SIM-Ciber: Uma Solução Baseada em Simulações Probabilísticas para Quantificação de Riscos e Impactos de Ciberataques Utilizando Relatórios Estatísticos

João Nunes, Muriel Franco, Eder Scheid, Geancarlo Kozenieski, Henrique Lindemann, Laura Soares, Jéferson Nobre, Lisandro Granville

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
{jdmnunes,mffranco,ejscheid,gkozenieski,
hlindemann,lrsoares,jcnobre,granville}@inf.ufrgs.br

Abstract. *The evolution of technologies and the growing dependence on digital devices increase cyber risks and cyber attacks, making it essential to understand the risks and their potential impacts from a technical and economic perspective. In this context, this article proposes SIM-Ciber, a solution for simulating risks and technical and financial impacts on companies. SIM-Ciber is based on cybersecurity reports and statistics from reputable companies (e.g., consultancies and service providers) and applies simulation techniques (e.g., Monte Carlo and Bayes Theorem) to understand the risks and impacts of cyberattacks on companies of different sizes, regions, and sectors. The feasibility of SIM-Ciber is demonstrated for Malware, Phishing, and DDoS attacks in different industry sectors, showing high accuracy for determining financial impacts based on real statistics.*

Resumo. *A evolução das tecnologias e a crescente dependência em dispositivos digitais aumentam os riscos cibernéticos e os ciberataques, tornando essencial para a compreensão dos riscos e de seus potenciais impactos a partir de uma perspectiva técnica e econômica. Neste contexto, este artigo propõe o SIM-Ciber, uma solução para simulação de riscos e impactos técnicos e financeiros em empresas. O SIM-Ciber se baseia em relatórios e estatísticas de cibersegurança de empresas reputadas (e.g., consultorias e provedores de serviços) e aplica técnicas de simulação (e.g., Monte Carlo e Teorema de Bayes) para compreender os riscos e impactos de ciberataques em empresas de diferentes tamanhos, regiões e setores. A viabilidade do SIM-Ciber é demonstrada para ataques de Malware, Phishing e DDoS em diferentes setores da indústria, mostrando alta precisão para determinar impactos financeiros com base em estatísticas reais.*

1. Introdução

A rápida evolução da tecnologia e a crescente dependência tecnológica de empresas e serviços traz consigo uma preocupação: a cibersegurança [Alawida et al. 2022]. O mundo corporativo é um dos maiores alvos de cibercriminosos, gerando assim impactos operacionais e econômicos às organizações e pessoas dependentes de sistemas de Tecnologia da Informação (TI). Neste contexto, as empresas devem estar atentas não apenas à perda de informações e disrupção de serviços, mas também aos impactos financeiros [Huang et al. 2023]. Portanto, a compreensão do nível de cibersegurança da empresa, assim como dos riscos de ataques e potenciais impactos econômicos consequentes, são cruciais para um bom planejamento e gestão de cibersegurança [Franco et al. 2023a].

Atualmente, abordagens têm sido propostas para simular e compreender diferentes fatores que influenciam nos comportamentos, riscos e impactos de ciberataques [Kavak et al. 2021]. Por exemplo, [Ahmed et al. 2022] utilizam grafos de ataque para simular

a probabilidade de ataques acontecerem, enquanto [Jawad and Jaskolka 2021] analisaram diferentes técnicas de simulação para mensurar impactos em sistemas industriais. Além disso, Cadeias de Markov são utilizadas em abordagens para descrever ameaças cibernéticas observadas, identificar vulnerabilidades comuns e gerar possíveis ações de defesa visando um uso otimizado dos recursos disponíveis [Gore et al. 2017]. Ambientes simulados também têm sido utilizados para fins educacionais, como, por exemplo, com *Cyber Ranges* que estão sendo utilizados como simuladores para treinamento em cibersegurança [Yamin and Katt 2022]. Assim, as simulações podem atuar como aliados na compreensão do comportamento, dos riscos e também dos impactos de ciberataques. Entretanto, ainda existem poucas soluções eficientes focadas em aspectos econômicos da cibersegurança [Kianpour et al. 2021].

É fundamental haver abordagens que permitam investigar a natureza e o comportamento de ciberataques, desde a sua motivação, vulnerabilidades, estratégias de defesa e também potenciais impactos socioeconômicos [Franco et al. 2022a]. Tais abordagens podem ser baseadas em métricas, modelos e ferramentas existentes para análise de riscos [Roldán-Molina et al. 2017] e investimentos [Gordon et al. 2021] em cibersegurança, de modo a agrupar informações relevantes para o processo de tomada de decisão em cibersegurança. Exemplos de informações incluem quais cenários aumentam os riscos de ser alvo de um ciberataque e também os impactos econômicos diretos e indiretos em caso de ataques realizados com sucesso [Franco et al. 2024]. Entretanto, as soluções da literatura não permitem uma simulação precisa de riscos técnicos e econômicos de ciberataques e não consideram características de empresas (*e.g.*, localização, setor e ativos em risco) nem dados históricos e estatísticos de ciberataques.

Neste artigo, é proposta a solução SIM-Ciber para a coleta e processamento de dados para quantificação e simulação dos riscos de empresas sofrerem determinados ciberataques e quais são seus potenciais impactos econômicos. Para isso, a solução SIM-Ciber *(i)* mapeia e utiliza dados reais e quantificáveis coletados de relatórios estatísticos e técnicos de cibersegurança publicamente disponíveis, *(ii)* implementa métodos probabilísticos (*e.g.*, Teorema de Bayes e Monte Carlo) para definir os riscos e suas relações em diferentes cenários e *(iii)* fornece um relatório sobre os possíveis impactos econômicos e fatores de risco aos quais empresas estão expostas, ajudando na compreensão das situações e na tomada de decisões estratégicas. Além disso, é proposto, como parte da solução, um modelo para classificação da qualidade dos relatórios de cibersegurança utilizados, permitindo assim uma melhor seleção das fontes de dados adicionadas. A validação do modelo de classificação de relatórios foi realizada utilizando métricas e pesos para a análise das fontes dos dados (*i.e.*, empresa ou instituição que coletou os dados presentes em relatórios). Para a avaliação da solução, foram utilizadas requisições simuladas de empresas com diferentes características (*e.g.*, setor e localização geográfica) e potenciais ciberataques, verificando assim a eficácia da SIM-Ciber em compreender os riscos dos ciberataques e seus impactos econômicos.

Este artigo está estruturado na seguinte forma: a Seção 2 discorre sobre os trabalhos relacionados; a Seção 3 introduz e detalha a solução SIM-Ciber, explicando seus componentes e funcionamento; já a Seção 4 foca na avaliação da metodologia proposta, descrevendo os testes utilizados e discutindo os resultados obtidos, e por fim, na Seção 5, apresentamos a conclusão e trabalhos futuros.

2. Trabalhos Relacionados

O planejamento de investimentos em cibersegurança tornou-se significativamente mais complexo com o aumento da digitalização das empresas e a vasta gama de soluções disponíveis no mercado [Havakhor et al. 2020]. A escassez de recursos e a falta de profissionais especializados intensificam esses desafios, especialmente para pequenas e médias empresas [Franco et al. 2023a]. A literatura busca suprir a demanda por ferramentas de

segurança acessíveis e intuitivas, capazes de atender não somente o público técnico, como também pessoas em cargos administrativos e financeiros [Kianpour et al. 2021].

Focado nos aspectos econômicos da cibersegurança, [Gordon et al. 2021] propõe e estende modelos econômicos para cálculo do investimento ótimo que uma organização precisaria fazer em segurança para proteger um conjunto de dados. Porém, o modelo não fornece métodos para análise de riscos das organizações. A ferramenta SECAdvisor foi proposta em [Franco et al. 2023b] como uma solução integrada para aplicação de conceitos de planajemaneto em cibersegurança, incluindo a utilização do modelo Gordon-Loeb e Return On Security Investment (ROSI). Já o modelo RCVaR [Franco et al. 2024] propõe uma metodologia para extrair e utilizar informações de risco provenientes de relatórios estatísticos e dados reais, por exemplo, os ciberataques mais comuns e suas perdas financeiras resultantes. Os dados extraídos são combinados com métodos econômicos para realizar a estimativa do custo que um ataque teria para uma empresa. Apesar de mais abrangente, o modelo não fornece simulações baseadas nos dados obtidos para a predição de riscos futuros.

Embora modelos econômicos possam ser aliados na compreensão de riscos cibernéticos, as análises de riscos ainda dependem de dados manuais que limitam a aplicação de modelos econômicos. Para análise de riscos, o uso de aprendizado de máquina (Machine Learning, ML) tem se tornado cada vez mais proeminente na literatura. O modelo CyRiPred [Kia et al. 2024], baseado em Common Vulnerabilities and Exposures (CVEs), primeiro identifica os principais grupos de risco aplicando técnicas de processamento de linguagem natural em sua base de conhecimento, e depois faz a predição da gravidade dos ataques futuros usando ML a partir de dados históricos da severidade e quantidade dos ataques. Em outro trabalho, A ferramenta proposta em [Subroto and Apriyana 2019] alimenta sua base de dados com conversas de usuários do Twitter em adição aos CVEs para a predição de riscos usando ML, tendo como motivação a justificativa de que a discussão sobre vulnerabilidades acontece de maneira mais ágil nas redes sociais. Ambas as soluções têm como foco as vulnerabilidades, sem fornecer simulações de riscos ou de impactos econômicos específicos a uma empresa.

Ainda usando IA, a ferramenta SecRiskAI [Franco et al. 2022b] determina o nível de exposição de uma empresa a ciberataques. Sua análise de riscos busca correlacionar características da empresa, como, por exemplo, setor de atuação, número de funcionários, e vulnerabilidades conhecidas. A ferramenta não utiliza dados de relatórios estatísticos em sua análise. Já [Jacobs et al. 2023] desenvolveram um modelo adaptativo que calcula pontuações para CVEs existentes se baseando em dados reais. Os autores analisaram relatórios totalizando 6,4 milhões de ataques, e observaram que, enquanto milhares de vulnerabilidades são conhecidas, apenas 6% delas são efetivamente exploradas. Com essas informações, o *Exploit Prediction Scoring System* (EPSS) é introduzido como um sistema de pontuação para previsão e priorização de vulnerabilidades. Os impactos econômicos das vulnerabilidades analisadas não são abordados. O EPSS, embora recente, tem sido utilizado como um aliado para priorização de riscos na indústria e também investigado pela academia como uma potencial base para integração com modelos econômicos.

A Tabela 1 resume os trabalhos relacionados encontrados na literatura. As soluções atuais focam na análise de riscos e vulnerabilidades utilizando métricas como Common Vulnerabilities and Exposures (CVE) e Exploit Prediction Scoring System (EPSS). Também existem soluções que utilizam de Inteligência Artificial (IA) para prever riscos de cibersegurança e inferir informações ausentes devido à falta de compartilhamento de informações sobre ciberataques. Porém, embora as magnitudes de impactos possam ser computadas com base nestas soluções, ainda são escassos os trabalhos que quantifiquem os reais impactos técnicos e econômicos de ciberataques. Ainda, embora existam trabalhos utilizando dados coletados de redes sociais, incidentes e de interações com especia-

Tabela 1. Comparação da SIM-Ciber com a Literatura

Solução	Objetivo	Relatórios Estatísticos	Simulações	Análise de Riscos	Impactos Econômicos
[Gordon et al. 2021]	Cálculo de investimento ótimo em cibersegurança	Não	Sim	Não	Sim
SECAdvisor [Franco et al. 2023b]	Planejamento de investimentos em cibersegurança utilizando modelos econômicos	Não	Sim	Não	Sim
RCVaR [Franco et al. 2024]	Calcular as possíveis perdas financeiras em caso de ciberataques	Sim	Não	Não	Sim
[Kia et al. 2024]	Classificação e predição de riscos usando informações de CVEs e dados da Wikipedia	Não	Não	Sim	Não
[Subroto and Apriyana 2019]	Predição de riscos e vulnerabilidades usando dados de redes sociais	Não	Não	Sim	Não
SecRiskAI [Franco et al. 2022b]	Análise de riscos de ciberataques em empresas usando IA	Não	Não	Sim	Não
EPSS [Jacobs et al. 2023]	Predição de riscos e priorização de vulnerabilidades usando EPSS	Sim	Sim	Sim	Não
SIM-Ciber (Este trabalho)	Classificação de relatórios, simulação de riscos e impactos econômicos	Sim	Sim	Sim	Sim

listas, a literatura ainda carece de trabalhos que utilizem dados de relatórios estatísticos reais de impactos econômicos de cibersegurança.

Portanto, existe uma oportunidade para soluções baseadas em relatórios estatísticos de empresas de cibersegurança [Franco et al. 2024] e também em simulações computacionais (*e.g.*, Monte Carlo e Teorema de Bayes) [Engström and Lagerström 2022] para inferir potenciais riscos e impactos econômicos de ciberataques em empresas.

3. Solução SIM-Ciber

A solução SIM-Ciber busca endereçar desafios encontrados na literatura para análise de riscos ao mapear e prover informações econômicas e técnicas para compreensão de riscos em empresas, ajudando assim na quantificação de riscos cibernéticos e seus impactos. A arquitetura da solução proposta é apresentada na Figura 1. A solução é dividida em três módulos, os quais utilizam técnicas probabilísticas para uma melhor garantia e precisão nas informações a serem entregues ao usuário. Detalhes dos processos, técnicas, dados recebidos e informações entregues ao usuário são explicados e apresentados ao longo desta seção.

O *Módulo de Relatórios* recebe como entrada dados de relatórios e estatísticas de cibersegurança relacionados aos setores da indústria e às localizações das empresas sendo analisadas, bem como dados sobre ciberataques e seus impactos. Esses dados são utilizados pelo SIM-Ciber e são utilizados conforme o exemplo a seguir: dado que um *Phishing* ocorra, a probabilidade de que a empresa atingida seja do setor financeiro é 13.2% [Zimperium 2023]. As Empresas de Consultoria (ECs) que fornecem esses dados são ponderadas por métricas, permitindo uma classificação das ECs por *Notas* e auxiliando na avaliação da confiabilidade dos dados coletados. Todo o conjunto de informações recebidos e inseridos nesse módulo são então transferidos para o *Módulo de Dados*.

O *Módulo de Dados* é responsável pelo gerenciamento do Banco de Dados. Nesse módulo, os dados são recebidos (*e.g.*, oriundos dos outros módulos e de *curadores*), organizados e verificados, sendo posteriormente armazenados no *Banco de Dados*. É nesse

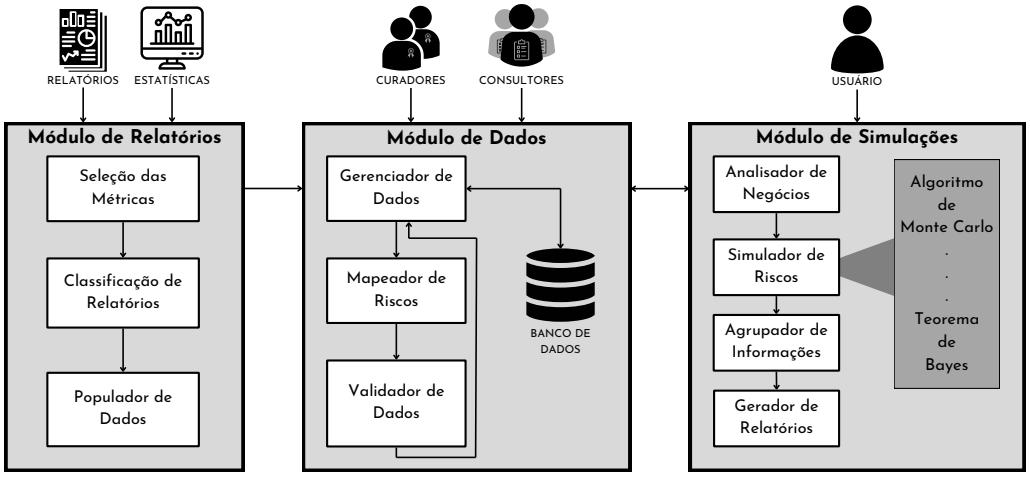


Figura 1. Arquitetura da Solução SIM-Ciber

módulo que ocorre o processo de separação e de organização dos dados em tangíveis e não tangíveis, preparando-os corretamente para a utilização no *Módulo de Simulações*. Dados tangíveis são aqueles utilizados para mensurar os impactos econômicos (*e.g.*, dado o sucesso de um ciberataque, o custo é de X milhões de dólares) e dados que não tangíveis são aqueles que auxiliam a entender estatisticamente o cenário da empresa (*e.g.*, a probabilidade de uma empresa do setor de comércio sofrer um ataque de *Phishing*). Os *curadores* são usuários qualificados (*e.g.*, especialistas em cibersegurança e administradores) e de boa reputação, com permissão para gerenciar os dados inseridos no *Banco de Dados*. Já os *consultores* são usuários autorizados que utilizam o SIM-Ciber para adicionar informações temporárias ao *Banco de Dados*, permitindo simulações de forma personalizada para empresas. Portanto, a separação dos dados em tangíveis e não tangíveis, assim como a presença de *curadores* e *consultores* no *Módulo de Dados* desempenham um papel crucial na manutenção orgânica, extensível e segura da SIM-Ciber e dos dados inseridos.

O *Módulo de Simulações* provê uma interface de interação com o usuário, projetada para processar requisições que contêm dados específicos de empresas. Ao receber uma requisição com informação de, por exemplo, localização e setor de uma empresa, o módulo realiza uma análise dos dados fornecidos, executando simulações para identificação e compreensão dos riscos e dos impactos técnicos e econômicos dos ciberataques. As simulações buscam, mediante filtros baseados nas características da empresa, os dados mais relevantes existentes no *Banco de Dados* que se adéquem ao perfil da empresa, executando algoritmos e técnicas probabilísticas, como, por exemplo, o algoritmo de Monte Carlo e Teorema de Bayes. Ao fim, são gerados relatórios (*e.g.*, probabilidades de ataques e custos envolvidos) no cenário de cibersegurança em relação à empresa analisada.

3.1. Módulo de Relatórios

Este módulo é responsável por receber os dados coletados de relatórios e estatísticas que serão utilizados para mensurar os riscos e impactos de um ciberataque numa empresa. As ECs que forneceram os dados são classificadas utilizando métricas, como demonstrado na Tabela 2, permitindo assim que as ECs sejam classificadas por *Notas*. Estas métricas foram definidas previamente através de revisão da literatura e da análise das principais características de distinções das empresas.

Tabela 2. Exemplos de ECs, Métricas e Notas

	Reputação ***(Rep)***	**Periodicidade** ***(Per)***	**Cobertura** ***(Cob)***	**Escopo** ***(Esc)***	**Abrangência dos Ataques** ***(Abr)***	**Metodologia de Pesquisa** ***(Met)***	**Total com Peso** ***(TP)***	**Notas**
Radware	1	1	2	1	2	2	4.5	Relevante
Verizon	2	2	2	2	2	2	6	Muito Relevante
Zayo	1	1	1	1	1	1	3	Relevante

O componente *Seleção de Métricas* atua adicionando valores às métricas, para o auxílio na compreensão da confiabilidade dos dados coletados das ECs dos relatórios e estatísticas. Nesta etapa, é empregado o uso de IA Generativa (*e.g.*, ChatGPT e Gemini) e também a verificação empírica, conferindo maior firmeza nas designações dos valores. As métricas analisadas são: *Reputação* da empresa, *Periodicidade* de publicação, *Cobertura*, *Escopo*, *Abrangência dos ataques* e *Metodologia* de pesquisa. As métricas são quantificadas, para cada EC, em valores entre 0 e 2. A descrição de cada métrica e seus respectivos valores estão indicados na Tabela 3.

Tabela 3. Métricas Definidas para Análise de Relatórios e Fontes de Dados

Métricas	**Definição**	**Valores**
Reputação	Classifica a EC em relação a sua reputação técnica e maturidade dos processos implementados	0= EC desconhecida 1= EC reconhecida nacionalmente 2= EC reconhecida mundialmente
Periodicidade	Verifica a frequência de publicações de dados da EC	0= Compilados de outras fontes 1= Publicação mensal/semestral 2= Publicação anual
Cobertura	Verifica o alcance do estudo dos relatórios publicados, em relação à um país/continente ou globalmente	0= EC não menciona 1= Cobertura local/continental 2= Cobertura global
Escopo	Avalia se a EC publica relatórios com dados de um único ou de múltiplos setores da indústria	0= EC não menciona 1= Setorial (único) 2= Multisetorial
Abrangência dos ataques	Indica se a EC publica relatórios sobre um tipo de ciberataque ou mais	0= EC não menciona 1= Apenas um tipo de ataque 2= Tipos variados de ataques
Metodologia de pesquisa	Tem como foco analisar se a EC utilizou métodos bem definidos para a coleta e fornecimento dos dados	0= Sem metodologia 1= Sem metodologia mas com inferências 2= Possuem metodologias e apresentam resultados completos

No componente *Classificação de Relatórios*, cada uma das ECs que fornecem os dados recebidos como entrada neste módulo são avaliadas, e cada uma das métricas recebe um peso referente ao seu nível de importância para o processo de classificação. Testes foram realizados nas *Notas* resultantes a partir das Fórmulas 1 e 2, onde foram analisadas estaticamente uma melhor conformidade com a realidade. Assim também, os pesos foram testados e definidos, na qual foram favorecidas as métricas mais relevantes, tais como a reputação e a metodologia de pesquisa, com a finalidade de ajustar a classificação. Assim, cada EC recebe uma *Nota*, que é calculada pela média ponderada das métricas e seus respectivos pesos, conforme descrito na Fórmula 2. A *Nota* é então utilizada para distinguir a credibilidade das ECs, permitindo que os usuários definam quais dados serão utilizados no processamento de requisições na SIM-Ciber. Por exemplo, alguns usuários podem utilizar todos os dados disponíveis durante as simulações, enquanto outros usuários desejam apenas utilizar dados oriundos de ECs com muita relevância. As ECs são classificadas com base nas *Notas* computadas, sendo definidas como: pouco relevante ($\mathbf{TP} \leq 3$), relevante ($3 \leq \mathbf{TP} \leq 5$) e muito relevante ($\mathbf{TP} > 5$).

Tabela 4. Exemplos de Dados Não Tangíveis Utilizados na Simulação

Condição A	Condição B	Probabilidade	Fonte
Malware	Setor de Comércio	21.74%	[Fortinet 2021]
Ransomware	Setor Financeiro	64%	[Sophos 2023]
Ciberataque	Pequenas e Médias Empresas	43%	[Verizon 2023]
DDoS	Brasil	1.75%	[Microsoft 2022]

Tabela 5. Exemplos de Dados Tangíveis Utilizados na Simulação

Condição A	Condição B	Valor	Métrica	Fonte
Custo	Ransomware	$ 170,404	Valor por Ataque	[Sophos 2021]
Custo — Brecha	Brasil	$ 1.22 M	Valor por Ataque	[IBM 2023]
Ransomware	-	$ 693.3 M	Ataques	[SonicWall 2023]

$$TS = \frac{Rep + Per + Cob + Esc + Abr + Met}{3} \tag{1}$$

$$TP = \frac{(Rep*P_{Rep})+(Per*P_{Per})+(Cob*P_{Cob})+(Esc*P_{Esc})+(Abr*P_{Abr})+(Met+*P_{Met})}{P_{Rep}+P_{Per}+P_{Cob}+P_{Esc}+P_{Abr}+P_{Met}} * 3 \tag{2}$$

Por fim, neste módulo, o *Populador de Dados* é responsável por receber e realizar a transferência de todos os dados (atualizados após as adições das métricas e das *Notas*) para o *Módulo de Dados*, onde são verificados e adicionados no *Banco de Dados*.

3.2. Módulo de Dados

Este módulo gerencia, filtra e prepara os dados para utilização nas simulações e análises realizadas pelo SIM-Ciber. Para realizar tais procedimentos, o módulo recebe como entrada os dados vindos do *Módulo de Relatórios* ou pelos *curadores* e *consultores*. Cada um dos dados recebidos possui os seguintes campos: EC geradora do dado, ano de coleta, *Nota* da EC, condições, valor (tangível ou não) e referência para a fonte dos dados. Cada condição mostra a razão de uma determinada situação ocorrer baseada no Teorema de Bayes [Chockalingam et al. 2017] e o valor indica a probabilidade de ocorrência do incidente ou o impacto financeiro, em caso do risco ocorrer. As Tabelas 4 e 5 mostram os tipos de dados que podem ser inseridos nesse módulo.

O *Gerenciador de Dados* é responsável por receber os dados do *Módulo de Relatórios*, dos *curadores*, dos *consultores*, dos dados filtrados pelo *Mapeador de Riscos* e também do *Módulo de Simulações*. É um componente central pois atua como um intermediário na relação de inserção no *Banco de Dados e* também na interação entre os módulos.

Os dados recebidos podem variar conforme a necessidade do modelo, sendo assim, é importante que sejam previamente mapeados quais dados são necessários. O modelo atual é composto por quatro campos relevantes: local, que é onde se encontram geograficamente as empresas (*e.g.*, América Latina (LATAM), Brasil, Estados Unidos); setores da indústria, que são onde as organizações atuam nos diferentes segmentos do mercado (*e.g.*, Comércio, Saúde e Financeiro); tipos de ciberataques, que, por definição, são meios e técnicas maliciosas utilizadas para realizar ataques cibernéticos [Snider et al. 2021] (*e.g.*, *Malware*, *Phishing* e *DDoS*) e os impactos, que são referentes às consequências operacionais e financeiras negativas provenientes do sucesso de um ciberataque [Huang et al. 2023] (*e.g.*, dados criptografados e disrupção de um serviço). Os dados são inseridos seguindo a lógica do Teorema de Bayes [Berger et al. 2020], onde visa compreender

a probabilidade de um evento ocorrer, dado previamente o conhecimento de outro evento. Novos dados podem ser adicionados ao Banco de Dados seguindo as etapas definidas na solução SIM-Ciber.

Os próximos componentes tratam diretamente com os dados. No *Mapeador de Riscos*, é realizada a separação dos dados recebidos como entrada e verifica quais são tangíveis (ou não) e quais são necessários para a análise. Essa verificação permite que os dados possam ser inseridos corretamente no *Banco de Dados*. E, por fim, o *Processador de Dados* verifica a conformidade dos dados, garantindo que os campos estão devidamente preenchidos antes de enviá-los para o *Gerenciador de Dados* para estarem disponíveis no *Banco de Dados*.

3.3. Módulo de Simulações

Este módulo é responsável por receber a requisição do usuário, analisar a empresa e realizar simulações para que seja gerado um relatório sobre os riscos potenciais impactos para a empresa. É o módulo de assimilação de conhecimento, pois utilizam simulações para a compreensão dos riscos e impactos de ciberataques via métodos probabilísticos, que permitem uma maior credibilidade nos dados gerados.

As requisições de empresas incluem o tipo de ataque, o setor da indústria e o local da empresa (*e.g.*, continente, região ou país) que se deseja verificar. Além disso, existe a possibilidade de adicionar palavras-chave após os indicativos, para especializar ainda mais a simulação, e também de escolher quais dados serão utilizados através da filtragem dos relatórios estatísticos disponíveis, utilizando assim as *Notas* das ECs fornecedoras de relatórios. Todas informações contidas na requisição são consideradas no momento da simulação de riscos e na preparação do relatório de análise.

A partir da requisição, o *Analisador de Negócios* tem por finalidade compreender o foco desejado da empresa, seu setor e localização, assim como os ataques a serem analisados. Com isso, é possível requisitar com precisão os dados presentes no *Gerenciador de Dados*, de forma que sejam relevantes para a avaliação da empresa. A precisão dos dados requisitados garante uma coerência com a requisição, pois, dado um grande banco de dados, é necessário que os dados a serem analisados sejam relacionáveis e permitam que as simulações sejam executadas corretamente.

Os dados requisitados no componente anterior chegam ao *Simulador de Riscos*, que os utiliza para calcular os riscos e impactos dos ciberataques em uma empresa com as características definidas na requisição. Para isso, são usadas técnicas matemáticas e probabilísticas, como o algoritmo de Monte Carlo e o Teorema de Bayes, para que a análise tenha uma maior confiabilidade estatística. Cada um dos métodos probabilísticos possuem suas aplicações específicas na solução SIM-Ciber.

O algoritmo de Monte Carlo permite calcular a probabilidade de eventos complexos que possuam vários fatores envolvidos (*e.g.*, simular a probabilidade de ataques em uma empresa com base na sua localização geográfica e setor), ou seja, em eventos incertos, baseados na realidade. Por outro lado, o Teorema de Bayes é um método analítico e atua na compreensão da probabilidade entre dois eventos ocorrerem, permitindo assim ajustar as expectativas sobre o evento principal. Portanto, esse teorema proporciona a geração de informações adicionais através da inferência, como, por exemplo, a probabilidade de ocorrer um ataque de *Phishing* no setor financeiro dado que já é conhecida a probabilidade de ocorrer *Phishing* e de uma empresa ser do setor financeiro (ambas probabilidades conhecidas separadamente). Assim, juntamente com a requisição, os conteúdos a serem simulados levam em conta os dados tangíveis e não tangíveis, a fim de compreender quais são as porcentagens de ocorrer os ciberataques e os seus impactos e também permitir uma estimativa dos custos econômicos da empresa, caso o ataque ocorra com

sucesso. Por fim, as informações geradas são enviadas adiante na solução e também armazenadas pelo *Gerenciador de Dados* para usos futuros de outros usuários.

Por fim, o *Agrupador de Informações* é responsável por receber as informações geradas e organizá-las para que sejam redirecionadas para o *Gerador de Relatórios*, e o *Gerador de Relatórios* recebe e formata as informações, gerando um relatório completo para o usuário. Este relatório dispõe de uma análise da empresa (informada na requisição do usuário), os riscos da empresa de sofrer os ciberataques (com base nas simulações realizadas previamente) e também de uma estimativa de perda financeira, decorrente dos ataques. O relatório também conta com recomendações de segurança às empresas, para poderem adotar métodos que as ajudem a ficar mais resilientes aos futuros ciberataques. Para tais recomendações, a solução pode ser integrada com sistemas de recomendação de proteção [Ferreira et al. 2023] e IA Generativa (*e.g.*, ChatGPT e Gemini).

4. Avaliação

A solução proposta foi avaliada em dois diferentes quesitos *(i)* precisão e capacidade de classificar relatórios relevantes para simulação de riscos e *(ii)* resultados das simulações de riscos de ataques acontecerem e seus potenciais impactos econômicos para empresas de diferentes setores. Para isso, foram utilizados dados coletados de relatórios estatísticos para simulação e também gerado empresas hipotéticas com diferentes configurações, situadas nos setores de Finanças, Saúde e Comércio. As avaliações realizadas e seus resultados são discutidos em detalhes, respectivamente, nas Seções 4.1 e 4.2.

Um protótipo do SIM-Ciber foi implementado utilizando *Python 3.11.9*, juntamente com as bibliotecas *Pandas 2.2.1* e *Numpy 1.26.4*. Além das tecnologias citadas, o Banco de Dados foi construído utilizando o *SQLite 3.38.5*. As avaliações foram executadas usando um computador com 8 GB de memória RAM, armazenamento HD de 1 TB, processador Intel de 5ª geração e com o sistema operacional *Debian* versão 12.5. O código-fonte e os resultados das avaliações se encontram publicamente disponíveis no repositório no Github[1], incluindo artefatos, dados e documentação.

4.1. Classificação dos Relatórios

Com o objetivo de verificar a precisão das *Notas* geradas, foi decidido avaliar as métricas utilizando pesos e para isso foram utilizadas duas fórmulas: Total com Soma (**TS**), que é uma média simples dos valores das métricas, e Total com Peso (**TP**), que é a média ponderada dos valores das métricas e seus respectivos pesos. As Equações 1 e 2 demonstram as fórmulas nas quais foram aplicados testes, de forma a variar os valores dos pesos e identificar o impacto na geração das *Notas*.

Os pesos variavam de zero a dez e seguiam a regra de que os pesos da reputação e da metodologia deverem ser superiores às demais métricas, pois foi definido previamente que tais métricas são as mais relevantes e que geram maior efeito na detecção na credibilidade dos dados gerados pelas Empresas de Consultoria (ECs). Foram gerados todos os pesos possíveis que satisfaziam a regra, de tal maneira que os resultados percebidos expressavam que a maioria dos conjuntos de pesos tiveram uma precisão de 90% ou mais e nenhuma manteve 100% das *Notas*, indicando que toda aplicação de pesos altera em uma pequena parcela nas *Notas* das EC (*cf.* Figura 2). Portanto, as *Notas* de ECs que estavam nos limites dos Totais (**TS** e **TP**) foram alteradas.

De todos os pesos possíveis, foi selecionado um dos conjuntos de pesos (10, 5, 8, 6, 5 e 10, para cada métrica respectivamente) com precisão de 95% para realizar trinta rodadas de testes e verificar o comportamento das *Notas* de forma mais específica. Em

[1]https://github.com/ComputerNetworks-UFRGS/SIM-Ciber

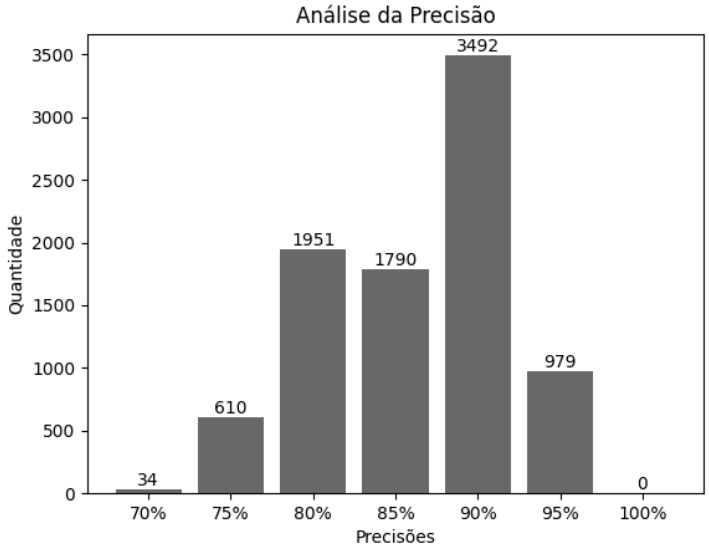


Figura 2. Análise das Precisões dos Pesos

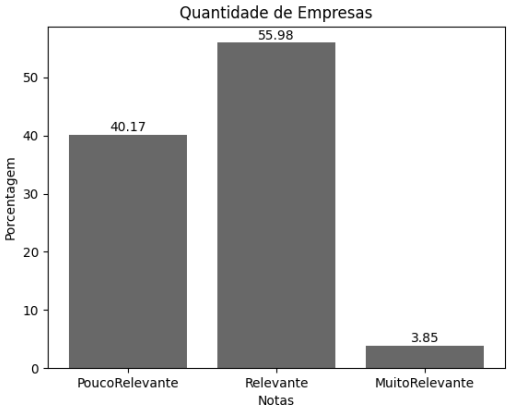
cada rodada, foram gerados dez mil ECs com métricas aleatórias, assim permitindo simular de forma abrangente todas as entradas possíveis para a solução SIM-Ciber e verificar o comportamento das *Notas* das ECs. Através dos testes, foram possíveis obter as seguintes conclusões:

- Houve uma concentração de ECs de classificações *Relevante* e *Pouco Relevante*, indicando que, para uma empresa ser do tipo *Muito Relevante*, os valores das métricas da empresa deviam ser altos. Este comportamento pôde ser observado na Figura 3(a);
- Grande parte das ECs mantiveram a mesma *Nota* após a aplicação dos pesos nas métricas e as ECs que trocaram de *Nota* tiveram sua *Nota* reduzida em um nível (*i.e.*, *Relevante* $\rightarrow$ *Pouco Relevante* e *Muito Relevante* $\rightarrow$ *Relevante*). Esta observação pôde ser verificada na Figura 3(b), onde ”P”significa *Pouco Relevante*; ”R”, *Relevante* e ”M”, *Muito Relevante*. A junção das letras indica as transições de *Notas* das ECs (*i.e.*, ”PR”= *Pouco Relevante* $\rightarrow$ *Relevante*);
- A percepção de transição foi ligeiramente suave, como demonstrado na Figura 3(c), onde a maioria manteve suas *Notas* após a aplicação dos pesos nas métricas e que poucos trocaram de *Notas*, indicados por ”Falsos Positivos”(*i.e.*, *Notas* melhoraram) e ”Falsos Negativos”(*i.e.*, *Notas* pioraram).

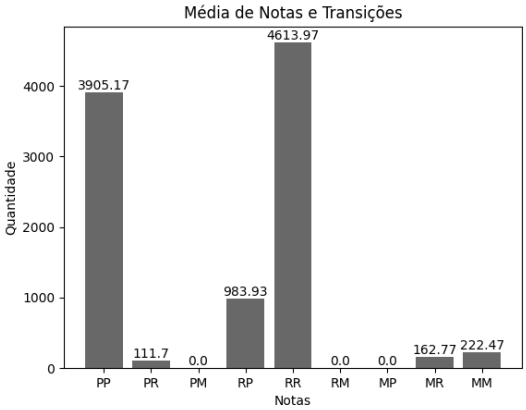
Após os testes realizados, foi decidido a utilização de pesos, pois identificou que essa decisão produziu efeitos positivos para a classificação das ECs, de forma a agregar qualidade nas informações geradas do SIM-Ciber. Assim, a solução foi aplicada utilizando os pesos 10, 5, 8, 6, 5 e 10 para cada métrica respectivamente, de maneira que a permitir uma melhor classificação das ECs no componente *Classificação dos Relatórios*, presente no *Módulo de Relatórios*.

4.2. Simulação de Riscos e Impactos

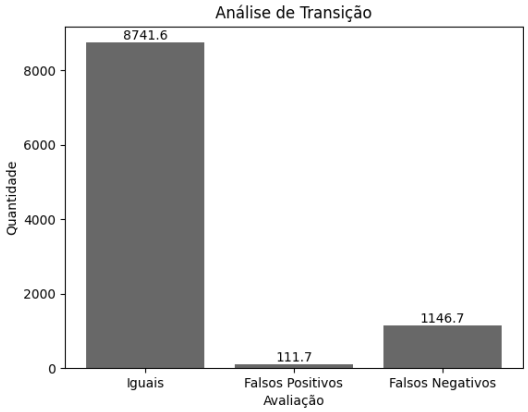
Com os dados prontos, classificados e adicionados ao Banco de Dados, foram realizadas as simulações de riscos e verificadas as suas precisões. As simulações se baseavam na compreensão do cenário de uma empresa a partir de uma requisição recebida como entrada no *Módulo de Simulações*, de forma a gerar resultados significativos e auxiliar nas tomadas de decisões das empresas. Essa subseção explica como as requisições estão formatadas e também avalia os resultados gerados pelas simulações.



(a) Média de ECs por Notas



(b) Média das Transições de Notas



(c) Análise das Transições de Notas

Figura 3. Avaliação da Classificação dos Relatórios

Simulações para analisar os riscos e impactos em cenários do setor financeiro, saúde e comércio foram realizadas, tendo os riscos de ataques de *Malware*, *Phishing* e *DDoS* avaliados. Os resultados apresentam os custos médios por tipo de ataque e as diferenças por setor. Todas as simulações são realizadas utilizando como base os relatórios estatísticos previamente coletados, totalizando 540 dados tangíveis e não tangíveis oriundos de 51 ECs e relatórios diferentes.

4.2.1. Configuração Inicial

Para a realização das simulações, foi necessária a geração de requisições, permitindo assim uma melhor compreensão dos resultados obtidos nas simulações. As requisições são apresentadas em arquivos no formato de texto (*.txt*) e seguem um modelo constituído de quatro linhas, com cada linha possuindo informações referentes à empresa ou ao ciberataque. Um exemplo da requisição é demonstrado na Tabela 6. A primeira linha contém a informação do setor da empresa; na segunda linha, a informação de quais ataques a empresa irá sofrer; na terceira linha, a localização geográfica da empresa, e na quarta linha, o nível de relevância dos dados utilizados nas simulações, com base nas *Notas* das ECs que fornecem os dados. Nas linhas abaixo, há também a possibilidade de inclusão

de palavras-chave que possam ajudar a tornar mais específica as simulações, como, por exemplo, o país da empresa.

Tabela 6. Exemplo de Requisição

Informações na Requisição	Significado
001	Setor da Empresa: Setor de Saúde
110	Tipo de Ataque: Malware e Phishing
0100	Localização Geográfica: LATAM
100	Relevância dos Dados: Todos
Brasil	Informações Extras: Brasil

Com o modelo de requisições padronizado, foram gerados exemplos de requisições com diferentes cenários suportados pelo SIM-Ciber (exceto as informações extras), servindo como base para as simulações e suas avaliações. Assim, a avaliação das simulações também pôde ocorrer e ajudar na análise da qualidade das informações expressas no relatório final.

4.2.2. Análise dos Riscos e Impactos

Após as configurações iniciais, todas as requisições geradas foram utilizadas na solução, onde foi possível perceber o comportamento de cada informação presente após serem testadas nas simulações. Tais simulações utilizam Monte Carlo, com cem mil rodadas cada, para estimar os possíveis custos que a empresa terá, decorrentes da sua localização geográfica, do seu setor na indústria e dos riscos dos ciberataques e de seus impactos. A lógica presente para a confirmação de um ataque é: em cada rodada, se uma probabilidade aleatória gerada for maior que a probabilidade de ocorrer o ataque nas situações informadas (com base nos dados do Banco de Dados), então o ataque ocorreu. Com o ataque confirmado, a mesma lógica é aplicada aos impactos referentes a cada ciberataque, conforme descrito na Tabela 7. Os impactos foram escolhidos a partir do entendimento da forma de atuação de cada um dos ataques, de forma a sintetizar as suas operações e as suas consequências técnicas. Por fim, no encerramento da rodada, com a confirmação do ataque e de seus impactos, é calculado o impacto financeiro possível que a empresa sofreria.

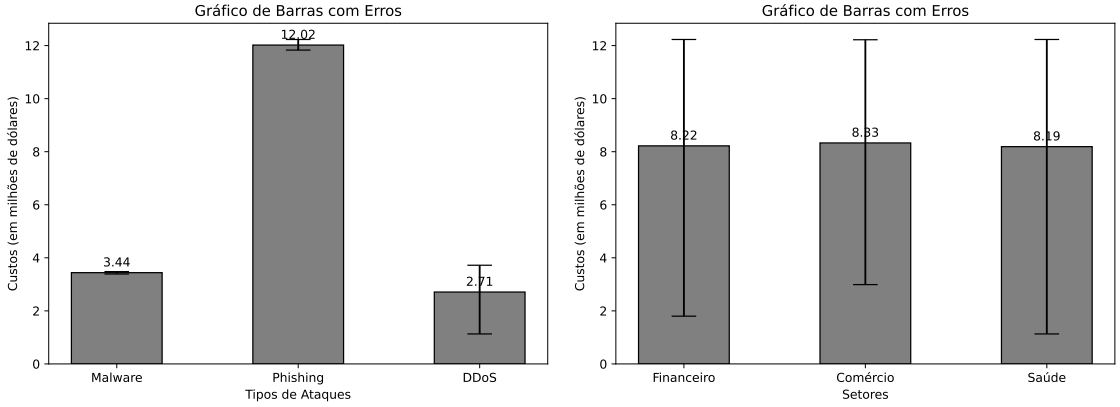
Tabela 7. Exemplos de Ciberataques Utilizados e seus Impactos Técnicos

Ciberataques	Impactos Técnicos
Malware	Vazamento de dados, dados criptografados, perda de desempenho ou e indisponibilidade do sistema
Phishing	Vazamento de dados, roubo de credenciais e perda de desempenho ou conectividade de sistemas
DDoS	Perda de desempenho, conectividade ou indisponibilidade do sistema

Após simular todas as rodadas com Monte Carlo, é encontrado o custo médio, assim como os custos mínimos e máximos que a empresa teria naquelas condições informados na requisição. Para uma análise comportamental da solução, para cada uma das requisições foram efetuadas cem rodadas de testes, com o objetivo de permitir uma avaliação da variedade das respostas das simulações e, assim, da capacidade do SIM-Ciber de compreender o cenário e fornecer informações relevantes para a empresa. As avaliações foram realizadas analisando os resultados gerados, visando disponibilizar uma representação gráfica dos impactos financeiros pelos setores e pelos tipos de ataques.

Em busca desse objetivo, para cada rodada dos testes foram armazenados os custos médios, conforme o ponto de vista a ser analisado. Para a análise dos resultados,

foram analisadas apenas as requisições contendo apenas um ciberataque por rodada (*e.g.*, requisições apenas com *Phishing*, sem *Malware* e sem *DDoS*). De maneira a facilitar a visualização dos valores nos gráficos, também foi realizada a divisão dos valores originais obtidos por um milhão (10^6).



(a) Custos Médios por Tipo de Ataque **(b)** Custos Médios por Setor

Figura 4. Análise dos Custos Médios

Analisando as informações por tipo de ataque, demonstrado pela Figura 4(a), podemos compreender que: o tipo de ataque *Phishing* é o que mais causa impacto financeiro e que, além dos custos de *Malware* e de *DDoS* serem próximos, é o *DDoS* que possui a maior variação de custo médio. A baixa variação de custo médio de *Malware* indica que o valor médio permanece, independentemente do setor. Assim, concluímos que *Phishing* ganha um plano de destaque no quesito financeiro perante os demais tipos de ataque.

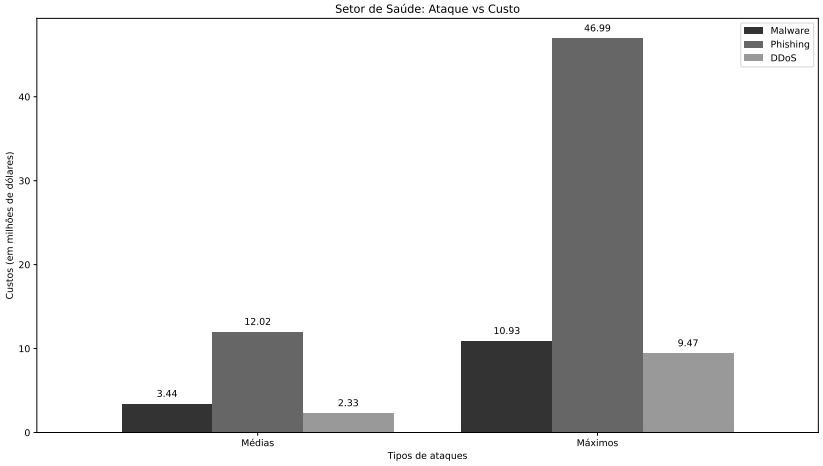


Figura 5. Custos por Tipo de Ataque no Setor de Saúde

Analisando as informações por setor, demonstrado pela Figura 4(b), podemos compreender que os custos médios são semelhantes, onde podemos deduzir que o impacto financeiro médio de um ataque não possui uma grande variação. O setor de saúde possui uma maior variação do valor mínimo de custo médio, possivelmente indicando que, quando uma empresa do setor sofre um ataque, há uma maior velocidade na solução

e consequentemente um menor impacto financeiro. Por tais inferências, constata-se que todos os tipos de ataques possuem relevância, independentemente do setor em que se encontra a empresa.

Para compreender de forma mais específica os custos por ataque, foi realizada uma análise apenas do setor de saúde, para ser possível uma melhor compreensão dos impactos financeiros de cada tipo de ataque. Os resultados foram apresentados na Figura 5, onde foram informados os custos médios e máximos separadamente de cada tipo de ataque. Assim, feita a análise, *Phishing* se mantém como o ataque que mais causa impacto financeiro (seguido do *Malware* e do *DDoS*) e que a diferença dos custos médios e máximos expressa o quão financeiramente danoso pode ser um ataque.

5. Conclusões e Trabalhos Futuros

A solução proposta demonstrou ser robusta e eficaz na coleta, processamento e análise de dados para quantificar e simular os riscos de ciberataques em ambientes corporativos, bem como na avaliação de seus impactos econômicos potenciais. Ao mapear dados reais provenientes de fontes confiáveis de cibersegurança e aplicar métodos probabilísticos avançados, como o Teorema de Bayes e Monte Carlo, a solução SIM-Ciber oferece uma visão abrangente dos riscos em diferentes cenários, proporcionando análises e *insights* cruciais para a tomada de decisões estratégicas por gestores e especialistas em cibersegurança. Além disso, a introdução de um modelo de classificação de qualidade para os relatórios utilizados amplia a confiabilidade das análises, considerando apenas as informações que sejam da escolha do usuário.

Os resultados das simulações permitem identificar que os impactos financeiros são significativos, independentemente do setor da empresa, e que as organizações devem estar financeiramente preparadas caso ocorra um ciberataque. O *Phishing* é o tipo de ataque que se mostrou mais custoso para uma organização, devido aos seus múltiplos impactos técnicos que podem interferir no funcionamento da empresa. O setor de saúde apresenta maior variação no valor mínimo do custo médio, indicando uma possível rapidez na solução dos ataques. Assim, todas as informações destacam um custo elevado resultante dos ataques, enfatizando a importância de estratégias de mitigação e preparação eficazes.

Como trabalhos futuros, a solução SIM-Ciber concede espaço para a verificação dos pesos atribuídos no cálculo das *Notas*, comparando o impacto nos resultados finais. Também pode ser expandida para incluir uma coleta mais abrangente de dados sobre as empresas examinadas (por exemplo, número de funcionários e ativos), bem como a inclusão de dados de mais setores e tipos de ataques. Simulações mais robustas podem ser desenvolvidas, utilizando modelos baseados em IA para quantificar riscos e impactos com diferentes níveis de granularidade, assim como para comparar com resultados das soluções já existentes. Além disso, interfaces intuitivas podem ser desenvolvidas para que o SIM-Ciber seja utilizado por usuários com diferentes níveis técnicos e também para a validação experimental, comparando com a realidade. Portanto, o trabalho proposto permite diversos caminhos de exploração e de desenvolvimento pela comunidade.

Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) sob processo número 2020/05152-7, do projeto PROFISSA, e faz parte do processo CNPq 316662/2021-6. Também faz parte do INCT de Redes Inteligentes de Comunicações e Internet das Coisas Inteligentes (ICoNIoT), financiado pelo CNPq (proc. 405940/2022-0) e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) Código Financeiro 88887.954253/2024-00.

Referências

Ahmed, M., Panda, S., Xenakis, C., and Panaousis, E. (2022). MITRE ATTCK-Driven Cyber Risk Assessment. In *17th International Conference on Availability, Reliability and Security (ARES)*, New York, NY, USA. Association for Computing Machinery.

Alawida, M., Omolara, A. E., Abiodun, O. I., and Al-Rajab, M. (2022). A deeper look into cybersecurity issues in the wake of covid-19: A survey. *Journal of King Saud University-Computer and Information Sciences*, 34(10):8176–8206.

Berger, C. E., de Boer, H. H., and van Wijk, M. (2020). Use of Bayes' Theorem in Data Analysis and Interpretation. In *Statistics and probability in forensic anthropology*, pages 125–135. Elsevier.

Chockalingam, S., Pieters, W., Teixeira, A., and van Gelder, P. (2017). Bayesian Network Models in Cyber Security: A Systematic Review. In *22nd Nordic Conference*, pages 105–122, Tartu, Estonia. Springer.

Engström, V. and Lagerström, R. (2022). Two Decades of Cyberattack Simulations: A Systematic Literature Review. *Computers Security*, 116:102681.

Ferreira, L., Silva, D. C., and Itzazelaia, M. U. (2023). Recommender Systems in Cybersecurity. *Knowledge and Information Systems*, 65(12):5523–5559.

Fortinet (2021). Retail Cybersecurity Statistics Not To Be Ignored. Fortinet, `https://www.fortinet.com/solutions/industries/retail/retail-cybersecurity-statistics`.

Franco, M. F., Granville, L. Z., and Stiller, B. (2023a). CyberTEA: a Technical and Economic Approach for Cybersecurity Planning and Investment. In *36th IEEE/IFIP Network Operations and Management Symposium (NOMS 2023)*, pages 1–6, Miami, USA.

Franco, M. F., Künzler, F., von der Assen, J., Feng, C., and Stiller, B. (2024). RCVaR: an Economic Approach to Estimate Cyberattacks Costs using Data from Industry Reports. *Computers & Security*, page 103737.

Franco, M. F., Lacerda, F. M., and Stiller, B. (2022a). A Framework for the Planning and Management of Cybersecurity Projects in Small and Medium-sized Enterprises. *Revista de Gestão e Projetos*, 13(3):1–25.

Franco, M. F., Omlin, C., Kamer, O., Scheid, E. J., and Stiller, B. (2023b). SECAdvisor: a Tool for Cybersecurity Planning using Economic Models.

Franco, M. F., Sula, E., Huertas, A., Scheid, E. J., Granville, L. Z., and Stiller, B. (2022b). SecRiskAI: A Machine Learning-Based Approach for Cybersecurity Risk Prediction in Businesses. In *2022 IEEE 24th Conference on Business Informatics (CBI)*, volume 1, pages 1–10, Amsterdam, Netherlands. IEEE.

Gordon, L. A., Loeb, M. P., and Zhou, L. (2021). Information Segmentation and Investing in Cybersecurity. *Journal of Information Security*, 12:115–136.

Gore, R., Padilla, J., and Diallo, S. (2017). Markov chain modeling of cyber threats. *The Journal of Defense Modeling and Simulation*, 14(3):233–244.

Havakhor, T., Rahman, M. S., and Zhang, T. (2020). Cybersecurity investments and the cost of capital. *SSRN Electronic Journal*, pages 1–48.

Huang, K., Wang, X., Wei, W., and Madnick, S. (2023). The Devastating Business Impacts of a Cyber Breach. `https://hbr.org/2023/05/the-devastating-business-impacts-of-a-cyber-breach`.

IBM (2023). Cost of a Data Breach Report 2023. https://www.ibm.com/downloads/cas/E3G5JMBP.

Jacobs, J., Romanosky, S., Suciu, O., Edwards, B., and Sarabi, A. (2023). Enhancing Vulnerability Prioritization: Data-Driven Exploit Predictions with Community-Driven Insights. In *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW 2023)*, pages 194–206, Delft, Netherlands. IEEE.

Jawad, A. and Jaskolka, J. (2021). Modeling and Simulation Approaches for Cybersecurity Impact Analysis: State-of-the-Art. In *Annual Modeling and Simulation Conference (ANNSIM)*, pages 1–12, Fairfax, USA.

Kavak, H., Padilla, J. J., Vernon-Bido, D., Diallo, S. Y., Gore, R., and Shetty, S. (2021). Simulation for Cybersecurity: State of the Art and Future Directions. *Journal of Cybersecurity*, 7(1):tyab005.

Kia, A. N., Murphy, F., Sheehan, B., and Shannon, D. (2024). A cyber risk prediction model using common vulnerabilities and exposures. *Expert Systems with Applications*, 237:121599.

Kianpour, M., Kowalski, S. J., and Øverby, H. (2021). Systematically Understanding Cybersecurity Economics: A Survey. *Sustainability*, 13(24):13677.

Microsoft (2022). DDoS Attack Trends and Insights. https://www.microsoft.com/en-us/security/blog/2023/02/21/2022-in-review-ddos-attack-trends-and-insights/.

Roldán-Molina, G., Almache-Cueva, M., Silva-Rabadão, C., Yevseyeva, I., and Basto-Fernandes, V. (2017). A Comparison of Cybersecurity Risk Analysis Tools. *Procedia Computer Science*, 121:568–575.

Snider, K. L., Shandler, R., Zandani, S., and Canetti, D. (2021). Cyberattacks, Cyber Threats, and Attitudes Toward Cybersecurity Policies. *Journal of Cybersecurity*, 7(1):tyab019.

SonicWall (2023). 2023 SonicWall Cyber Threat Report. https://www.sonicwall.com/resources/white-papers/2023-sonicwall-cyber-threat-report/.

Sophos (2021). The State of Ransomware 2021. https://assets.sophos.com/X24WTUEQ/at/k4qjqs73jk9256hffhqsmf/sophos-state-of-ransomware-2021-wp.pdf?cmp=120469.

Sophos (2023). The State of Ransomware in Financial Services 2023. https://news.sophos.com/en-us/2023/07/13/the-state-of-ransomware-in-financial-services-2023/.

Subroto, A. and Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning. *Journal of Big Data*, 6(50):1–19.

Verizon (2023). 2023 Data Breach Investigations Report. https://www.verizon.com/business/en-gb/resources/reports/dbir/.

Yamin, M. M. and Katt, B. (2022). Modeling and Executing Cyber Security Exercise Scenarios in Cyber Ranges. *Computers Security*, 116:102635.

Zimperium (2023). 2023 Global Mobile Threat Report. https://www.zimperium.com/global-mobile-threat-report/.

Todos os links foram acessados e estavam válidos em Agosto de 2024.