiskIQ. Destas ferramentas, foram descartas a Polyswarm e RiskIQ, uma vez que não eram as mais adequadas para analisar o risco associado aos IoC referentes a endereços IPv4 através da metodologia deste trabalho. Além do mais os planos gratuitos destas ferramentas não permitem ter acesso à avaliação do risco em termos qualitativos e quantitativos.

[2] Ficheiro retirado do site https://github.com/OpenCTI-Platform/docker/blob/master/docker-compose.yml a 06/03/2022
[3] Ficheiro retirado do site https://github.com/OpenCTI-Platform/docker/blob/master/.env.sample a 06/03/2022
[4] Informação retirada do site https://www.uuidgenerator.net/version4 a 06/03/2022

As restantes cinco ferramentas destacam-se por terem planos gratuitos que permitem retirar uma panóplia de informação útil sobre as ameaças e, mesmo utilizando estes planos básicos, são passíveis de serem integradas na OpenCTI da organização. As características gerais destas ferramentas estão descritas no Anexo A da presente dissertação.

De forma a uniformizar os resultados provenientes das plataformas em análise, e com o objetivo de realizar comparações estatísticas, foi necessário efetuar adaptações ao tipo de classificações atribuídas pelas ferramentas. Portanto, a classificação quantitativa associada ao risco obtida da ferramenta IBM X-Force Exchange foi traduzida para uma escala qualitativa, permitindo assemelhar-se à plataforma Neutrino que apresenta originalmente os resultados em ambos os formatos. Os intervalos de valores apresentados pela IBM X-Force Exchange foram traduzidos na seguinte classificação qualitativa:

- **"low"** (0< Risco <4);
- **"medium"** (4≤ Risco ≤ 6);
- **"high"** (6 < Risco ≤ 8);
- **"critical"** (8 < Risco ≤ 10)

Relativamente à ferramenta Greynoise, de acordo com a documentação, além da classificação qualitativa devolvida por esta ferramenta ("*benign"*, "*malicious"*, "*unknown"*), é necessário ter em consideração outros dois parâmetros reportados nos resultados, "*Riot"* e "*Noise"* [62]. O primeiro é um atributo desenvolvido pela plataforma em questão, que permite informar os utilizadores sobre endereços IPv4 utilizados por serviços comerciais e que, muito provavelmente, não estão a atacar a rede onde são detetados. Quanto ao segundo parâmetro, "*Noise*", está associado aos endereços IPv4 que costumam rastrear a internet, através da correlação de dados sobre estes endereços, é possível classificar as suas atividades como suspeita ou não suspeita. Consequentemente, devido à relevância destas duas variáveis e de acordo com as indicações da plataforma em questão, para atribuir um *score* quantitativo adequado a um endereço IPv4 deve-se ter em conta a conjugação destes três elementos. Como tal, a interpretação destes parâmetros e a sua aplicabilidade ao *ThreatScore* (TS) será descrita no subcapítulo 3.4.2.

3.2.2.1. Seleção de endereços IPv4 maliciosos

Para validar as fontes externas foram realizados dois ensaios independentes com um universo total de 7608 endereços IPv4 recolhidos, sem a ocorrência de duplicados. As amostras referentes ao primeiro e segundo ensaio foram de 2687 e 4921 endereços IPv4, respetivamente. Os endereços foram obtidos de *blacklists* que são atualizadas diariamente com novos endereços suspeitos, desta forma as fontes

consultadas para o ensaio 1 foram: Snort.org[5], *Internet Storm Center*[6]. Para o ensaio 2, foram consultadas as seguintes fontes: GitHub[7], *Emerging Threats*[8] e plataforma *Clean Talk*[9]. Além destas fontes consultadas para cada um dos ensaios, houve uma fonte em comum para ambos os ensaios, a *Binary Defense*[10].

Por conseguinte, com o objetivo de filtrar os 2687 endereços IPv4 selecionados para o ensaio 1, estes foram submetidos a análise na plataforma AbuseIPDB[11] e o critério de seleção baseou-se na interseção de duas regras (Quadro 3.1):

1. **Regra 1** - *Score* reportado pela plataforma AbuseIPDB igual a 100. A racional desta regra é selecionar os endereços IPv4 sobre os quais existe um maior grau de certeza relativamente às suas ações maliciosas [63]. Além do mais, esta premissa permitirá reduzir os falsos positivos na amostra em análise. Consequentemente, dos 2687 endereços, somente 1164 cumpriam este critério.
2. **Regra 2** - Rácio entre o total de reportes que classificaram um determinado endereço IPv4 como malicioso e o número de utilizadores distintos que reportaram esse mesmo endereço como malicioso. Este valor deverá ser igual ou inferior a 3. A racional desta regra é selecionar os endereços IPv4 que foram reportados por diferentes utilizadores/organizações, ou seja, maioritariamente tentam atacar diferentes alvos e não apenas uma organização. Consequentemente é dedutível que é um endereço que está referenciado como maligno entre diferentes organizações e, portanto, existe maior confiança sobre as suas más intenções. Após aplicar este segundo critério de seleção, a amostra final foi de 243 endereçosIPv4.

O ensaio 2 foi semelhante em termos de desenho experimental ao ensaio 1, com a exceção de que além da amostra de 4921 endereços IPv4 submetidos a análise na plataforma AbuseIPDB, estes também foram analisados na ferramenta VirusTotal[12]. O objetivo desta dupla análise foi validar o grau da ameaça dos indicadores recolhidos com recurso a duas ferramentas bastante utilizadas na área de cibersegurança e, mais uma vez, reduzir a interferência de falsos positivos. Posteriormente, os critérios de seleção da amostra final assentaram nas Regras 1 e 2 descritas no ensaio anterior, bem como numa terceira regra adicional: **Regra 3** - Mínimo 3 reportes como maliciosos na plataforma VirusTotal. A conjunção das 3 regras resultou na Condição – C2 - descrita no Quadro 3.1.

[5] Informação retirada do site https://*Snort.org* a 20/03/2022
[6] Informação retirada do site https://isc.sans.edu a 20/03/2022
[7] Informação retirada do site https://github.com/stamparm/ipsum a 01/04/2022
[8] Informação retirada do site https://EmergingThreats.com a 01/04/2022
[9] Informação retirada do site https://cleantalk.org a 01/04/2022
[10] Informação retirada do site https://binarydefense.com a 20/03/2022 e 01/04/2022
[11] Informação retirada do site https://www.abuseipdb.com/ a 20/03/2022 e 01/04/2022
[12] Informação retirada do site https://www.virustotal.com/gui/home/search a 01/04/2022

Quadro 3.1 - Resultados das Regras para a seleção de endereços IPv4 maliciosos em ambos os ensaios. Fonte: elaboração do autor.

	Regra	SubTotal	Condição	Total
Ensaio 1	**R1.** Reportados no AbuseIPDB com um score =100	1164	**C1** = R1 ∩ R2	243
	R2. (Rácio Total Reportes para um IP / número de utilizadores distintos a reportar esse mesmo IP) <= 3	1841		
Ensaio 2	**R1.** Reportados no AbuseIPDB com um score =100	3054	**C2** = R1 ∩ R2 ∩ R3	535
	R2. (Rácio Total Reportes para um IP / número de utilizadores distintos a reportar esse mesmo IP) <= 3	2600		
	R3. Reportados como maliciosos no VirusTotal >=3	1263		

Os ensaios 1 e 2 foram modelados em *Business Process Model and Notation* (BPMN), para permitir visualizar graficamente a diferença entre os processos de seleção dos endereços IPv4 em ambos os ensaios. Como tal, os ensaios 1 e 2 estão representados nas Figuras 3.4 e 3.5, respetivamente.

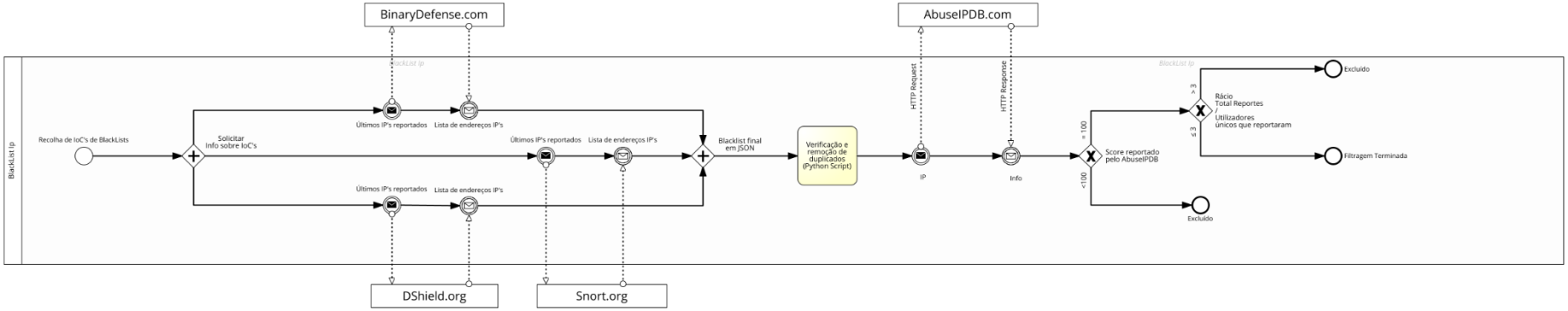

Figura 3.4 - BPMN do Ensaio 1. Fonte: elaboração do autor.

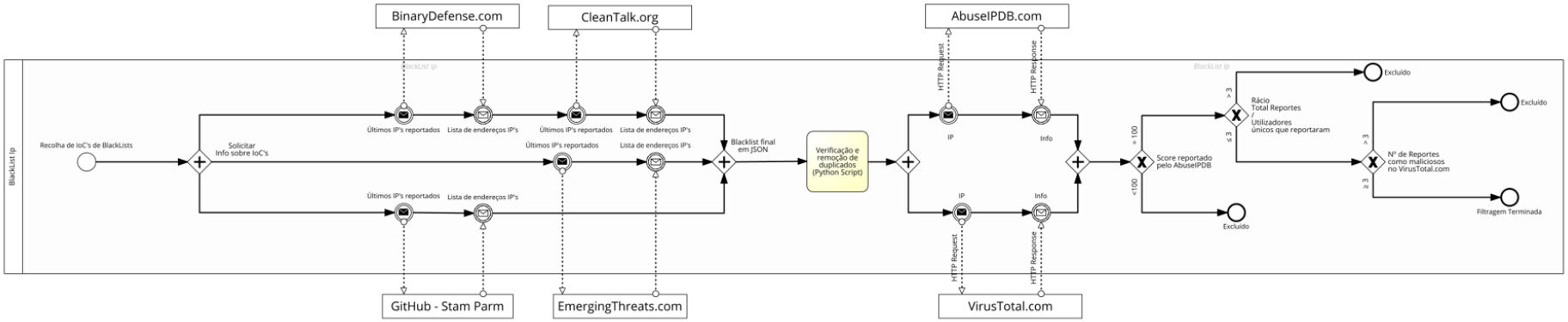

Figura 3.5 - BPMN do Ensaio 2. Fonte: elaboração do autor.

3.2.2.2. Seleção das Fontes Externas

As amostras finais dos ensaios 1 e 2 realizados na secção anterior, contabilizaram 243 e 535 endereços IPv4, respetivamente. Estes endereços foram analisados nas 5 plataformas externas em avaliação, mencionadas anteriormente. No ensaio 1, a plataforma GreyNoise e Neutrino destacaram-se ao detetar 49% e 91% dos endereços IPv4 como ameaças, respetivamente. Por outro lado, a plataforma MyIP detetou apenas 7% dos endereços IPv4 como ameaça. Quanto às plataformas IBM X-Force Exchange e Pulsedive, estas consideraram que 43% e 14% destes endereços, respetivamente, representavam um risco médio ou superior (Quadro 3.2).

Relativamente ao ensaio 2, as ferramentas que apresentaram melhores resultados foram a GreyNoise e a Neutrino que reportaram 71% e 99% destes endereços, respetivamente, como ameaças. Em contrapartida, a plataforma MyIP apresentou os piores resultados ao detetar apenas 6% dos endereços como ameaças. Quanto às restantes ferramentas, a IBM X-Force Exchange reportou 66% dos endereços IPv4 com risco médio ou superior, sendo que 31% destes tinham um risco crítico. Por sua vez, a plataforma PulseDive classificou 40% dos endereços sem risco e outros 40% com risco baixo.

Em suma, teoricamente, todos os endereços analisados eram maliciosos de acordo com as plataformas de triagem (AbuseIPDB e VirusTotal) utilizadas na seleção das amostras. Todavia, existiram diferenças estatisticamente significativas na classificação atribuída pelas ferramentas à amostra de endereços, ou seja, não foram todos classificados como tal. De acordo com os resultados reportados e resumidos no Quadro 3.2, as ferramentas que se destacaram pela positiva e melhoraram o desempenho entre ensaios, foram a Neutrino e a GreyNoise. Adicionalmente, a IBM X-Force Exchange apesar de não ter tido resultados tão bons quanto estas duas ferramentas, também melhorou consideravelmente do ensaio 1 para o ensaio 2.

Quadro 3.2 - Compilação de Resultados dos Ensaios 1 e 2. Fonte: elaboração do autor.

Ferramenta	Resultado (Ensaio 1 / Ensaio 2)		Hipótese Nula do Teste Estatístico (H0)	Teste Estatístico	Decisão (Sig.[a])
IBM XForce Exchange	**"low"** (0< Risco <4) = 137 / 236; **"medium"** (4≤ Risco ≤ 6) = 34 / 79; **"high"** (6 < Risco ≤ 8) = 24 / 53; **"critical"** (8 < Risco ≤ 10) = 48 /167		As categorias qualitativas definidas na variável IBM ocorrem com iguais probabilidades.	One-Sample Chi-Square	Rejeitar H0 (0.000)
GreyNoise	**"Malicious"**= 118 / 380	**"Benign"**= 1 / 155	As categorias = "malicious"; "unknown" e "benign" definidas na variável GreyNoiseScoreClassification ocorrem com probabilidades iguais de 0,5 e 0,5.	One-Sample Chi-Square	Rejeitar H0 (0.000)
	"Unknown"= 124 / 0				
MyIP	**"Yes"** = 16 / 34	**"No" =** 227 / 501	As categorias "yes" e "no" definidas na variável MyipBlackList ocorrem com iguais probabilidades de 0,5 e 0,5.	One-Sample Binomial	Rejeitar H0 (0.000)
Neutrino	**"Listed"** = 220 / 528	**"Not Listed"** = 23 / 7	As categorias "listed" e "not listed" definidas na variável NeutrinoListedBlockList ocorrem com iguais probabilidades de 0,5 e 0,5.	One-Sample Binomial	Rejeitar H0 (0.000)
Pulsedive	**"unknown"** (Risco = 0) = 0 / 2; **"none"** (Risco = -1) = 74 / 216; **"low"** (Risco = 1) = 60 / 212; **"medium"** (Risco =2) = 34 / 76; **"high"** (Risco = 3) = 1 / 6; **"critical"** (Risco = 4) = 1 / 10; "**not found**" = 73 / 13		As categorias definidas na variável PulsediveRisk ocorrem com iguais probabilidades	One-Sample Chi-Square	Rejeitar H0 (0.000)

3.2.3. Dados Recolhidos das Fontes Externas

Uma vez selecionadas as plataformas a utilizar como fonte de informação externa para avaliar o risco dos IoC suspeitos, o conector "*OsThreatEnrichment*" será responsável por recolher os dados destas plataformas. A recolha de informação processa-se através de mensagens de *Hypertext Transfer Protocol* (HTTP), ou seja, o conector ao receber um novo indicador de compromisso para análise, solicitará através do protocolo em questão informação ao servidor das fontes externas.

Os dados recolhidos de cada uma das fontes externas, representados no diagrama da Figura 3.6, são de diferentes tipos consoante a informação existente. Como tal, o risco da ameaça reportado por estas plataformas pode ser disponibilizado de forma qualitativa e/ou quantitativa. Posteriormente, estes dados serão normalizados para serem utilizados pelo algoritmo implementado no presente conector e que é responsável por calcular a pontuação da ameaça.

A plataforma Neutrino e IBM XForce Exchange disponibilizam labels associadas à atividade do endereço IPv4, ou seja, caracterizam o tipo de ações no qual este foi detetado. Desta forma, as labels serão recolhidas pelo conector para contextualizar informação sobre a ameaça. Adicionalmente, será recolhida informação pertinente sobre o IoC, disponibilizada pela plataforma IBM XForce Exchange. Esta informação será disponibilizada pelo conector na plataforma OpenCTI no formato de um relatório, que estará disponível para o analista consultar sempre que necessário.

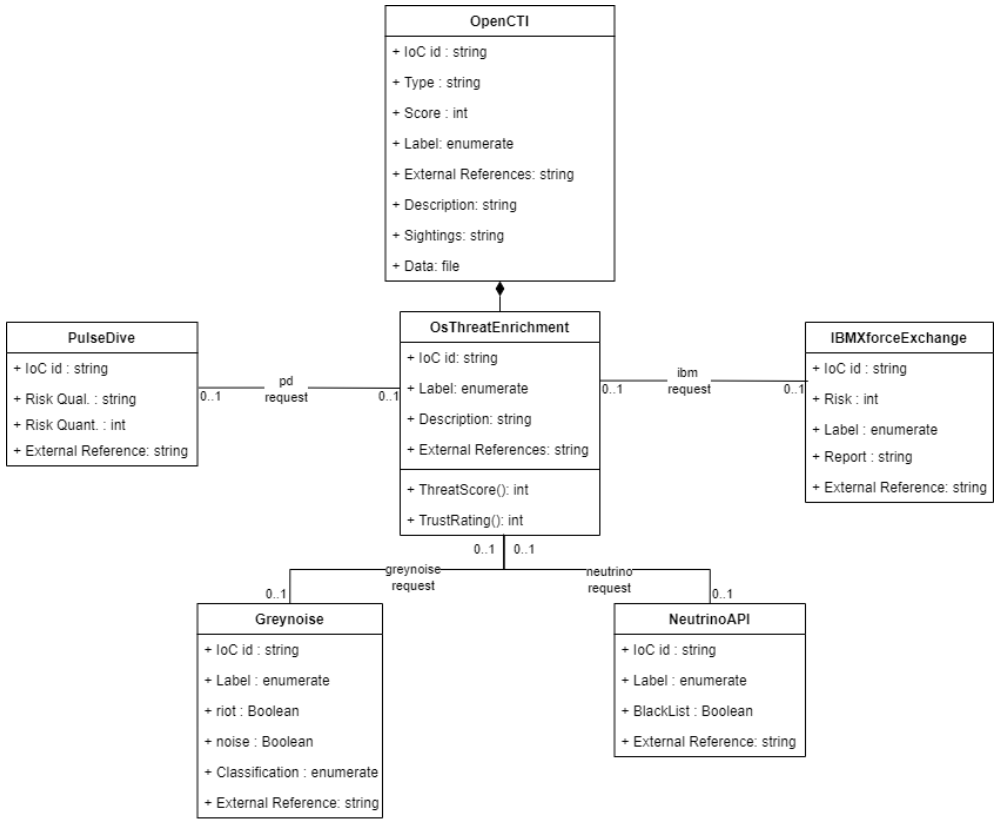

Figura 3.6 - Diagrama de Classes com atributos recolhidos das Plataformas Externas para a plataforma OpenCTI. Fonte: elaboração do autor.

3.3. Pontuação de Risco da Ameaça - *ThreatScore*

O IoC referente a um endereço IPv4 será analisado para verificar três propósitos: o primeiro é calcular qual o risco da ameaça - *ThreatScore* (TS); o segundo é calcular o nível de confiança do risco calculado e por último contribuir para a produção de *threat intelligence* da própria organização, que poderá ser utilizada em análises de *Big Data*.

O cálculo do risco de um IoC que é considerado suspeito pelos sistemas de defesa da rede da organização OutSystems é realizada pelo conector *OsThreatEnrichment,* desenvolvido no presente trabalho. A fórmula de cálculo foi determinada com base em duas fontes de informação, internas e externas, obtidas através do registo histórico do indicador de compromisso nos registos da organização e em registos provenientes de plataformas externas, respetivamente. Os parâmetros extraídos destas fontes de informação, bem como a sua contribuição para o *ThreatScore*, serão abordados com mais detalhe nos subtópicos seguintes.

3.3.1. Registos Internos da Rede da Organização

Através dos registos recolhidos da rede da organização referentes a um determinado endereço IPv4, será necessário ter em consideração a variável *History* (*His).* Esta variável baseia-se no espaço temporal entre observações do mesmo IoC, ou seja, em que momentos um determinado endereço IPv4 foi detetado nos sistemas de defesa da organização, e consequentemente, despoletou eventos suspeitos. É mensurável em dias, como tal representa a diferença de tempo entre as duas ocorrências. Os valores atribuídos à variável *His* com base na última observação de um dado IoC, submetido a análise, estão descritos na Quadro 3.3. O peso desta variável é representado pela incógnita $w2$.

Quadro 3.3 - Valores da Variável History de acordo com o intervalo de tempo. Fonte: elaboração do autor.

Data atual – Data última observação = *x* dias	**Valor da variável *His***
$x \leq 7$	100
$7 < x \leq 14$	75
$14 < x \leq 30$	50
$30 < x \leq 45$	25
Nunca observado ou $x > 45$	0

Outra variável extraída da plataforma OpenCTI, é o *ThreatScore* atribuído anteriormente ao IoC em análise. Na fórmula 3.1, esta variável está representada pelo parâmetro "*prevThreatScore"* e o peso atribuído esta variável está representado pela incógnita $w1$.

Estas duas variáveis são obtidas da plataforma OpenCTI e trabalhadas através da seguinte fórmula:

$$\text{Fonte Interna} = [(\text{prevThreatScore} * w1) + (\text{His} * w2)] \quad (3.1).$$

Onde,

prevThreatScore = [0, 100];

His = {0, 25, 50, 75, 100};

w1 = 0.5;

w2 = 0.5.

Resumindo, os registos presentes na plataforma OpenCTI - fonte interna - contribuem para o cálculo do *ThreatScore* somente quando ocorre a análise de um endereço IPv4 que já conste na plataforma. Consequentemente, um IoC que não conste na plataforma OpenCTI e que seja introduzido pela primeira vez, não terá em conta a informação proveniente das fontes internas. Nesta situação, o *ThreatScore* será calculado apenas com base nas fontes externas (equação 3.2), como veremos no tópico seguinte.

3.3.2. Registos Históricos Externos Associados

O registo histórico externo sobre um endereço IPv4, assume um papel de destaque para permitir atingir os três propósitos descritos anteriormente, sobretudo no âmbito de produção de *threat intelligence*, uma vez que permite obter conhecimento sobre a atividade dos atores associados e/ou responsáveis pelos indicadores em análise. Como tal, com o objetivo de ter mais confiança no nível de risco relativo a um determinado endereço IPv4, além dos registos da própria organização, é igualmente importante consultar o registo histórico de ferramentas externas de *threat intelligence*, já validadas anteriormente. Das 5 fontes analisadas, a ferramenta MyIP foi excluída devido aos fracos resultados apresentados no âmbito deste estudo, como tal foram utilizadas as seguintes: IBM X-Force Exchange, Neutrino, Pulsedive e GreyNoise. A contribuição das fontes externas para o *ThreatScore,* é calculada com base na equação 3.2.

$$\text{F. Ext.} = [(\text{Neutrino} * \text{w'1}) + (\text{GreyNoise} * \text{w'2}) + (IBM * \text{w'3}) + (Pulsedive * \text{w'4})] * 100 \quad (3.2).$$

A equação (3.2) sugere que as variáveis representadas pelo nome das plataformas correspondem a valores binários (0 ou 1) no caso da Neutrino e da GreyNoise. Como tal, 1 indica a existência de um endereço IP na base de dados, enquanto 0 representa a ausência de um endereço IP nas bases de dados da plataforma. "Não listado" não significa que um endereço IPv4 não esteja associado a atividades maliciosas, apenas não há registos de atividades dessa tipologia associadas a este. Quanto às variáveis representadas pela IBM e Pulsedive, estas assumem os valores do risco que atribuem ao endereço IPv4 submetido para análise, ou seja, entre 0 e 10 e -1 e 4, respetivamente.

As incógnitas w' 1, w' 2, w' 3 e w' 4 representam os pesos máximos, ordenados de forma decrescente, atribuídos às plataformas Neutrino, GreyNoise, IBM X-Force Exchange e Pulsedive, respetivamente. Como consequência dos resultados quantitativos devolvidos pelas plataformas IBM X-Force Exchange e Pulsedive não estarem na mesma escala, optou-se por normalizar os resultados destas duas ferramentas e convertê-los para valores tendo em conta o peso máximo que foi atribuído a cada uma destas ferramentas (w' 3 e w' 4). As fórmulas de conversão podem ser consultadas no Quadro 3.5.

No caso da plataforma Neutrino, esta ferramenta reporta os resultados de forma booleana, ou seja, 1 se o endereço IP está listado na *blacklist* e 0 caso não esteja, o que corresponde aos valores de 0.4 e 0, respetivamente, tendo em consideração o valor máximo atribuído a esta plataforma.

Quanto à plataforma GreyNoise que tem o segundo maior peso (w' 2), no caso do endereço IPv4 estar listado, esta classifica a ameaça através de uma variável qualitativa, "*malicious*", "*unknown"* e "*benign".* Porém, além da ameaça, este apresenta nos resultados duas variáveis booleanas extra, *Riot* e *Noise*, que devem ter-se em consideração na conversão do *score* qualitativo para quantitativo, tal como descrito anteriormente. De acordo com as indicações da plataforma [64], deve tomar-se a decisão em concordância com os valores registados no Quadro 3.4. Podem ser consultadas informações adicionais sobre as ferramentas utilizadas, bem como a normalização dos resultados obtidos das mesmas, no Anexo A - Ferramentas Externas Consultadas - da presente dissertação.

Quadro 3.4 - GreyNoise - Conversão dos resultados qualitativos para quantitativos. Fonte: elaboração do autor.

Observação	Riot	Noise	Recomendação (Valor de *w*' 2)	Racional
GreyNoise não tem informação sobre o endereço IPv4	False	False	Definir como máxima prioridade, ou seja, máximo risco atribuído (0.3).	Poderá tratar-se de uma tentativa de ataque propositada e não apenas uma tentativa oportunista de um ator à procura de vulnerabilidades.
GreyNoise não tem informação sobre o endereço IPv4. Todavia, este está atribuído a um serviço comercial que o opera	True	False	Definir como prioridade média, ou seja, risco médio atribuído (0.15).	É improvável que esse endereço IP esteja a realizar uma atividade maliciosa, a menos que o provedor tenha a sua infraestrutura comprometida ou permita que fontes externas adicionem conteúdo malicioso na mesma.
Classificação "*benign"*	-	True	Definir como prioridade baixa, ou seja, atribuir o valor mais baixo (0.1).	Tentativa de "varredura" oportunista por um ator conhecido. Como o ator está categorizado como "benigno", as suas ações podem ser consideradas desse cariz na maioria dos casos.
Classificação "*malicious"*	-	True	Definir como prioridade alta, ou seja, retirar um valor ao máximo (0.3).	Poderá ser uma tentativa oportunista de um ator que está a verificar a Internet. No entanto, poderá estar subjacente uma intenção maliciosa.
Classificação "*unknwown"*	-	True	Definir como prioridade média, ou seja, risco médio atribuído (0.15).	Provavelmente foi apenas uma tentativa oportunista de um ator que está a verificar a Internet. A atividade não parece ter intenção maliciosa com base no que é observado pela GreyNoise. Não é necessariamente preocupante para a organização. Todavia, não foi de um ator conhecido, consequentemente deve ser tratado com cautela.
Qualquer Classificação	False	True	Definir como máxima prioridade, ou seja, máximo risco atribuído (0.3).	Uma conexão de saída foi feita para um dispositivo conhecido que está a "varrer" a Internet. Independentemente da classificação do endereço IP, nos dados disponibilizados pela GreyNoise, este é possivelmente um comportamento indesejado e deve ser investigado de imediato.

3.3.3. Pontuação da Ameaça – *ThreatScore*

Este passo consiste na agregação das pontuações provenientes das equações Fonte Interna (equação 3.1) e Fontes Externas (equação 3.2), de forma a produzir o *ThreatScore* do endereço IPv4 em análise. Por conseguinte, quando o endereço IPv4 já existe na plataforma OpenCTI, foi definido que os *scores* provenientes das fontes internas e externas contribuiriam equitativamente para o *ThreatScore*, ou seja, 50% cada uma das fontes, representados por α e β, respetivamente. A única exceção é quando o endereço IPv4 ainda não existe na plataforma e está a ser analisado pela primeira vez. Nessa situação α e β assumem os valores de 0 e 1, respetivamente, como consequência do que foi explicado no tópico 3.4.1 – Registos Internos da Rede da Organização.

Como tal, temos:

$$\text{Threat}Score(TS) = \alpha(Score\ Fonte\ Interna) + \beta(Score\ Fontes\ Externas) \quad (3.3).$$

Onde, o *Score Fonte Interna* e o *Score Fontes Externas* presentes na equação 3.3, correspondem à equação 3.1 e 3.2, respetivamente. Portanto a equação completa para o cálculo do TS traduz-se na equação 3.4, que é representada na seguinte forma:

$$T.S = \alpha[(prevTS * w1) + (His * w2)] + \beta[(Neutrino * w'1) + (GreyNoise * w'2) + (IBM * w'3) + (PulseDive * w'4) * 100] \quad (3.4).$$

Em forma de resumo, os pesos correspondentes a cada uma das incógnitas encontram-se descritos na Quadro 3.5.

Quadro 3.5 - Valor das Incógnitas e pesos atribuídos às diferentes variáveis. Fonte: elaboração do autor.

Variável	**valor**	
$\boldsymbol{\alpha}$	{0, 0.5}	
$\boldsymbol{\beta}$	{0.5, 1}	
IBM	$\frac{X}{10}$	*X*= [0, 10]
Pulsedive	$\frac{abs(X)}{4}$	*X*= [-1, 4]
Neutrino; GreyNoise	{0, 1}	
Peso	**valor**	
$\boldsymbol{w1}$	0.5	
$\boldsymbol{w2}$	0.5	
$\boldsymbol{w'1}$	0.4	
$\boldsymbol{w'2}$	{0.3, 0.15, 0.1} – ver Quadro 3.4	
$\boldsymbol{w'3}$	0.2	
$\boldsymbol{w'4}$	0.1	

O *ThreatScore* traduz a ameaça de um endereço IPv4 para a organização, como tal ao registar a pontuação da ameaça na plataforma OpenCTI, permite que este valor seja dinâmico e, sempre que seja realizado um novo pedido, este seja atualizado. Além do mais, a utilização deste valor de reputação tem outras vantagens: decidir que tipo de privilégios podem ser atribuídos aos utilizadores que possuem os endereços em causa; utilizado em *use cases* nos *softwares* SIEM ou SOAR e ter uma visão sobre a evolução da ameaça ao longo do tempo nas redes da organização.

3.4. Pontuação de Confiança – *TrustRating*

Após calcular a pontuação da ameaça - *ThreatScore* - é importante determinar a pontuação de confiança do sistema de atribuição de *score*. A pontuação de confiança - *TrustRating* - é diretamente proporcional à informação proveniente das fontes externas, ou seja, quanto mais um endereço IPv4 está listado externamente, mais confiança existirá sobre o *ThreatScore* do mesmo. Este parâmetro é calculado através do somatório dos pesos atribuídos às fontes de informação externas. Consequentemente, estes valores estão em concordância com os pesos definidos anteriormente para o cálculo do *ThreatScore*, ou seja, as variáveis representadas pelo nome das plataformas na equação 3.5 continuam a assumir valores binários, entre 0 e o peso máximo atribuído a essa plataforma. Por exemplo, se o endereço IPv4 estiver listado na plataforma IBM X-Force Exchange, como o peso desta plataforma é de 0.2, a pontuação de confiança será incrementada em 20%.

A pontuação de confiança traduz-se na seguinte equação:

$$\textit{TrustRating}\ (\text{TR})\ \% = [\ (\text{Neutrino} + \text{GreyNoise} + \text{IBM} + \text{Pulsedive})] * 100 \qquad (3.5).$$

Onde,

Neutrino = {0, 0.4};

GreyNoise = {0, 0.3};

IBM = {0, 0.2};

Pulsedive = {0, 0.1}.

3.5. Atualização da Pontuação da Ameaça

Para tornar o sistema de atribuição de pontuação das ameaças eficiente, a pontuação não é computada constantemente e de forma repetida, isto é, se tiver sido calculada previamente, fica armazenada na instância do IoC e será reutilizada quando necessário. Quando ocorrer uma nova solicitação para analisar um observável já existente na plataforma OpenCTI, as plataformas externas

só serão consultadas para atualizar o *ThreatScore* e o *TrustRating* da ameaça, sempre que se verificar pelo menos uma das seguintes premissas:

- IoC com *Trust Rating* ≤ 50;
- *ThreatScore* calculado há 1 dia ou mais.

Caso contrário, o *ThreatScore* será atualizado tendo em conta os valores recolhidos na última análise, ou seja, o valor da componente - $\beta(Score\ Fontes\ Externas)$ - da equação 3.3 e, que é calculado através da equação 3.2, manter-se-á inalterado. Paralelamente, o *TrustRating* também não sofrerá alterações, uma vez que está exclusivamente relacionado com as plataformas externas. Concludentemente, só será recalculado o parâmetro $\alpha(Score\ Fonte\ Interna)$ da equação 3.3, através da equação 3.1 (Figura 3.7).

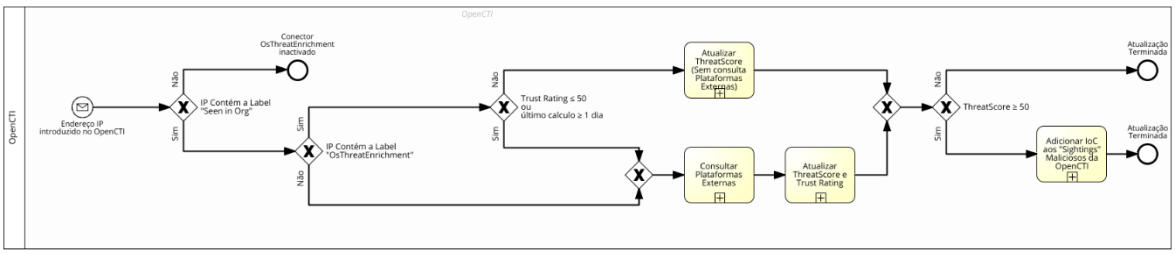

Figura 3.7 - Premissas para a atualização do ThreatScore. Fonte: elaboração do autor.

A racional de recalcular a pontuação da ameaça baseia-se no facto do IoC ser reincidente através de uma atividade suspeita que despoletou os mecanismos de alerta que monitorizam a rede da organização. Como tal, esta situação repercutirá efeitos na variável *History* que contribui para o *Threatscore* calculado a partir da plataforma OpenCTI, como demonstrado anteriormente. Logo, esta situação culminará com o cálculo do *ThreatScore* atualizado.

Para determinar o *ThreatScore* atualizado (*updatedThreatScore)* aplica-se o conceito de média ponderada, que possibilita agregar a nova pontuação com a calculada previamente. Este modelo de cálculo permite penalizar de forma mais célere um IoC que passe a desempenhar uma atividade maliciosa, como é o caso das ameaças do dia zero, e, ao mesmo tempo, retardar a despenalização de um IoC que já não apresente atividade suspeita. Desta forma, tomemos como exemplo um IoC que outrora não era suspeito e atualmente é considerado uma ameaça elevada, esta situação estará refletida no novo *ThreatScore,* uma vez que a pontuação da ameaça atual tem maior preponderância face ao *ThreatScore* anterior. Em contrapartida, um IoC que antes era suspeito e à data não apresenta atividade maliciosa, mantém-se sob suspeita, uma vez que o *ThreatScore* anterior é favorecido face ao atual. As premissas para adequar a equação à atualização do *ThreatScore* estão definidas no Quadro 3.6. De forma a facilitar a interpretação do Quadro 3.6, a terminologia "*previousThreatScore"* é referente à pontuação anterior que esse mesmo endereço IPv4 possuía antes do novo cálculo do *ThreatScore*.

Quadro 3.6 - Equações para atualizar o ThreatScore (UpdatedThreatScore). Fonte: elaboração do autor.

Condição	**Fórmula**
(*ThreatScore* > 50 || *ThreatScore* < *30*) && *previousThreatScore* > *ThreatScore*	$UpdatedTS$ = [0.7 * $prevTS$ + 0.3 * $ThreatScore$] (3.6)
(*ThreatScore* > 50 || *ThreatScore* < *30*) && *previousThreatScore* <= *ThreatScore*	$UpdatedTS$ = [0.3 * $prevTS$ + 0.7 * $ThreatScore$] (3.7)
30 ≤ *ThreatScore* ≤ 50 && 30 ≤ *previousThreatScore* ≤ 50	$UpdatedTS$ = [0.5 * $prevTS$ + 0.5 * $ThreatScore$] (3.8)

O parâmetro *TrustRating* também será recalculado para apresentar um *status* atualizado, caso tenha existido uma consulta das ferramentas externas.

CAPÍTULO 4

Resultados e Discussão

O mecanismo de reputação através de pontuações é aplicado ao nível da internet desde que começaram a existir compras online, ou seja, há mais de duas décadas. Numa fase posterior, este sistema foi estendido para a relação *peer-to-peer*, muito popular pelos *downloads* ilegais que este tipo de comunicação permitia e pelas ameaças que tinha subjacentes (*malware*), de forma a validar a confiança e qualidade de cada par [65]. Desde então, este mecanismo tem sido amplamente adotado em diversas áreas, o mesmo sucede na área de cibersegurança com a partilha de *threat intelligence*.

A necessidade de criar um mecanismo autónomo e fidedigno de atribuição de pontuação de risco aos indicadores de compromisso, urge devido aos milhares de eventos diários que um analista SOC tem de investigar. Desta forma, a automatização de processos de recolha de *threat intelligence* e a modelação da mesma pronta a ser consumida pela máquina e pelo humano é de enorme relevância.

No caso específico da plataforma OpenCTI, o conector que atribui pontuações aos IoC do tipo endereços IPv4 tem de ser do tipo "*internal-enrichment*", geralmente, é o AbuseIPDB, caso esteja implementado. Todavia, a pontuação atribuída por esta plataforma baseia-se nos reportes dos utilizadores, isto é, cada reporte tem um impacto baixo no *score* total de forma a ser incrementado progressivamente, o que faz com que este valor seja muito conservativo [63]. Como tal, a inexistência de um conector que atribuísse uma pontuação de ameaça aliada a uma pontuação de confiança, e que não dependesse apenas do reporte dos utilizadores de uma única plataforma, exacerbou a necessidade de implementar um conector que atendesse a esse objetivo e aos demais especificados pela organização em estudo.

4.1. Conector "*OsThreatEnrichment*"

O conector desenvolvido ao longo deste trabalho, designado de "*OsThreatEnrichment"*, foi implementado com sucesso e pertence à categoria "*internal-enrichment"* da plataforma OpenCTI. Atualmente, o *container* da plataforma está implementado no *docker* que se encontra em execução na *cloud* da AWS, sendo possível aceder através do endereço sob o domínio da OutSystems[13] conforme representado na Figura 4.1.

[13] https://openctidev.outsystemssec.com/dashboard

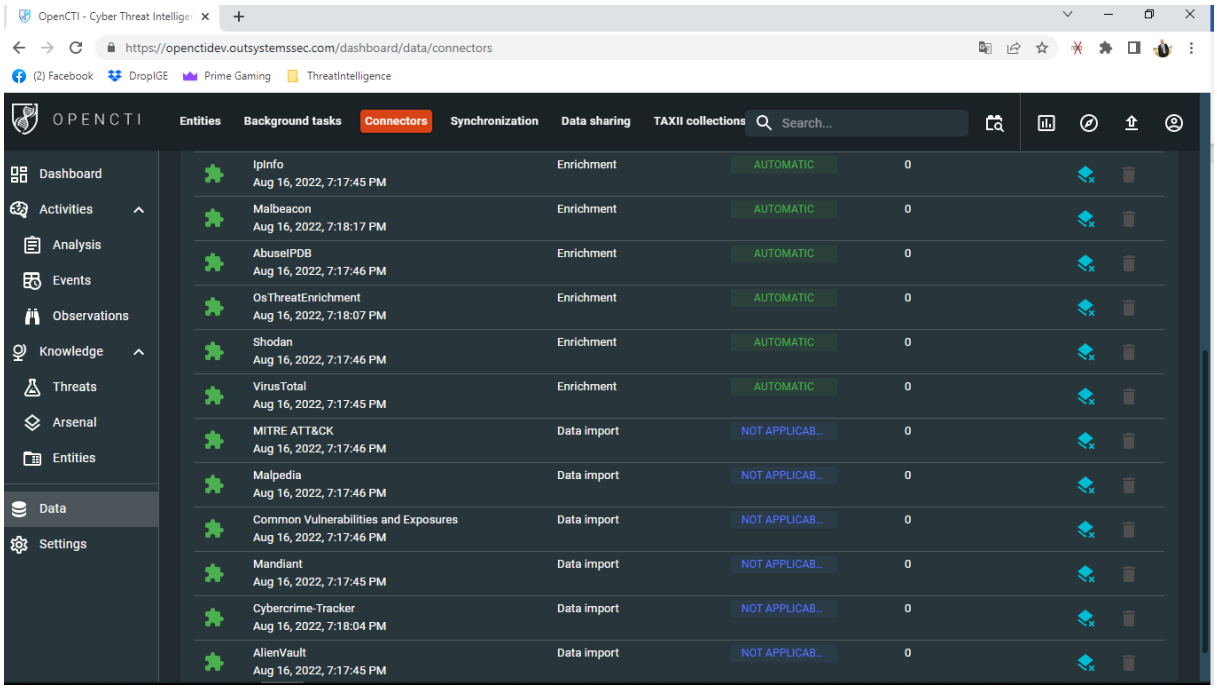

Figura 4.1- Imagem referente à plataforma OpenCTI sob o domínio da organização OutSystems. Fonte: elaboração do autor.

Conforme descrito no Capítulo 3 – Arquitetura de Integração e Metodologia - da presente dissertação, este conector permite obter informação de forma célere e fidedigna sobre o nível de ameaça de endereços IPv4 que sejam submetidos para análise pela organização. Portanto, sempre que existam suspeitas sobre um determinado IoC deste tipo ou sobre o qual sejam pretendidas mais informações, este é adicionado à plataforma OpenCTI com a *label "Seen in Org"* e, automaticamente, o conector é executado. Por sua vez, são consultadas plataformas externas de *threat intelligence* com o objetivo de calcular a pontuação da ameaça (*ThreatScore*). Adicionalmente, é calculado uma pontuação de confiança (*TrustRating*) que está relacionada com o grau de certeza da pontuação atribuída pelo algoritmo implementado.

O conector integra outras funcionalidades que se revelaram uma mais-valia de acordo com os analistas da organização OutSystems. Por exemplo, o automatismo de importação de *labels*, que permite direcionar investigação dos analistas para determinar o que motiva a presença daquele IoC na rede da organização, bem como as possíveis intenções do ator que possui aquele endereço IPv4. Além do mais, este tipo de dados são uma mais-valia para criar regras de inferência nos sistemas de defesa da rede ou para relacionar com outros tipos de informações de forma a retirar ilações [66], [67], resultando em medidas reativas e proativas mais adequadas perante a ameaça identificada [68], [69].

Este trabalho está intimamente correlacionado com outros dois projetos a serem desenvolvidos na organização. O primeiro corresponde a um *bot,* designado de "*SOARBOT",* responsável por adicionar à plataforma OpenCTI indicadores de compromisso detetados nos registos da rede da organização e recolher o *ThreatScore*, ou seja, baseia-se numa relação estática em que há apenas um *post* e *request*; o segundo corresponde a um algoritmo de deteção de ameaças que visa interligar os sistemas de defesa da organização com a plataforma OpenCTI. Contudo, distingue-se do primeiro por ser dinâmico, em que a interação ocorre proporcionalmente às deteções do IoC nos registos de rede, permitindo acompanhar a evolução da ameaça e contribuir para a atualização constante do *ThreatScore* através dos reportes de deteção do endereço em questão.

4.2. Estudo dos Dados Analisados pelo Conector

O conector ficou operativo a partir do dia 18 de julho de 2022, e desde então, tem estado a consumir dados fornecidos pelas ferramentas de defesa da organização. Como fora demonstrado na arquitetura do trabalho no Capítulo 3, o conector além de calcular o *ThreatScore* e *TrustRating,* possui outros automatismos:

- Recolha de dados para contextualizar a ameaça e averiguar a atividade dos endereços IPv4 submetidos para análise (Figura 4.2 e 4.3);
- Adiciona uma descrição personalizada consoante o IoC analisado, por exemplo em quais das ferramentas consultadas foi encontrado, bem como as hiperligações para consultar mais informações sobre o IoC nas mesmas (Figura 4.2). Além do mais, sempre que um determinado IoC seja analisado pela primeira vez, o campo "*ThreatScore previous"* é igualado a N/A ("*Not Available"*). Consequentemente, numa próxima análise solicitada para este endereço em específico, este campo será preenchido com o valor do *ThreatScore* que apresentava no momento do pedido e será calculado um novo *ThreatScore* que surge na descrição como "*ThreatScore New"* (Figura 4.2). Este mecanismo permite ter uma perceção sobre a evolução da ameaça;
- Efetua registos históricos no campo dos avistamentos ("*sightings"*), referente ao quantitativo de vezes que o índice de compromisso foi analisado e apresentou um *ThreatScore* superior a 50, ou seja, foi considerado uma ameaça maliciosa. Além disso é associado um "*Confidence Level"* qualitativo a esse avistamento, cujo valor é o calculado no *TrustRating*, contudo é convertido automaticamente pela plataforma para uma variável qualitativa. Caso seja comprovado que o IoC não era de cariz malicioso, a plataforma permite que o analista reporte esse avistamento como falso positivo (Figura 4.3);

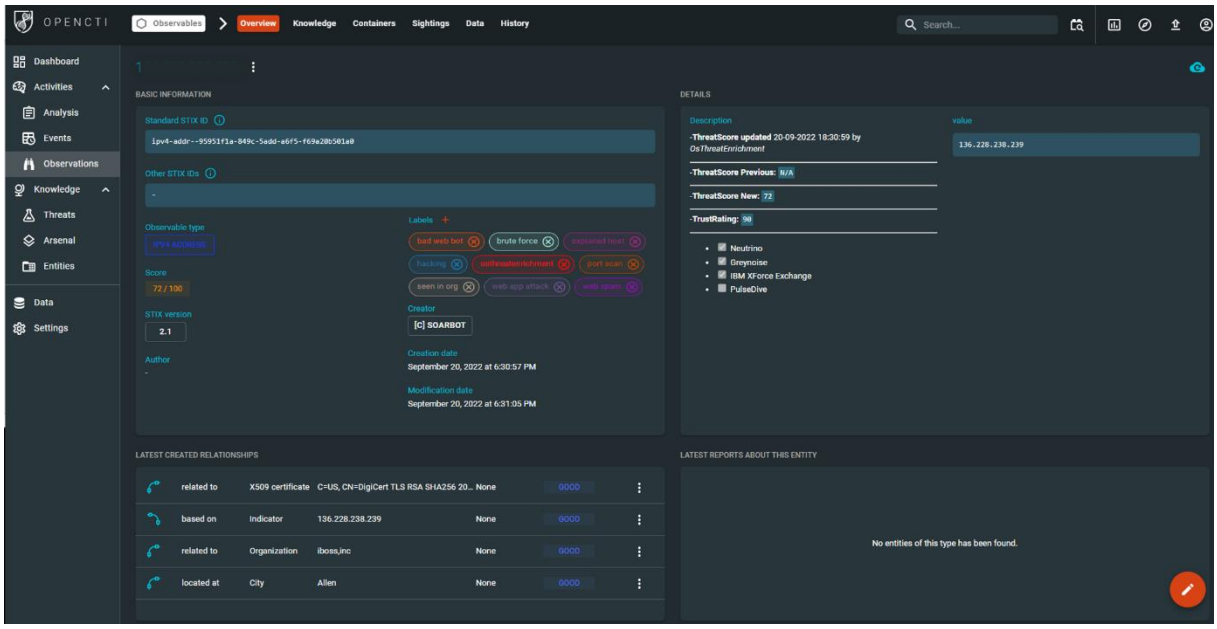

Figura 4.2 - Vista principal do índice de compromisso em análise. Fonte: elaboração do autor.

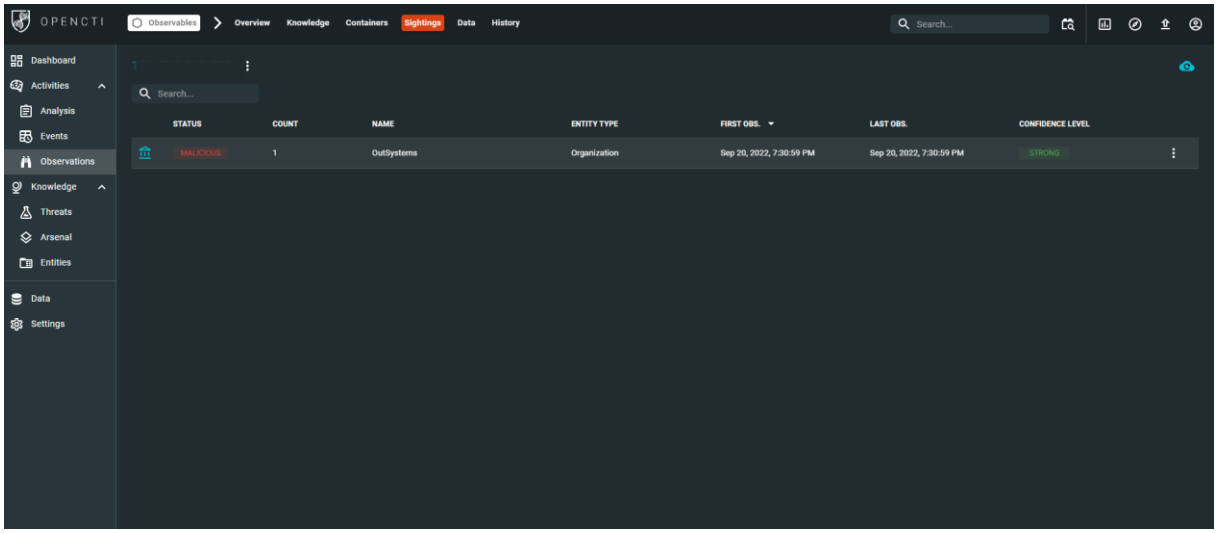

Figura 4.3 - Vista secundária referente aos avistamentos (Sightings) do índice de compromisso. Fonte: elaboração do autor.

Durante 90 dias foram analisados um total de 542 dados referentes a endereços IPv4 suspeitos (Figura 4.4). Dos eventos analisados, a média do *ThreatScore* atribuído é igual a 36 pontos, como é possível observar na Figura 4.5. Paralelamente, a média da pontuação de *TrustRating* é de 90%, o que significa que na maioria das vezes foi possível atribuir uma pontuação a partir das fontes externas consultadas, com exceção da plataforma Pulsedive que nem sempre possui informação sobre os endereços IPv4 analisados. Porém, esta situação já tinha sido avaliada na metodologia da presente dissertação durante a fase de validação das ferramentas, e manteve-se a decisão de manter a plataforma entre as eleitas, uma vez que a ferramenta disponibiliza dados relevantes quando se trata de endereços que estão associados a atividades maliciosas.

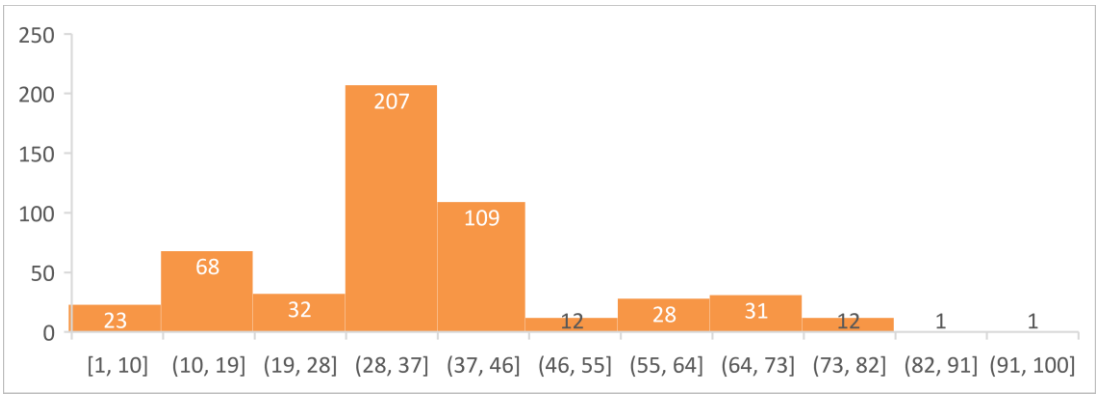

Figura 4.4 - Distribuição do ThreatScore atribuído aos 542 eventos analisados, dividido em intervalos de 9 unidades. Fonte: elaboração do autor.

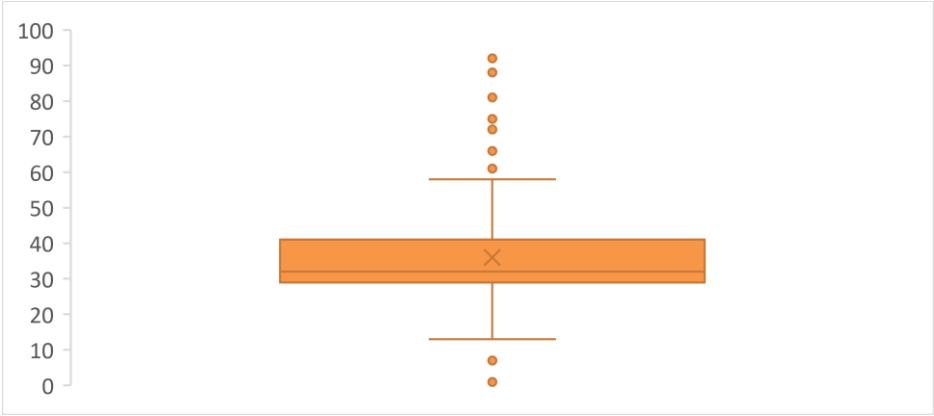

Figura 4.5 - Boxplot referente ao valor ThreatScore atribuído aos eventos analisados. Fonte: elaboração do autor.

Dos endereços IPv4 analisados, através das *labels* importadas das plataformas IBM X-Force Exchange e Neutrino, podemos apurar a que tipo de atividades estão associados. De acordo com representação gráfica da Figura 4.6, é possível constatar que 13% dos endereços analisados estão associados a atividades de "*hacking*". Posteriormente, com uma percentagem equivalente a 12%, surge a *label* "*web app attack*" associada a atividades criminosas. Em terceiro lugar, estão duas classificações a de "*brute force*" e a "*dynamic ips*", sendo que esta última é o termo utilizado para endereços IPv4 temporários atribuído a dispositivos conectados à internet. Estes endereços temporários podem originar dúvidas, uma vez que a maioria dos utilizadores domésticos são endereços IPv4 dinâmicos que são atribuídos pelos provedores do serviço de internet (*Internet Service Provider* – ISP), consequentemente os mesmos não estão exclusivamente associados atores maliciosos [70]. Portanto, as medidas de prevenção poderão passar pela deteção precisa destes blocos de endereços IPv4 dinâmicos e correlacionar com outros IoC.

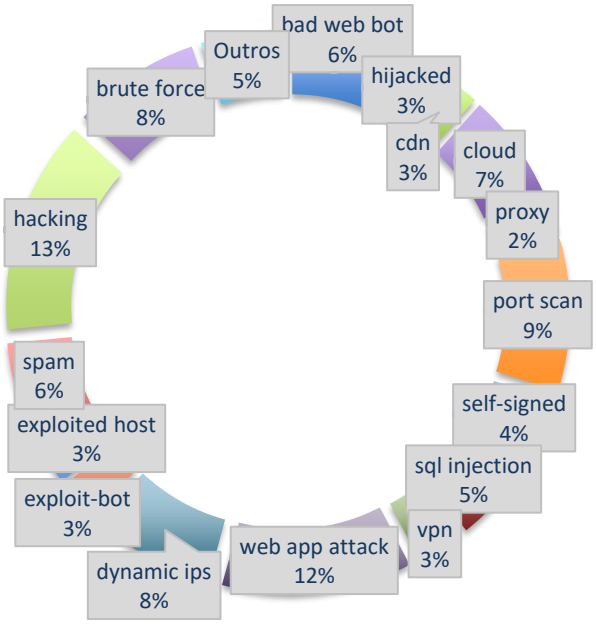

Figura 4.6 - Labels associadas aos endereços IPv4 analisados. Fonte: elaboração do autor.

Através da análise da Figura 4.7, é possível verificar a origem geográfica dos endereços que foram considerados suspeitos pelos sistemas de defesa da organização e, por sua vez, resultaram em eventos que foram submetidos na plataforma OpenCTI. A origem geográfica destes endereços é maioritariamente dos Estados Unidos da América e da Índia, com 36% e 12% dos eventos, respetivamente. Estes resultados estão em concordância com os dados recolhidos durante a seleção das ferramentas de *threat intelligence*, obtidos a partir das plataformas AbuseIPDB e VirusTotal, em que 36% dos novos endereços reportados como maliciosos eram originários dos EUA.

Todavia, ao analisar a Figura 4.8 que representa o *ThreatScore* médio atribuído aos endereços IPv4 analisados e distribuídos por país, as maiores ameaças não são provenientes dos que apresentam maior número de endereços IPv4 suspeitos. Como tal, os eventos com *ThreatScore* médio mais elevado são provenientes de endereços pertencentes a países do continente asiático, assim como Singapura e Filipinas. No continente europeu, destacam-se países do leste europeu, como a Ucrânia, Polónia e Chéquia. O Luxemburgo é uma exceção, pois tratou-se de um único evento analisado, cujo valor de *ThreatScore* atribuído foi de 62 pontos.

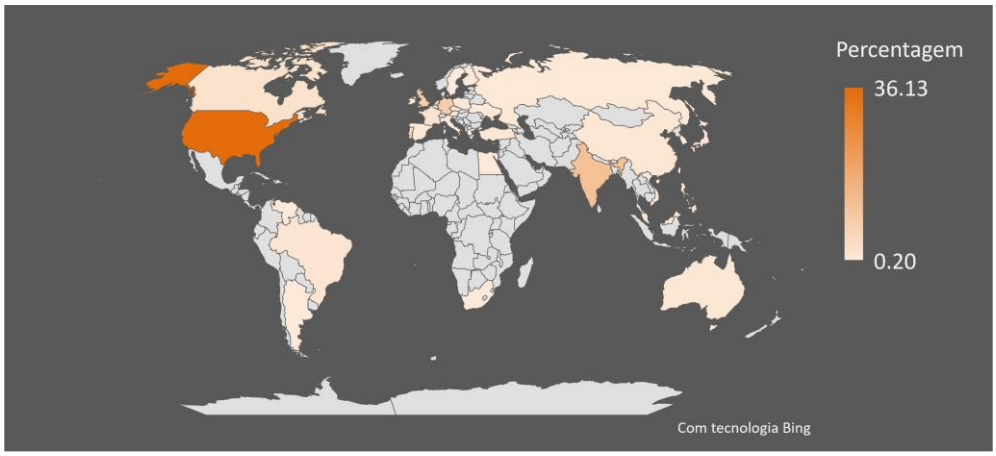

Figura 4.7 - Distribuição geográfica dos eventos em percentagem, considerado o total da amostra de 542 endereços IPv4.
Fonte: elaboração do autor.

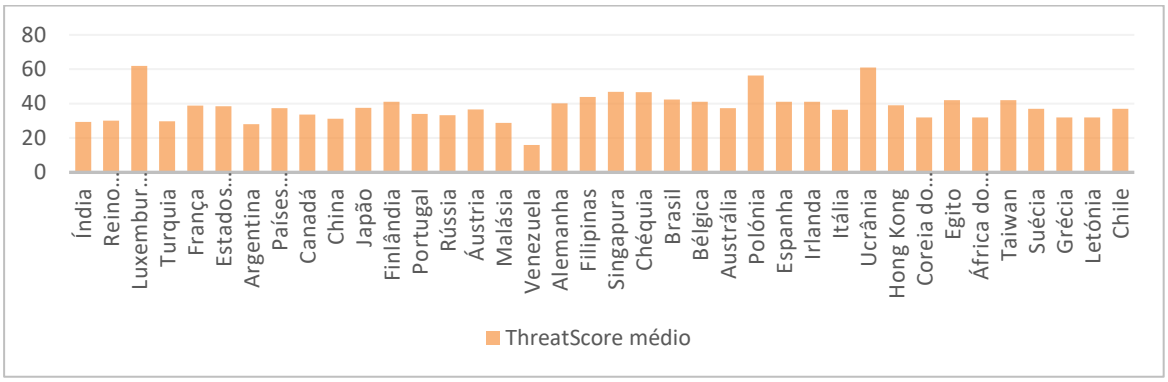

Figura 4.8 - Média do ThreatScore atribuído aos eventos de cada país. Fonte: elaboração do autor.

4.3. Validação do Modelo

Para validar o modelo de reputação desenvolvido, não se recorreu às *blacklists* anteriores utilizadas para construir o mesmo, de forma a não existir enviesamento dos resultados. Cmo tal, a organização - OutSystems - disponibilizou duas listas de endereços IPv4 classificados como benignos e malignos, provenientes de uma *whitelist* e de uma *blacklist*, respetivamente. Estes endereços foram exportados para a plataforma OpenCTI e avaliados pelo conector implementado, dando origem a uma matriz de confusão que se encontra representada no Quadro 4.1. Para construir a presente matriz, foram considerados endereços IPv4 malignos todos os que obtiveram um *ThreatScore* superior a 50, os demais foram classificados como benignos.

Quadro 4.1 Matriz de Confusão com os endereços IPv4 analisados. Fonte: elaboração do autor.

		Valores Preditos		
		Maligno	**Benigno**	
Valores Reais	**Maligno**	44 Verdadeiros Positivos	0 Falsos Positivos	44
	Benigno	8 Falsos Negativos	50 Verdadeiros Negativos	58
		52	50	

No total foram analisados 102 endereços IPv4, dos quais 50 endereços foram reportados corretamente como benignos, ou seja, todos os que constituíam a *whitelist*. Quanto aos malignos, foram classificados 44 dos 52 endereços corretamente, com um *ThreatScore* médio de 64 pontos. Consequentemente, ocorreram 8 falsos negativos, o que se traduz num valor de *accuracy* de 92%. Outro parâmetro passível de análise é a taxa de verdadeiros positivos, ou seja, quão bom é o modelo a prever endereços IPv4 malignos, neste caso o resultado é de 85%. Por último, a precisão do modelo é de 100%, ou seja, todos os endereços IPv4 malignos foram identificados corretamente, o que significa que não houve nenhum falso positivo. O facto de não ter sido reportado qualquer falso positivo, muito provavelmente, deve-se à escolha dos endereços IPv4 para afinar o modelo durante a metodologia deste trabalho. Recorde-se que durante a filtragem dos endereços na plataforma AbuseIPDB, optou-se pela seleção dos quais existia certeza sobre as suas intenções maliciosas, para evitar a existência de falsos positivos na amostra.

Estes resultados estão ao nível de outros estudos de reputação revistos na literatura, em que o maior problema é identificar ameaças do dia zero (*zero-day)*, como se pode comprovar pelos falsos negativos. Quanto à avaliação do modelo através da matriz de confusão, e*xistem* estudos que recorrem a diferentes metodologias para desenvolver sistemas de reputação de IoC, cujos resultados apresentados são inferiores[12], outros idênticos [71], [72]e alguns que chegam perto de 100% de *accuracy* [11].

No geral, os modelos de reputação têm como objetivo avaliar determinada característica de forma quantitativa ou qualitativa. Neste caso específico, trata-se da reputação dos endereços IPv4 a partir de diferentes fontes de informação, ou seja, a reputação é usada para representar a opinião que uma entidade recebe das restantes entidades na rede de forma quantitativa (*ThreatScore)* [14] . Portanto, para suportar a validação do presente modelo de reputação, é legítimo recorrer a um dos estudos pioneiros nesta matéria, como é o caso do trabalho desenvolvido por Dingledine *et al*. (2000) [73], no qual propõe um conjunto de critérios básicos para avaliar a qualidade e solidez dos sistemas de reputação [14], [65], [73]. Tais princípios foram inicialmente definidos para classificar as ligações *peer-to-peer* (P2P) e, posteriormente, aplicações comerciais existentes na internet, bem como redes móveis *ad hoc*. Sendo assim, os critérios de maior relevância aplicáveis na validação dos sistemas de reputação à base de *threat intelligence*, são os seguintes:

a) Preciso para um desempenho a longo prazo – O sistema deve refletir a confiança (a probabilidade de precisão) de uma determinada pontuação, e ao mesmo tempo deve distinguir entre uma nova entidade de qualidade desconhecida e uma entidade com má performance de longo termo;
b) Ponderado tendo em consideração o desempenho atual – O sistema reconhece e reflete tendências recentes no desempenho da entidade. Por exemplo, uma entidade que teve um bom comportamento por muito tempo, mas repentinamente adotou maus comportamentos, deve ser reconhecida como tal e não mais confiável;
c) Eficiente – o sistema deve conseguir recalcular uma pontuação rapidamente, além do mais é importante que os cálculos possam ser realizados de forma incremental;
d) Verificável – não deve haver pontuações dúbias, pois estas devem ser suportadas por dados;

Os requisitos a), b) e d) são cumpridos através da consulta dinâmica em tempo real de diferentes fontes de *threat intelligence*, bem como através do valor calculado com base no registo histórico do IoC analisado (*ThreatScore)* e a confiança sobre este (*TrustRating)*. Quanto ao requisito descrito na alínea c), conforme reportado em diversos estudos, é despendido muito tempo na análise de pacotes associados a endereços IPv4, como tal a ameaça calculada pelo conector permite reduzir esse custo ao evitar esse tipo de inspeções. Como resultado, possibilita uma melhoria de desempenho substancial, bem como permite penalizar apenas conexões com um elevado grau de ameaça, ao invés de todos os endereços suspeitos.

A plataforma escolhida para incorporar na infraestrutura da organização, a OpenCTI, bem como o conector "OsThreatEnrichment", permitem ir ao encontro dos três princípios referentes à *threat intelligence* publicados por Henry Dalziel (2014) [74]: (i) deve ser relevante para a entidade que a recebe, (ii) acionável, (iii) e de valor na perspetiva da organização. Os pontos (i) e (iii) são cumpridos a partir do momento em que os indicadores de compromisso resultam em eventos suspeitos submetidos para análise, como tal toda a informação obtida relativamente a essas entidades, será útil na tomada de decisão de os descartar ou, em última instância, resultar num incidente [66]. Quanto ao ponto (ii), este também é cumprido, uma vez que vai ao encontro dos critérios definidos pela Agência da União Europeia para Cibersegurança (ENISA). De acordo com esta entidade, ser acionável requer o cumprimento dos seguintes parâmetros [75]:

- Relevante – o IoC poderá ter impacto ou comprometer a infraestrutura da organização;
- Oportunidade temporal – *threat intelligence* sobre IoC é relevante quando partilhada e/ou obtida em tempo real, isto é, a *threat intelligence* referente a atividade maliciosa recente e partilhada num curto espaço de tempo, tem maior utilidade do que a *threat intelligence* descoberta e publicada alguns meses após o início da atividade maliciosa;
- Precisão - o recetor da informação deve ter capacidade de processar a mesma num curto espaço de tempo;
- Variedade – Essencial usar várias ferramentas e fontes para detetar o IoC;
- Digestibilidade – A informação sobre o IoC deverá ser de fácil análise e processamento nas diferentes ferramentas importantes para a defesa e partilha de *threat intelligence*;
- Plenitude – A informação sobre a ameaça ou IoC deverá ser o mais completa possível. No entanto, mesmo que incompleta, deverá ter utilidade no seio da organização recetora de forma a ser acionável em complemento com outra informação interna ou externa.

4.4. Limitações do Modelo

O modelo desenvolvido apresenta algumas limitações, entre as quais destacam-se:

- A *framework* apresentada foi otimizada para indicadores de compromisso do tipo endereço IPv4. Como tal, para estender a outra tipologia terão de ser validadas novamente algumas destas ferramentas que permitem a análise de outros tipos de IoC, como é o caso da plataforma IBM X-Force Exchange, Pulsedive e Neutrino, assim como avaliar plataformas adicionais para aprimorar o modelo;

- A técnica desenvolvida permite um cálculo rápido e eficiente do *ThreatScore*, ou seja, é possível obter um resultado em menos de 6 segundos com um baixo consumo de recursos da máquina, o que corrobora a sua eficiência. Além do mais, a solução apresentada não tem custos associados, como tal apresenta limitações quanto ao número de solicitações mensais, estando esta situação relacionada com a utilização dos planos gratuitos disponibilizados pelas fontes consultadas. Até à data esta não foi uma limitação para a OutSystems. No entanto, poderá tornar-se em organizações que sejam mais suscetíveis ao interesse de atores mal-intencionados ou que não tenham sistemas de defesa eficazes na triagem dos indicadores de compromisso, resultando num número elevado de solicitações. Caso se verifique este cenário, poderá ser ultrapassado investindo em planos pagos, sem a necessidade de alterar a *framework* desenvolvida;
- Este conector é despoletado automaticamente na plataforma OpenCTI sempre que é criado um novo observável com a *label "seen in org"*. Todavia, se o IoC já existe na plataforma OpenCTI e é introduzido como novo observável, de forma a impedir a existência de duplicados, a OpenCTI não adiciona essa nova instância. Consequentemente, é crucial que o mecanismo que interliga os sistemas de defesa da organização com o presente conector seja capaz de efetuar novas solicitações referentes a um IoC que fora detetado novamente nos eventos suspeitos. Esta lacuna será corrigida brevemente, através de um automatismo que está a ser desenvolvido num trabalho de mestrado paralelo a este, como fora descrito anteriormente;
- Existem abordagens que recorrem a técnicas baseadas em *machine learning*, como está ilustrado no Anexo B. Estas técnicas podem ser uma mais-valia na previsão de ameaças e na tomada de ação para mitigá-las, como tal devem ser encaradas como um complemento ao presente trabalho. No entanto, apesar das vantagens inerentes a essas metodologias, continua a haver muitas limitações nestes sistemas de reputação, sobretudo ao nível das vulnerabilidades associadas ao dia zero*,* uma vez que estes tipos de abordagens não categorizam com precisão os ataques do dia zero [9];
- A *accuracy* do sistema de reputação apresenta limitações relativamente aos falsos negativos. Esta limitação está intimamente relacionada com as ameaças do dia zero, um problema comum a trabalhos que têm como objeto de estudo a reputação de indicadores de compromisso, como fora referido durante a revisão bibliográfica. Este parâmetro poderá ser melhorado se conjugado com outros mecanismos, tais como a inteligência artificial. Consequentemente, este tipo de complementaridade terá impacto no desempenho do modelo, uma vez que poderá requerer mais tempo e capacidade de processamento para analisar os indicadores do compromisso.

CAPÍTULO 5

Conclusões e Trabalhos Futuros

5.1. Conclusões

O presente trabalho foi realizado com o objetivo de centralizar numa única plataforma a aquisição, produção e partilha de *threat intelligence*, de forma a permitir a avaliação das ameaças detetadas pelos sistemas de defesa da organização - OutSystems - de uma forma célere e precisa. Como tal, é importante clarificar que a *framework* desenvolvida é de prevenção de ameaças e não de proteção direta contra as mesmas, ou seja, é um mecanismo complementar às defesas da organização que permite fornecer informações credíveis para a tomada de decisão.

A OpenCTI foi a plataforma designada como ferramenta central para atingir os objetivos definidos em colaboração com a organização em estudo. De forma a tirar o máximo partido desta plataforma, foi desenvolvido um conector que permite recolher dados de ferramentas externas e, consequentemente, através de um algoritmo quantificar o nível de ameaça (*ThreatScore*) do indicador de compromisso em análise, bem como o nível de confiança (*TrustRating*) da pontuação atribuída. Esta metodologia foi aplicada apenas a indicadores de compromisso do tipo endereço IPv4.

Como produto deste trabalho é possível retirar diversas ilações, entre as quais, destaca-se o poder de gestão de informação sobre ameaças que a plataforma OpenCTI possui através da importação de novos indicadores de compromisso e enriquecimento de dados; a compatibilidade para integrar novos conectores desenvolvidos de acordo com as estratégias e pretensões das organizações, bem como a capacidade de potenciar as ferramentas utilizadas na cibersegurança através da integração e criação de automatismos com base nestes conectores. Quanto aos resultados obtidos através do conector "OsThreatEnrichment", destacamos as seguintes funcionalidades e vantagens comparativamente a outros conectores existentes para a plataforma:

- Evolução e Registo histórico dos indicadores de compromisso detetados na rede da organização, em virtude da natureza volátil dos IoC ao longo do tempo [11], [76]. Esta é uma vantagem comparativamente a outros conectores da plataforma em questão, mas também a alguns modelos de reputação que não contemplam a evolução histórica dos IoC internamente nas redes da organização, baseando-se apenas em informações externas;
- Agregação e correlação de dados de *threat intelligence* provenientes de diversas plataformas externas com o intuito de calcular uma pontuação da ameaça. Enquanto que os demais baseiam-se apenas nas suas fontes;

- Validação da fórmula de cálculo *ThreatScore* e *TrustRating* de forma a ser o mais precisa possível;
- O algoritmo apresentado é dinâmico, na medida em que reavalia a ameaça tendo em conta o histórico do IoC na rede interna e em fontes externas;
- *TrustRating* associado à pontuação de ameaça, o que permite aos analistas ter uma confiança adicional na estrutura de reputação desenvolvida;
- Enriquecimento dos indicadores de compromisso suspeitos submetidos a análise com dados sobre o tipo de atividade maliciosa a que estão associados, neste caso as *labels*, o que permitirá ao analista priorizar os eventos suspeitos e direcionar a sua investigação;
- Integração de relatórios provenientes de plataformas externas com informação adicional;
- Compatibilidade e integração com automatismos utilizados na resolução de eventos detetados no sistema de defesa da organização. Esta funcionalidade permite minimizar o tempo despendido pelo analista na procura de informação sobre o IoC que gerou o evento e proporciona uma tomada de decisão rápida e segura. Recorde-se que esta é uma das principais causas de fadiga dos analistas;
- Possibilidade de partilha em tempo real de indicadores de compromisso detetados na rede da organização, bem como o nível de ameaça associado.

Apesar das vantagens inumeradas anteriormente, este sistema, tal como outros sistemas de reputação, apresenta algumas vulnerabilidades:

- O algoritmo apresentado é válido apenas para endereços IPv4, como tal para outros tipos de IoC poderá não ser a solução mais adequada. No entanto, a metodologia apresentada é apropriada para selecionar as plataformas a utilizar para recolha de *threat intelligence*, sendo necessário redefinir a fórmula de cálculo e os pesos de cada variável;
- Como apresentado na revisão bibliográfica, o endereço IPv4 é dos indicadores mais fáceis de adulterar por parte dos atacantes. Todavia, é dos que permite tomar ações mais rápidas para bloquear ameaças;

- O sistema de reputação é vulnerável a ameaças do dia zero, ou seja, em caso de inexistência de informação sobre o endereço em análise, a pontuação da ameaça será baixa e resultará num falso negativo. De forma a minimizar esta fragilidade, as organizações deverão investir em *softwares* de defesa eficientes, bem como em analistas treinados que possam reverter estes falsos negativos. Esta situação enaltece a importância de o algoritmo desenvolvido ter em conta o registo histórico interno dos indicadores de compromisso, pois é uma forma mais célere de detetar este tipo de ameaças e aumentar o *ThreatScore* das mesmas.

Ao longo deste trabalho foram dadas respostas às questões de investigação formuladas no Capítulo 1 – Introdução - da presente dissertação. No entanto, essas respostas serão sumariadas nos seguintes tópicos:

- **RQ1 (problema)**: Quais os problemas associados à gestão e integração de *threat intelligence* nas organizações? Os problemas estão relacionados com as políticas e estratégias de negócio que as organizações adotam, uma vez que a maioria investe pouco na cibersegurança e, quando o faz, investe essencialmente em ferramentas de segurança. O presente estudo demonstra que é necessário apostar na partilha de informações sobre ameaças em tempo real, de forma a minimizar os danos causados por estas. Quanto mais cedo a partilha, mais útil e eficiente será essa informação.
- **RQ2.1 (solução)**: Quais as potenciais soluções para a gestão e integração de *threat intelligence* nas organizações? As soluções passam por implementar uma ferramenta que possibilite centralizar e correlacionar as informações provenientes de fontes internas da própria organização com as fontes externas, ou seja, que tenha capacidade de consumir e produzir *threat intelligence*, como é o caso da OpenCTI. Além do mais, a ferramenta deverá ser versátil ao permitir a integração com as diversas ferramentas existentes no SOC.
- **RQ2.2 (solução):** De que forma estas soluções podem ser implementadas? Para implementar uma solução deste tipo, primeiramente é necessário ter claro o tipo de negócio da organização e recolher informações sobre as ameaças existentes na área em questão, bem como as vulnerabilidades associadas. Subsequentemente, é necessário avaliar a capacidade de defesa da organização contra ataques de atores maliciosos. Por último, após a identificação destas características, devem ser priorizados os indicadores de compromisso a monitorizar de acordo com o interesse para a organização e, consequentemente, adotar políticas proativas com base na *threat intelligence* produzida a partir de fontes internas e externas.

- **RQ2.3 (solução):** De que forma a(s) solução desenvolvida(s) ajudam a proteger a organização? A solução apresentada possibilita que a organização resolva de forma célere eventos suspeitos despoletados pelos mecanismos de segurança, através da recolha de *threat intelligence* e avaliação do risco. Adicionalmente, não carece da interferência do analista durante o processo de investigação, permitindo que este direcione a sua atenção para os eventos de maior gravidade.
- **RQ3.1 (avaliação):** Até que ponto os problemas típicos associados à gestão e integração de *threat intelligence* nas organizações podem ser resolvidos? Antes de ser adotada qualquer tipo de solução, as organizações devem de ter uma visão clara sobre os centros de operações de segurança e as necessidades existentes na constituição destes, uma vez que não existe um padrão quanto à configuração destes centros. De uma forma resumida, existem as ferramentas de defesa de primeira linha e as ferramentas auxiliares, como é o caso das plataformas de *threat intelligence.* Um mecanismo de defesa eficiente baseia-se na harmonia existente entre estes diferentes tipos de ferramentas. Como tal, ao longo deste trabalho, foi demonstrado que a solução implementada permite resolver os problemas de gestão e integração de *threat intelligence*, através da utilização de informação pertinente para efetuar a avaliação do risco dos eventos suspeitos detetados e, por conseguinte, disponibilizar essa informação às ferramentas de defesa ou ao analista para resolvê-los. Adicionalmente, a matriz de confusão permitiu avaliar esta solução e sustentar a afirmação anterior quanto à eficiência da solução apresentada.

5.2. Trabalhos Futuros

Em suma, a área da cibersegurança é altamente dinâmica em termos de ameaças e de novas ferramentas disponíveis que auxiliam na proteção das organizações. Como fora referido anteriormente, o problema não está nas ferramentas em si, mas na extração de informação útil, uma vez que diversidade não é sinónimo de qualidade. Portanto, esta ferramenta, como muitas outras, requer adaptação constante a este tipo de ameaças voláteis e é essencial desenvolver/aprimorar os seguintes objetivos em trabalhos futuros:

- Recolher dados de tentativas de ataques à organização e avaliar a pontuação de ameaça atribuídas aos IoC extraídos desses incidentes. Consequentemente, avaliar o método de reputação e, se necessário, ajustar os parâmetros que constituem o modelo de forma a melhorar o modelo de reputação;

- Analisar novas listas de endereços IPv4 malignos e benignos, calcular a matriz de confusão de forma a reavaliar o desempenho do algoritmo e, se necessário, ajustar os valores dos parâmetros que incorporam o mesmo;
- Conectar esta ferramenta a outros sistemas de defesa que avaliem as ameaças de forma quantitativa ou qualitativa, e introduzir essa variável no algoritmo apresentado com o objetivo de aumentar a precisão;
- Identificação de padrões ou características comuns aos indicadores de compromisso associados a atividades malignas, que permitam desenvolver automatismos à base de regras de inferência ou *playbooks*;
- Aplicar técnicas de *machine learning* para aumentar a precisão da avaliação do risco, tendo como base dados claros e concisos que permitam instruir o sistema. Por exemplo, no caso dos endereços IPv4, os endereços com má reputação habitualmente partilham atributos similares, algumas características selecionadas para estes sistemas são [12]: *Autonomous System Number* (ASN*); Internet Service Provider* (ISP); país de registo; Tipo de utilizador (comercial ou privado); região onde o endereço está localizado; tipo de *browser* (*Tor* ou outro);
- Estender este sistema de reputação aos restantes indicadores de compromisso. Para tal, a metodologia a utilizar poderá ser a mesma, com a particularidade que terão de ser avaliadas as plataformas mais fidedignas para recolher *threat intelligence* referente ao tipo de IoC que se pretende avaliar e, posteriormente, ajustar os pesos das ferramentas no algoritmo proposto;
- Partilhar a *threat intelligence* produzida pela organização, a partir desta *framework*, com organizações externas em tempo real. Como fora referido, a oportunidade temporal é terminante para as organizações adotarem medidas proativas, ao invés de medidas reativas que retardam a mitigação dos danos causados pelas ameaças. Além do mais, esta é uma valência da plataforma OpenCTI, a partilha de informação com organizações externas através de uma linguagem padrão, neste caso a STIX, que está pronta a ser "consumida" por outras ferramentas e ser convertida em *threat intelligence* operacional.

Referências Bibliográficas

[1] J. Jang-Jaccard e S. Nepal, «A survey of emerging threats in cybersecurity», em *Journal of Computer and System Sciences*, 2014, vol. 80, n. 5, pp. 973–993. doi: 10.1016/j.jcss.2014.02.005.

[2] Y. Li e Q. Liu, «A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments», *Energy Reports*, vol. 7, pp. 8176–8186, Nov. 2021, doi: 10.1016/j.egyr.2021.08.126.

[3] V. Adewopo, B. Gonen, e F. Adewopo, «Exploring Open Source Information for Cyber Threat Intelligence», em *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, Dez. 2020, pp. 2232–2241. doi: 10.1109/BigData50022.2020.9378220.

[4] IBM, «Cost of a data breach 2022 A million-dollar race to detect and respond», *Cost of a Data Breach Report 2022*, 2022. https://www.ibm.com/security/data-breach (acedido Jan.10, 2022).

[5] Centro Nacional de Cibersegurança, «Riscos & Conflitos 2021 », *Relatório Cibersegurança em Portugal*, 2021. https://cibercrime.ministeriopublico.pt/sites/default/files/documentos/pdf/

relatorio_riscos_conflitos2021.pdf (acedido Jan. 12, 2022).

[6] D. Schlette, M. Vielberth, e G. Pernul, «CTI-SOC2M2 – The quest for mature, intelligence-driven security operations and incident response capabilities: CTI-driven SOC capability maturity model», *Comput Secur*, vol. 111, Dez. 2021, doi: 10.1016/j.cose.2021.102482.

[7] B. Biswas, A. Mukhopadhyay, S. Bhattacharjee, A. Kumar, e D. Delen, «A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums», *Decis Support Syst*, vol. 152, Jan. 2022, doi: 10.1016/j.dss.2021.113651.

[8] H. Hettema, «Rationality constraints in cyber defense: Incident handling, attribution and cyber threat intelligence», *Comput Secur*, vol. 109, Out. 2021, doi: 10.1016/j.cose.2021.102396.

[9] N. Usman *et al.*, «Intelligent Dynamic Malware Detection using Machine Learning in IP Reputation for Forensics Data Analytics», *Future Generation Computer Systems*, vol. 118, pp. 124–141, Mai. 2021, doi: 10.1016/j.future.2021.01.004.

[10] G. Sakellariou, P. Fouliras, I. Mavridis, e P. Sarigiannidis, «A Reference Model for Cyber Threat Intelligence (CTI) Systems», *Electronics (Switzerland)*, vol. 11, n. 9, Mai. 2022, doi: 10.3390/electronics11091401.

[11] D. Preuveneers e W. Joosen, «Sharing Machine Learning Models as Indicators of Compromise for Cyber Threat Intelligence», *Journal of Cybersecurity and Privacy*, vol. 1, n. 1, pp. 140–163, Fev. 2021, doi: 10.3390/jcp1010008.

[12] Arya Renjan, Karuna Pande Joshi, Sandeep Nair Narayanan, e Anupam Joshi, *DAbR: Dynamic Attribute-based Reputation scoring for Malicious IP Address Detection*, IEEE ISI2018. IEEE International Conference on Intelligence and Security Informatics , 2018. doi: 978-1-5386-7848-0/18/$31.00.

[13] R. Sharifnya e M. Abadi, «A novel reputation system to detect DGA-based botnets», em *Proceedings of the 3rd International Conference on Computer and Knowledge Engineering, ICCKE 2013*, Set. 2013, pp. 417–423. doi: 10.1109/ICCKE.2013.6682860.

[14] Z. Zhang, Y. Kadobayashi, e F. Naït-Abdesselam, «Toward an evaluation framework for reputation systems in autonomic computing networks», em *2009 4th International Conference on Communications and Networking in China, CHINACOM 2009*, 2009, pp. 785–789. doi: 10.1109/CHINACOM.2009.5339707.

[15] R. Vinayakumar, K. P. Soman, e P. Poornachandran, «Evaluating deep learning approaches to characterize and classify malicious URL's», em *Journal of Intelligent and Fuzzy Systems*, 2018, vol. 34, n. 3, pp. 1333–1343. doi: 10.3233/JIFS-169429.

[16] A. de Melo e Silva, J. J. C. Gondim, R. de Oliveira Albuquerque, e L. J. G. Villalba, «A methodology to evaluate standards and platforms within cyber threat intelligence», *Future Internet*, vol. 12, n. 6, Jun. 2020, doi: 10.3390/fi12060108.

[17] OutSystems, «Informação sobre OutSystems», *OutSystems Website*. https://www.outsystems.com/company/ (acedido Nov. 10, 2021).

[18] A. Dennis, R. Jones, D. Kildare, e C. Barclay, «Design science approach to developing and evaluating a national cybersecurity framework for Jamaica», *Electronic Journal of Information Systems in Developing Countries*, vol. 62, n. 1, pp. 1–18, Mar. 2014, doi: 10.1002/j.1681-4835.2014.tb00444.x.

[19] A. Hevner e J. Park, «Design Science in Information Systems Research», 2004. Available: https://www.researchgate.net/publication/201168946

[20] Avast, «What is the History and Future of Network Security?», *History and future of network security*. https://www.avast.com/business/resources/future-of-network-security#pc (acedido Jan. 12, 2022).

[21] S. Panda, «Usefulness and Impact of Big Data in Libraries: An Opportunity to Implement Embedded Librarianship», 2021, pp. 45–60. doi: 10.31235/osf.io/fsau9.

[22] M. Vielberth, M. Glas, M. Dietz, S. Karagiannis, E. Magkos, e G. Pernul, «A Digital Twin-Based Cyber Range for SOC Analysts», em *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021, vol. 12840 LNCS, pp. 293–311. doi: 10.1007/978-3-030-81242-3_17.

[23] U. Noor, Z. Anwar, J. Altmann, e Z. Rashid, «Customer-oriented ranking of cyber threat intelligence service providers», *Electron Commer Res Appl*, vol. 41, Mai. 2020, doi: 10.1016/j.elerap.2020.100976.

[24] T. R. Reshmi, «Information security breaches due to ransomware attacks - a systematic literature review», *International Journal of Information Management Data Insights*, vol. 1, n. 2, p. 100013, Nov. 2021, doi: 10.1016/j.jjimei.2021.100013.

[25] S. Kramer e J. C. Bradfield, «A general definition of malware», *Journal in Computer Virology*, vol. 6, n. 2, pp. 105–114, 2010, doi: 10.1007/s11416-009-0137-1.

[26] J. M. Borky e T. H. Bradley, «Protecting Information with Cybersecurity», em *Effective Model-Based Systems Engineering*, Springer International Publishing, 2019, pp. 345–404. doi: 10.1007/978-3-319-95669-5_10.

[27] F. Thomaz, C. Salge, E. Karahanna, e J. Hulland, «Learning from the Dark Web: leveraging conversational agents in the era of hyper-privacy to enhance marketing», *J Acad Mark Sci*, vol. 48, n. 1, pp. 43–63, Jan. 2020, doi: 10.1007/s11747-019-00704-3.

[28] Simson. Garfinkel, Gene. Spafford, e Simson. Garfinkel, *Practical UNIX and Internet security*. O'Reilly & Associates, 1996.

[29] N. Kharlamova, S. Hashemi, e C. Træholt, «Data-driven approaches for cyber defense of battery energy storage systems», *Energy and AI*, vol. 5, Set. 2021, doi: 10.1016/j.egyai.2021.100095.

[30] M. A. Rajab, J. Zarfoss, F. Monrose, e A. Terzis, «A Multifaceted Approach to Understanding the Botnet Phenomenon», 2006.

[31] Q. Gu e P. Liu, «Denial of Service Attacks».

[32] M. Mohd e S. Abstrak, «User Awareness in Handling Computer Viruses Incident for Windows Platform», 2007.

[33] Z. Long, L. Tan, S. Zhou, C. He, e X. Liu, «Collecting Indicators of Compromise from Unstructured Text of Cybersecurity Articles using Neural-Based Sequence Labelling», Jul. 2019. Available: http://arxiv.org/abs/1907.02636

[34] D. Makrushin, «Indicators of Compromise as an Instrument for Threat Intelligence». Available: https://www.researchgate.net/publication/349211330

[35] M. S. Abu, S. R. Selamat, A. Ariffin, e R. Yusof, «Cyber threat intelligence – Issue and challenges», *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, n. 1, pp. 371–379, Abr. 2018, doi: 10.11591/ijeecs.v10.i1.pp371-379.

[36] N. Miloslavskaya, «A Brief Evolution of Network Protection Tools and Methods», em *Procedia Computer Science*, Jul. 2021, vol. 190, pp. 590–596. doi: 10.1016/j.procs.2021.06.069.

[37] N. Miloslavskaya, «Security operations centers for information security incident management», em *Proceedings - 2016 IEEE 4th International Conference on Future Internet of Things and Cloud, FiCloud 2016*, Set. 2016, pp. 131–138. doi: 10.1109/FiCloud.2016.26.

[38] M. Vielberth, F. Bohm, I. Fichtinger, e G. Pernul, «Security Operations Center: A Systematic Study and Open Challenges», *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3045514.

[39] D. R. Tsai, W. C. Chen, Y. C. Lu, e C. W. Wu, «A trusted security information sharing mechanism», em *Proceedings - International Carnahan Conference on Security Technology*, 2009, pp. 257–260. doi: 10.1109/CCST.2009.5335529.

[40] R. A. Maalem Lahcen, B. Caulkins, R. Mohapatra, e M. Kumar, «Review and insight on the behavioral aspects of cybersecurity», *Cybersecurity*, vol. 3, n. 1. Springer Science and Business Media B.V., Dez. 01, 2020. doi: 10.1186/s42400-020-00050-w.

[41] G. Settanni, Y. Shovgenya, F. Skopik, R. Graf, M. Wurzenberger, e R. Fiedler, «Acquiring Cyber Threat Intelligence through Security Information Correlation», em *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, 2017, pp. 1–7. doi: 10.1109/CYBConf.2017.7985754.

[42] C. Feng, S. Wu, e N. Liu, «A user-centric machine learning framework for cyber security operations center», em *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2017, pp. 173–175. doi: 10.1109/ISI.2017.8004902.

[43] N. Miloslavskaya, «Analysis of siem systems and their usage in security operations and security intelligence centers», em *Advances in Intelligent Systems and Computing*, 2018, vol. 636, pp. 282–288. doi: 10.1007/978-3-319-63940-6_40.

[44] S. Mihindu e F. Khosrow-Shahi, «Collaborative Visualisation embedded Cost-efficient, Virtualised Cyber Security Operations Centre», em *Proceedings of the International Conference on Information Visualisation*, Set. 2020, vol. 2020-September, pp. 153–159. doi: 10.1109/IV51561.2020.00078.

[45] C. Islam, M. A. Babar, e S. Nepal, «Architecture-centric support for integrating security tools in a security orchestration platform», em *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12292 LNCS, pp. 165–181. doi: 10.1007/978-3-030-58923-3_11.

[46] «A SANS Survey Who's Using Cyberthreat Intelligence and How?», 2015. Available: www.sans.org/reading-room/whitepapers/analyst/analytics-intelligence-survey-2014-35507

[47] E. Lindahl, S. O'Hara, e Q. Zhu, «A Multi-Agent System of Evidential Reasoning for Intelligence Analyses», em *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, 2007. doi: 10.1145/1329125.1329461.

[48] A. Elitzur, R. Puzis, e P. Zilberman, «Attack Hypothesis Generation», em *2019 European Intelligence and Security Informatics Conference (EISIC)*, 2019, pp. 40–47. doi: 10.1109/EISIC49498.2019.9108886.

[49] J. Pastor-Galindo, P. Nespoli, F. Gomez Marmol, e G. Martinez Perez, «The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends», *IEEE Access*, vol. 8, pp. 10282–10304, 2020, doi: 10.1109/ACCESS.2020.2965257.

[50] S. Samtani, M. Abate, V. Benjamin, e W. Li, «Cybersecurity as an Industry: A Cyber Threat Intelligence Perspective», em *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, Springer International Publishing, 2019, pp. 1–20. doi: 10.1007/978-3-319-90307-1_8-1.

[51] A. Cartagena, G. Rimmer, T. van Dalsen, L. Watkins, W. H. Robinson, e A. Rubin, «Privacy Violating Opensource Intelligence Threat Evaluation Framework: A Security Assessment Framework for Critical Infrastructure Owners», em *2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020*, Jan. 2020, pp. 494–499. doi: 10.1109/CCWC47524.2020.9031172.

[52] W. Tounsi e H. Rais, «A survey on technical threat intelligence in the age of sophisticated cyber attacks», *Computers and Security*, vol. 72. Elsevier Ltd, pp. 212–233, Jan. 01, 2018. doi: 10.1016/j.cose.2017.09.001.

[53] H. Al-Mohannadi, I. Awan, e J. al Hamar, «Analysis of adversary activities using cloud-based web services to enhance cyber threat intelligence», *Service Oriented Computing and Applications*, vol. 14, n. 3, pp. 175–187, Set. 2020, doi: 10.1007/s11761-019-00285-7.

[54] Y. Ghazi, Z. Anwar, R. Mumtaz, S. Saleem, e A. Tahir, «A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources», em *Proceedings - 2018 International Conference on Frontiers of Information Technology, FIT 2018*, Jan. 2019, pp. 129–134. doi: 10.1109/FIT.2018.00030.

[55] M. Landauer, F. Skopik, M. Wurzenberger, W. Hotwagner, e A. Rauber, «A Framework for Cyber Threat Intelligence Extraction from Raw Log Data», em *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, Dez. 2019, pp. 3200–3209. doi: 10.1109/BigData47090.2019.9006328.

[56] J. L. Lewis, G. F. Tambaliuc, H. S. Narman, e W.-S. Yoo, «IP Reputation Analysis of Public Databases and Machine Learning Techniques».

[57] B. Stojkovski, G. Lenzini, V. Koenig, e S. Rivas, «What s in a Cyber Threat Intelligence sharing platform?», em *ACM International Conference Proceeding Series*, Dez. 2021, pp. 385–398. doi: 10.1145/3485832.3488030.

[58] S. Bromander *et al.*, «Investigating Sharing of Cyber Threat Intelligence and Proposing A New Data Model for Enabling Automation in Knowledge Representation and Exchange», *Digital Threats: Research and Practice*, vol. 3, n. 1, Mar. 2022, doi: 10.1145/3458027.

[59] ANSSI - Agence Nationale de la Sécurité des Systèmes d'Information, «OpenCTI – The Open Source Solution for Processing and Sharing Threat Intelligence Knowledge», *OpenCTI Platform*. https://www.ssi.gouv.fr/en/actualite/opencti-the-open-source-solution-for-processing-and-sharing-threat-intelligence-knowledge/ (acedido Fev. 10, 2022).

[60] OpenCTI Development Team, «GitHub OpenCTI», *GitHub*. https://github.com/OpenCTI-Platform/opencti (acedido Fev. 10, 2022).

[61] OpenCTI Contributors, «OpenCTI Architecture», *OpenCTI Public Knowledge Base*. https://www.notion.so/Architecture-5ce8241eac7e4e24906249e9595314cd (acedido Fev. 15, 2022).

[62] GreyNoise Development Team, «Applying GreyNoise Data to Your Analysis», *GreyNoise Guide*, 2022. https://docs.greynoise.io/docs/applying-greynoise-data-to-your-analysis (acedido Mar. 20, 2022).

[63] AbuseIPDB Development Team, «AbuseIPDB - Frequently Asked Questions», *AbuseIPDB Guide*, 2022. https://www.abuseipdb.com/faq.html (acedido Mar. 20, 2022).

[64] Neutrino Development Team, «IP Blocklist», *Neutrino API Docs*, 2022. https://www.neutrinoapi.com/api/ip-blocklist/ (acedido Mar. 20, 2022).

[65] A. Jøsang, «Trust and Reputation Systems», em *Foundations of Security Analysis and Design IV*, 2007, pp. 209–245.

[66] M. Faiella, G. Gonzalez-Granadillo, I. Medeiros, R. Azevedo, e S. Gonzalez-Zarzosa, «Enriching Threat Intelligence Platforms Capabilities». Available: https://csirtgadgets.com/collective

[67] W. Lalouani e M. Younis, «Multi-observable reputation scoring system for flagging suspicious user sessions», *Computer Networks*, vol. 182, Dez. 2020, doi: 10.1016/j.comnet.2020.107474.

[68] B. Coskun, «(Un)wisdom of crowds: Accurately spotting malicious ip clusters using not-so-accurate IP blacklists», *IEEE Transactions on Information Forensics and Security*, vol. 12, n. 6, pp. 1406–1417, Jun. 2017, doi: 10.1109/TIFS.2017.2663333.

[69] J. Wang *et al.*, «A comprehensive security operation center based on big data analytics and threat intelligence PoS(ISGC2021)028», 2021. Available: https://pos.sissa.it/

[70] T. Nakamori, D. Chiba, M. Akiyama, e S. Goto, «Detecting dynamic IP addresses and cloud blocks using the sequential characteristics of PTR records», *Journal of Information Processing*, vol. 27, pp. 525–535, 2019, doi: 10.2197/ipsjjip.27.525.

[71] M. Ali, S. Shiaeles, G. Bendiab, e B. Ghita, «Malgra: Machine learning and N-GRAM malware feature extraction and detection system», *Electronics (Switzerland)*, vol. 9, n. 11, pp. 1–20, Nov. 2020, doi: 10.3390/electronics9111777.

[72] N. Serketzis, V. Katos, C. Ilioudis, D. Baltatzis, e G. Pangalos, «Improving forensic triage efficiency through Cyber Threat Intelligence», *Future Internet*, vol. 11, n. 7, Jul. 2019, doi: 10.3390/fi11070162.

[73] Andrew. Oram, *Peer-to-peer : harnessing the benefits of a disruptive technology*. O'Reilly, 2001.

[74] H. Dalziel, «How to Define and Build an Effective Cyber Threat Intelligence Capability», em *Syngress*, H. Dalziel, Ed. Boston: Syngress, 2014. doi: https://doi.org/10.1016/B978-0-12-802730-1.00017-X.

[75] Luc. Dandurand *et al.*, *Actionable information for security incident response.* ENISA, 2015. doi: 10.2824/38111.

[76] A. Iklody, G. Wagener, A. Dulaunoy, S. Mokaddem, e C. Wagner, «Decaying Indicators of Compromise», *ArXiv*, Mar. 2018. Available: http://arxiv.org/abs/1803.11052

[77] IBM Development Team, «IBM X-Force Exchange - Frequently Asked Questions», *IBM X-Force Exchange - Information Source*, 2022. https://exchange.xforce.ibmcloud.com/faq#information_source (acedido Mar. 15, 2022).

[78] MyIP Development Team, «MyIP Guide», *MyIP Documentation - About us*, 2022. https://myip.ms/info/about/About_Us.html (acedido Mar. 20, 2022).

[79] Pulsedive Development Team, «Pulsedive Overview», *Pulsedive Guide* , 2022. https://pulsedive.com/about/ (acedido Mar. 20, 2022).

[80] L. Bilge, E. Kirda, C. Kruegel, e M. Balduzzi, «EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis», 2011.

[81] Y. Zhao, B. Lang, e M. Liu, «Ontology-based unified model for heterogeneous threat intelligence integration and sharing», Set. 2017, pp. 11–15. doi: 10.1109/ICASID.2017.8285734.

[82] I. Vacas, I. Medeiros, e N. Neves, «Detecting Network Threats using OSINT Knowledge-Based IDS», em *Proceedings - 2018 14th European Dependable Computing Conference, EDCC 2018*, Nov. 2018, pp. 128–135. doi: 10.1109/EDCC.2018.00031.

[83] S. Gong, J. Cho, e C. Lee, «A Reliability Comparison Method for OSINT Validity Analysis», *IEEE Trans Industr Inform*, vol. PP, p. 1, Set. 2018, doi: 10.1109/TII.2018.2857213.

[84] R. Vinayakumar, P. Poornachandran, e K. P. Soman, «Scalable Framework for Cyber Threat Situational Awareness Based on Domain Name Systems Data Analysis», 2018.

Anexo A

Ferramentas Externas Consultadas

	Características	Limitações do Plano Básico	Fontes dos Dados	Formato do Score/Classificação
GreyNoise[62]	• *Label's* para caracterizar as atividades dos IP's; • Exporta dados no formato JSON; • Classificação qualitativa do IP de acordo com a sua atividade; • Fornece informação sobre dois parâmetros, *Riot* e *Grey*, úteis para auxiliar a averiguar a ameaça dos endereços IPv4 categorizados como "*Uknown*";	• 50 pesquisas **diárias** através da API; • Ausência de *score* quantitativo que permita comparar o grau de ameaça entre diferentes endereços;	• Sensores próprios para analisar o comportamento dos atores associados a cada endereço IPv4 analisado; • Enriquecimento dos dados para dar contexto, a partir de parcerias com terceiros.	3 tipos de classificação: • *Malicious* – comportamentos maliciosos por parte dos atores detetados*;* • *Benign* – quando se conhece as intenções do ator que possui determinado endereço IPv4; • *Uknown* – Quando não são cumpridos os critérios para serem classificados com as classificações anteriores;
IBM X-Force Exchange[77]	• Score do endereço varia de acordo com quantia de evidências e cronologia das mesmas; • Exporta dados no formato JSON;	• 5000 pesquisas **mensais** através da API;	• Bases de dados internos e parcerias com terceiros;	Risco qualitativo e quantitativo: • **"low"** (0< Risco <4); • **"medium"** (4≤ Risco ≤ 6); • **"high"** (6 < Risco ≤ 8); • **"critical"** (8 < Risco ≤ 10)

	• 75 categorias diferentes para classificar os endereços (Bot, Malware, Spam, etc); • Permite avaliar outros IoC, por exemplo domínios.			
MyIP[78]	• Exporta dados no formato JSON; • Registo histórico completo e descritivo sobre endereço IPv4;	• 150 pesquisas **mensais** através da API; • Não indica o tipo de *blacklist* para o qual o IPv4 está listado;	• Bases de dados próprias construídas com base na monitorização dos endereços que visitam as suas páginas de internet;	• Booleano que indica se o endereço IPv4 está listado em alguma *blacklist*;
Neutrino[64]	• Exporta dados no formato JSON; • Endereços IPv4 são removidos automaticamente das *blacklists* após 7 dias sem atividade maliciosa; • 12 tipos de categorias maliciosas de endereços IPv4; • Permite avaliar outros IoC, por exemplo domínios e endereços de email.	• 50 pesquisas **diárias** através da API;	Bases de dados próprias obtidas de 3 fontes essenciais: • Redes autónomas (*bots*, *crawlers* e *honeypots*) que recolhem dados continuamente; • Informações obtidas de diferentes fontes, tais como *firewalls* e sistemas IDS;	• Registo do endereço IPv4 em cada uma das 12 *blacklists* é um booleano;

			• Dados provenientes de fontes públicas, tais como listas negras e informações recolhidas de *threat intelligence.*	
Pulsedive [79]	• Endereços IPv4 considerados como ameaça, ficam automaticamente com uma *label* "retired" após 2 semanas sem atividade maliciosa; • Exporta em STIX 2.1 e JSON; • Permite avaliar outros IoC, por exemplo domínios, endereços de email e URL's.	• 1000 pesquisas **diárias** através da API.	• Dados provenientes de reportes de utilizadores e outras fontes públicas; • Dados recolhidos por "*nodes"* do Pulsedive ajudam a avaliar o risco do IoC, este risco é avaliado com base em decisões que os analistas tomaram na presença desse IoC.	Risco qualitativo e quantitativo: • *Unknown* (0); • *None* (-1)*; • *Low* (1); • *Medium* (2); • *High* (3); • *Critical* (4)*;* *Este valor será utilizado com módulo, de forma a não afetar o *Trust Rating* e o *Threat Score*.

Anexo B

Trabalhos Científicos Semelhantes

Tipos de IoC analisados (*input*)	**Tipo de *threat intelligence* recolhida para desenvolver o modelo**	**Características do modelo**	**Resultado (*output*)**	**Fonte**
Domínios	Domínios maliciosos: *malwaredomains.com*; *Phishtank* e *software* de análise de *malware* – Anubis.	• Sistema EXPOSURE que analisa o DNS passivamente; • Paralelamente, são recolhidos domínios maliciosos e benignos de diferentes fontes; • Caracterização de parâmetros comuns entre os domínios benignos e os malignos; • Modelo utilizado para *machine learning.*	Binário qualitativo: *Benigno* ou *Maligno.*	[80]
Domínios e endereços IPv4	Não aplicável.	• Análise de tráfego DNS para detetar hospedeiros infetados com *botnet*; • Identificação dos endereços IP associados, utilizados como servidores de comando e controlo do *botnet*.	Quantitativo: *Score* de reputação negativo, isto é, quanto mais negativo pior é.	[13]
Endereços IPv4, domínios maliciosos, atores responsáveis por ameaças, *malware*, ferramentas de intrusão, URL's maliciosos, *exploit*	*Threat intelligence* recolhida de diferentes fontes e armazenada como eventos na ferramenta IntelMQ. Tipos de IoC recolhidos: endereços IPv4 maliciosos; domínios maliciosos; entidades responsáveis por ameaças; *Malware;* ferramentas de intrusão;	• Modelo caracterizado por ter dois módulos: módulo da plataforma IntelMQ e módulo de uniformização dos dados de acordo com a ontologia; • Armazenamento na base de dados (MongoDB) em coleções mediante a	Criação de instâncias de entidades, semelhante a uma plataforma de *threat intelligence*, que permite partilhar ameaças.	[81]

	URL maliciosos; *Exploit*; vulnerabilidades identificadas.	sua ontologia e prontos a serem integrados por outras ferramentas.		
Endereços IPv4	*Blacklist* de endereços IPv4	• *Clustering* de endereços IP em maliciosos e benignos, a partir de *blacklists*.	Binário qualitativo: *Benigno; Malicioso.*	[68]
URL's	*Blacklists* de URL's de *malware*: *MalwareURL; MalwareDomains*; *MalwareDomainList;* *Blacklists* de URL's de *phishing*: *Phishtank* e *OpenPhish*.	• Extração de características de URL's maliciosos e benignos; • Classificação de URL's com recurso a *machine learning,* desenvolvido a partir das características extraídas.	Binário qualitativo: *Benigno; Malicioso*	[15]
Endereços IPv4	Dados OSINT obtidos da plataforma IntelMQ.	Modelo denominado *IDSoSint*, composto por 3 módulos: • módulo de correlação de eventos a partir dos IoC's obtidos; • módulo dos indicadores de ataque (IoA); • módulo de configuração do *Intrusion Detection System* (IDS) através de regras e *blacklists*.	Produção de IoC que são processados nos 3 módulos da ferramenta de forma a produzir regras automáticas de configuração do IDS.	[82]
Endereços IP, domínios e *hashes* (MD5)	CTI recolhida de diferentes fontes: *Threat Crowd*; *VirusTotal; Cymon; Open Threat Exchange.*	Sistema composto por 3 camadas: • Camada de processamento dos dados CTI; • Camada de normalização dos dados; • Camada de confiabilidade.	Modelo que visa avaliar a credibilidade da *threat intelligence*, de forma a ser utilizada na indústria, integração com outras ferramentas e partilhada com maior confiança.	[83]
Endereços de IPv4 e domínios	*Malware*: dados recolhidos de *blacklist* públicas não especificadas.	• Análise de dados recolhidos dos sensores DNS, *Border Gateway Protocol* (BGP);	Quantitativo*: score* de reputação atribuído ao domínio, isto é, quanto maior, mais confiável é.	[84]

	Endereços IP e domínios associados a *botnet* recolhidos de fontes públicas não especificadas.	• Análise em tempo real num algoritmo de *machine learning* que permite atribuir um *score* de reputação.	Qualitativo: IP associado ao domínio classificado como *Malicioso* ou *Não Malicioso*.	
Endereços IPv4	*Blacklists* de endereços IP disponibilizadas pela Talos (http://talosintel.com/feeds/ip-filter.blf) Endereços de IP associados a domínios que estão na *blacklist* da página de internet hpHosts (https://www.hosts-file.net/)	Sistema de classificação *Dynamic Attribute based Reputation (*DAbR): • Recolha de endereços IP existentes em diversas *blacklists;* • Extração de metacaracterísticas; • Treino de modelo para classificar endereços maliciosos de forma autónoma;	Quantitativo: *scores* de reputação compreendidos numa escala de 0 a 10, ou seja, quanto maior melhor a reputação.	[12]
Combinação de IoC de baixo e alto nível, entre os quais se destaca endereços IPv4, URL, artefactos de rede, campanhas de ataque e *Common Vulnerability Exposure* (CVE).	Recolha de IoC de diferentes fontes plataformas de *Threat Intelligence Platforms* (TIP).	Sistema constituído por dois módulos: • módulo de IoC recolhidos de fontes externas OSINT e correlacionados entre si com recurso à plataforma *Malware Information Sharing Platform* (MISP)– resulta em IoC compostos; • módulo que correlaciona os IoC compostos com a informação recolhida das outras ferramentas de segurança que analisam a rede das organizações.	Quantitativo: *score* da ameaça (*score threat*).	[66]
Endereços IPv4	VirusTotal; MyIP.ms; Apility.io; AbuseIPDB; Shodan; Censys	Ferramenta *Automated IP Reputation Analyzer Tool* (AIPRA): • Verificação em plataformas externas públicas dos endereços IP detetados; • Recolha de geolocalização dos endereços de IP detetados, obtido a partir da plataforma AbuseIPDB e aplicação de *machine learning*;	Binário qualitativo: *Good;* *Bad*	[56]

URL; domínios e endereços IPv4	*Blacklists:* Cymus; Talos; *Whitelists:* Alexa	Modelo *multi-observable session reputation* (MuSeR): • Atribuição de um *reputation score* a cada pedido efetuado para início de sessão; • Utilização de *machine learning* para avaliar observáveis/IoC envolvidos nos pedidos para inícios de sessão com padrões maliciosos conhecidos; • Consiste em 3 fases: - Análise inter-observáveis; - Pontuação do pedido do utilizador; - *Session scoring.*	Binário qualitativo para categorizar os observáveis: *Good;* *Bad.* Quantitativos (valores entre 0 e 1): *Score* dos observáveis, que corresponde à probabilidade da sua categorização qualitativa; *Request score* – resulta da relação entre os observáveis analisados; *Session score* – baseia-se nos *request scores* associados à atividade do utilizador.	[67]
Endereços IPv4 *e malware*	Histórico associado aos endereços IP recolhido das plataformas de CTI: *VirusTotal; MyIP; AlienVault Open Threat Excahnge*(OTX). *Malware* associados aos endereços de IP através da plataforma *Cuckoo Sandbox*.	• Recolha de dados através de *honeypots;* • Análise dos dados recolhidos, recolha de endereços IP e infeções de *malware;* • Correlação dos dados recolhidos com dados de *threat intelligence* provenientes de plataformas externas, com o objetivo de desenvolver e aprimorar mecanismos de *machine learning.*	Reputação e risco atribuídos endereços IP.	[9]