UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

JOÃO DAVI MARTINS NUNES

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

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação

Orientador: Prof. Dr. Luciano Paschoal Gaspary
Co-orientador: Dr. Muriel Figueredo Franco

Porto Alegre
2024

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. Carlos André Bulhões Mendes
Vice-Reitora: Profª. Patricia Helena Lucas Pranke
Pró-Reitora de Graduação: Profª. Cíntia Inês Boll
Diretora do Instituto de Informática: Profª. Carla Maria Dal Sasso Freitas
Coordenador do Curso de Ciência de Computação: Prof. Marcelo Walter
Bibliotecário-chefe do Instituto de Informática: Alexsander Borges Ribeiro

“Tudo posso naquele que me fortalece.”

— PAULO (FILIPENSES 4.13)

AGRADECIMENTOS

Primeiramente, quero agradecer a Jesus! Durante a montanha-russa da minha vida, Ele me sustentou e possibilitou todas as coisas. Jesus me deu um sentido de vida e me amou incondicionalmente. Sem Ele, nada disso seria possível.

Agradeço também à minha família - Marcos e Jucimere, Jennifer e Gean (e Neemias) - por não desistirem de mim. Por estarem ao meu lado nos momentos bons e ruins, por me sustentarem durante este tempo, por não se cansarem de me ajudar durante minhas crises de saúde, por sempre desejarem o melhor para mim nas áreas pessoal, profissional e espiritual. Amo vocês!

Agradecimento ao Lisandro Granville, meu orientador, por ter me escolhido e acreditado em mim. Obrigado também Muriel Franco e Éder Scheid por serem motivações e referências, por estarem sempre dispostos a transmitir conhecimento, a iluminar e a servirem como inspiração na trilha acadêmica para mim, por serem influências positivas no meu desenvolvimento e desempenho profissional e me ensinarem a almejar grandes patamares.

Obrigado Marcos Landi, Felipe Didio, Anderson Rieger, Afaqii, Cleiton, Diego, Matheus, Marcelo, Arthur e Nicolas pelos bons tempos juntos na faculdade. Ao pessoal da UFSM (Rodrigo Amaral, José Viriato, Gabriel Righi e Lucas Micol), saudades dos estudos coletivos, de resenhas, de parcerias de trabalho até altas horas, de um tempo leve na universidade e pelo tempo juntos.

Não esquecendo de meus discipuladores Felipe Mattos e Enzo Brocker, e também da Carolina Leal e Patrícia Oliveira: amo vocês! Um abraço ao meus demais amigos. Obrigado por serem minha paciência, meu descanso, minhas risadas, minhas memórias.

SIM-Ciber: Probabilistic Simulations for Risks and Impacts Quantification of Cyberattacks Using Statisticals Reports

ABSTRACT

The evolution of technologies, the reliance on digital devices and the growing circulation of data on the Internet all contribute to linking human and business activities, making them vulnerable to what can happen on the network. This interconnectivity increases cyber risks and also the attractiveness of cyber attacks, making it essential to understand these risks and their potential impacts from a technical and economic perspective. In this context, this paper proposes SIM-Ciber, a solution for simulating technical and financial risks and impacts on companies. SIM-Ciber is based on cybersecurity reports and statistics from reputable companies (e.g. consultancies and service providers) and runs probabilistic and simulation methods (e.g. Monte Carlo Method and Bayes Theorem) to understand the risks and impacts of cyberattacks on companies from different locations and sectors. The viability 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. Future work is also proposed that could expand the solution's capacity through improvements and the cooperation of organizations in providing recent and relevant data.

Keywords: Cybersecurity, Cybersecurity Economics, Monte Carlo, Bayes Theorem, Risks Quantification.

RESUMO

A evolução das tecnologias, a dependência em dispositivos digitais e a crescente circulação de dados na Internet contribuem na vinculação das atividades humanas e empresariais, tornando-os vulneráveis ao que pode acontecer na rede. Essa interconectividade aumenta os riscos cibernéticos e também o atrativo pela realização de ciberataques, tornando, pois, essencial a compreensão desses riscos e de seus potenciais impactos a partir de uma perspectiva técnica e econômica. Neste contexto, este trabalho propõe o SIM-Ciber, uma solução para simulação de riscos e impactos técnicos e financeiros em empresas. O SIM-Ciber baseia-se em relatórios e estatísticas de cibersegurança de empresas reputadas (e.g., consultorias e provedores de serviços) e executa métodos probabilísticos e de simulação (e.g., Método de Monte Carlo e Teorema de Bayes) para compreender os riscos e impactos de ciberataques em empresas de diferentes localidades 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. Igualmente, propõem-se trabalhos futuros que possam ampliar a capacidade da solução através de melhorias e pela cooperação das organizações no fornecimento de dados recentes e relevantes.

Palavras-chave: Cibersegurança. Economia de Cibersegurança. Monte Carlo. Teorema de Bayes. Quantificação de Riscos.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas

CVE Common Vulnerabilities and Exposures

DDoS Distributed Denial of Service

DNS Domain Name System

EC Empresa de Consultoria

EPSS Exploit Prediction Scoring System

HD Hard Disk

IA Inteligência Artificial

IP Internet Protocol

MC Método de Monte Carlo

MitM Man-in-the-Middle

ML Machine Learning

OSI Open Systems Interconnection

RAM Random Access Memory

RNP Rede Nacional de Ensino e Pesquisa

SMTP Simple Mail Transfer Protocol

SQL Structured Query Language

TB Teorema de Bayes

TCP Transmission Control Protocol

TI Tecnologia da Informação

TP Total com Peso

TS Total com Soma

SUMÁRIO

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, é crucial para um bom planejamento e gestão de cibersegurança (FRANCO; GRANVILLE; STILLER, 2023).

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; 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; PADILLA; DIALLO, 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; 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; KOWALSKI; ØVERBY, 2021).

É fundamental haver abordagens que permitam investigar a natureza e o comportamento de ciberataques, desde a sua motivação, vulnerabilidades, estratégias defensivas e também potenciais impactos socioeconômicos (FRANCO; LACERDA; STILLER, 2022). Tais abordagens podem ser baseadas em métricas, modelos e ferramentas existentes para análise de riscos (ROLDáN-MOLINA et al., 2017) e investimentos (GORDON; LOEB; ZHOU, 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 trabalho, 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 (*e.g.*, 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 trabalho está estruturado na seguinte forma: o Capítulo 2 aborda a teoria e definição dos conceitos utilizados neste trabalho; o Capítulo 3 discorre sobre os trabalhos relacionados; o Capítulo 4 introduz e detalha a solução *SIM-Ciber*, explicando seus componentes e funcionamento; já o Capítulo 5 foca na avaliação da metodologia proposta, descrevendo os testes utilizados e discutindo os resultados obtidos, e por fim, no Capítulo 6, apresentamos as conclusões e os trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo é focado na descrição de conceitos e tecnologias utilizadas na solução *SIM-Ciber* que são relevantes para a compreensão completa deste trabalho. Tais conceitos que serão relatados são os riscos dos ciberataques, assim como os métodos probabilísticos Monte Carlo e Teorema de Bayes.

2.1 Ciberataques

O amplo uso da Internet, além do constante tráfego e grande armazenamento de dados relevantes, confere espaço para atores mal-intencionados se empenharem na exploração de vulnerabilidades de sistemas, sejam por motivos políticos, por ganhos financeiros ou simplesmente por diversão (VARONIS, 2021). Comportamentos de funcionários e composições de dispositivos em empresas podem torná-las mais suscetíveis a ciberataques, portanto é imprescindível a compreensão de atuação do invasor.

Cada tipo de ataque possui suas características únicas e compreender como os ciberataques são aplicados no processo de invasão, e também como atuam em conjunto, é importante na atuação preventiva no contexto de cibersegurança. Ataques podem atuar nas fases iniciais, outros no meio e outros no fim, englobando todas as informações obtidas para fomentar o objetivo do invasor e alcançar o sucesso final. Por conseguinte, deve-se compreender as facetas e as sutilezas daquilo que podem compor a invasão cibernética, para estar preparado perante as dificuldades a serem enfrentadas.

Na atualidade, diversos outros tipos de ciberataques e variantes são criados, de maneira que o comportamento da técnica de invasão varie por dispositivo, por nível de acesso requisitado e por esforço empenhado pelo atacante. Alguns conhecidos são a injeção SQL, o *Jailbreak* e o ataque *Man-in-the-Middle* (MitM), porém nesse trabalho será discutido sobre o *malware*, *phishing* e *DDoS*. Será discutido sobre suas definições, modos de atuação e impactos gerados, assim como demonstrações de casos reais.

2.1.1 Malware

Também conhecido por código malicioso, o *malware* é um *software* que foi deliberadamente projetado pelo invasor com intenção prejudicial (GANDOTRA; BANSAL;

SOFAT, 2014) aos usuários, organizações e sistemas de computadores e de telecomunicações. Esse tipo de ataque se beneficia de softwares e/ou sistemas defasados, falta de medidas de proteção (*e.g.*, antivírus e *firewall*), erros nas políticas de controle de acesso e principalmente na desatenção dos usuários, nos quais realizam *downloads* indevidos por meio de *links* maliciosos. Os impactos podem abranger a coleta de senhas e informações confidenciais do usuário ou empresa, criptografia de arquivos e softwares, corrupção do sistema operacional e também perda de desempenho (máquina e/ou rede). Portanto, o *malware* cria uma ameaça à disponibilidade da Internet, à integridade e à privacidade dos usuários.

Dependendo da finalidade e forma de execução, o *malware* pode ser representado de diferentes formas, que variam na maneira de propagação e ação. Alguns dos exemplos existentes são: o *ransomware*, que é um código elaborado para restringir o acesso a um sistema ou a dados até que seja feito um pagamento de resgate, de forma a satisfazer o desejo do invasor (OZ et al., 2022); o *worm*, que é um código malicioso capaz de se replicar e de se autopropagar utilizando meios de comunicação, como o e-mail e o protocolos de rede TCP/IP (BELLAMY; HUTCHINSON; WELLS, 2007); e, por fim, o *Trojan*, o qual se disfarça por *software* legítimo, porém é maléfico ao dispositivo e ao usuário (FAGHANI; NUGYEN, 2017). Portanto, percebe-se que o *malware* possui diversas ferramentas para gerar danos técnicos e perdas financeiras (IBM, 2023), compelindo a preparação de uma segurança efetiva por parte das empresas e usuários.

Em 2017, ocorreu um dos maiores casos reais experienciados de *malware*. O *ransomware* chamado de *WannaCry* (Figura 2.1) foi projetado para usufruir de uma vulnerabilidade do sistema operacional *Windows*, de maneira a criptografar dados e requerir da vítima um valor de resgate pago em *bitcoin* (KASPERSKY, 2024). O ataque atingiu diversos países e cerca de 230 mil computadores foram afetados, além da estimativa de perda financeira por volta de quatro bilhões de dólares ao redor do mundo. Um exemplo claro da força desse tipo de ataque e da importância da cibersegurança no cenário global.

2.1.2 Phishing

Por definição (ALABDAN, 2020), *phishing* é uma técnica de engenharia social que, usando várias metodologias, visa influenciar o alvo do ataque a revelar informações pessoais, tais como endereço de e-mail e senha, ou até informações financeiras. O nome tem origem na palavra em inglês *fishing* (no português, pesca), que semelhantemente usa

Figura 2.1: Ransomware WannaCry

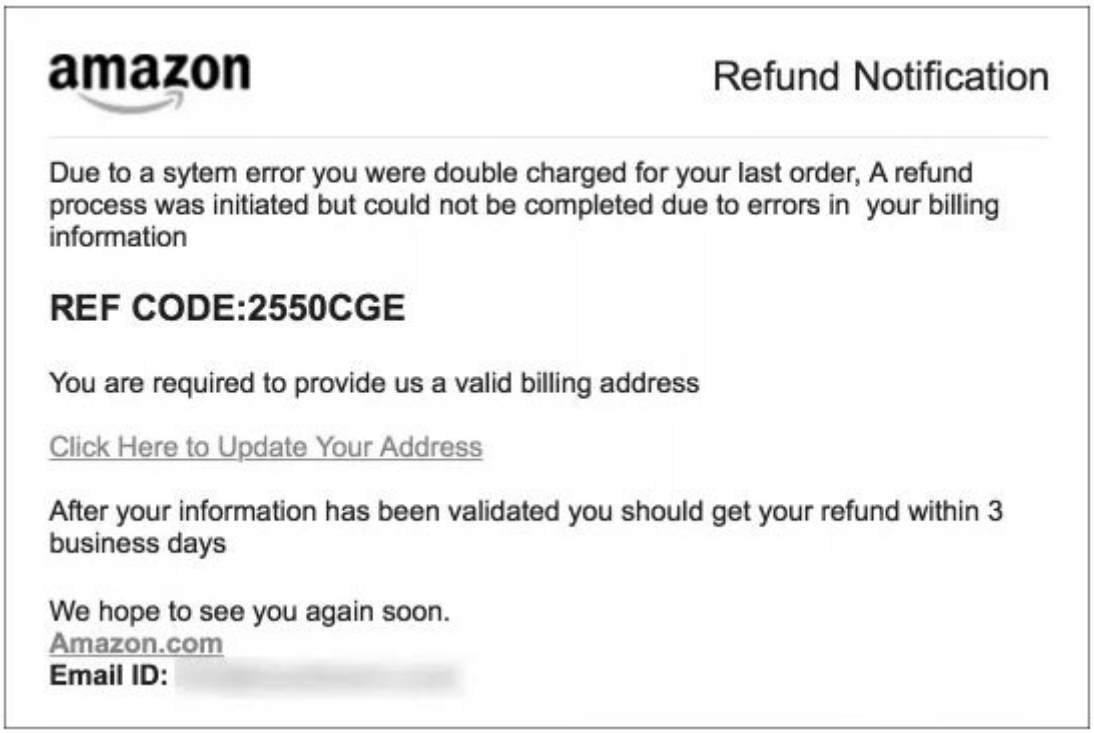
Fonte: (BBC, 2017)

uma isca para atrair a vítima. A falta de atenção dos usuários é um dos maiores motivos para um prejudicial desfecho (VERIZON, 2023), além de poder ser motivado por razões financeiras ou pessoais. Esse tipo de ataque é usado como vetor inicial para outros tipos de ataques, onde podem explorar a rede e os computadores através do acesso não autorizado.

O modo de atração pode variar, assim como o nível de informação prévia que o invasor possui antes de realizar o ataque e o alvo desejado. Dentre os tipos conhecidos, podemos evidenciar o *spear phishing* - atraindo o usuário através da necessidade de ação imediata por parte da vítima e a captura de informações por um canal (por exemplo, um portal falso) (AKAMAI, 2024) - e o *whaling* - um ataque focalizado em alvos executivos corporativos de alto nível através de e-mails, mensagens de texto ou chamadas telefônicas fraudulentas, para que assim seja autorizado grandes pagamentos ou a coleta de informações confidenciais e valiosas (IBM, 2024a). Como atua na fase inicial, o ataque de *phishing* pode contar com técnicas de outros tipos de ciberataques, como o *malware* (WORLD, 2020), em outros momentos da invasão. Por ser intrusivo e operar visando obter dados sensíveis das empresas e dos usuários (como, por exemplo, credenciais), esse tipo de ataque concede espaço para exploração e agravação dos possíveis impactos, além de gerar perda de produtividade nos dispositivos das vítimas.

Sendo o e-mail uma das formas mais comuns de meio de comunicação do *phishing*, como consta na Figura 2.2, é possível compreendermos a sua atuação. Através de uma isca, o atacante se passa pela empresa *Amazon* informando ao usuário sobre o estado atual

Figura 2.2: Exemplo de Phishing



Fonte: (UNIVERSITY, 2024)

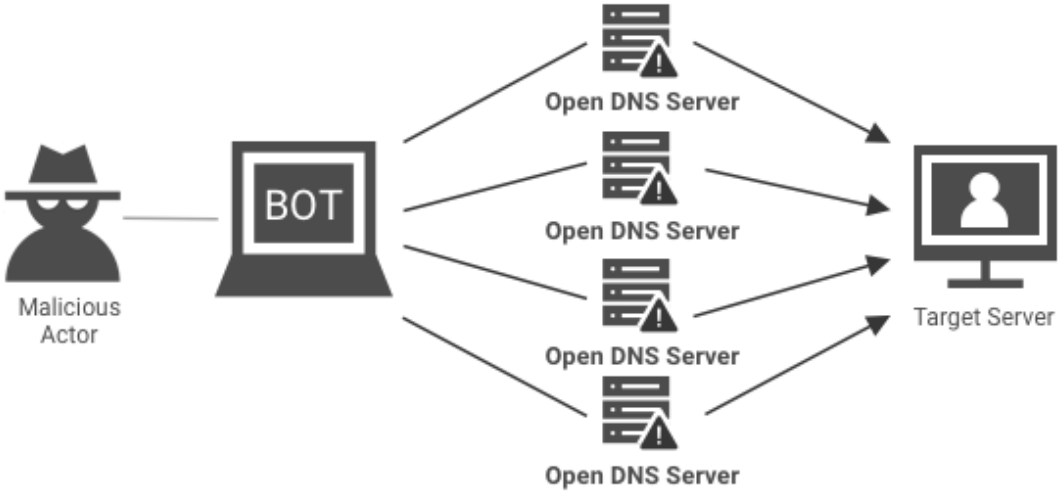
de sua fatura. A estrutura do e-mail, o logotipo da empresa original, as informações contidas e os links falsos trabalham em conjunto para convencer a vítima a cometer um erro e conceder informações pessoais ao invasor. Por fim, podemos entender como a engenharia social concede oportunidades de pessoas agirem maleficamente em seu próprio benefício.

2.1.3 DDoS

O *DDoS* (*Distributed Denial of Service* ou Ataque de Negação de Serviço Distribuído, em português) é um método malicioso de ataques simultâneos a partir de múltiplos sistemas a um alvo específico, de forma a tornar os recursos do computador ou da rede indisponíveis e suspendendo serviços conectados na Internet (LI et al., 2023). O *DDoS* envolve muito mais do que o sucesso do ataque, mas sim as questões políticas e de mercado em torno do setor afetado (concorrência), dos interesses globais, do *hacktivismo* e até por vingança. Esse tipo de ataque é altamente prejudicial para as empresas, independentemente do tamanho, pois a parada de serviços implica em diversos impactos, além do financeiro.

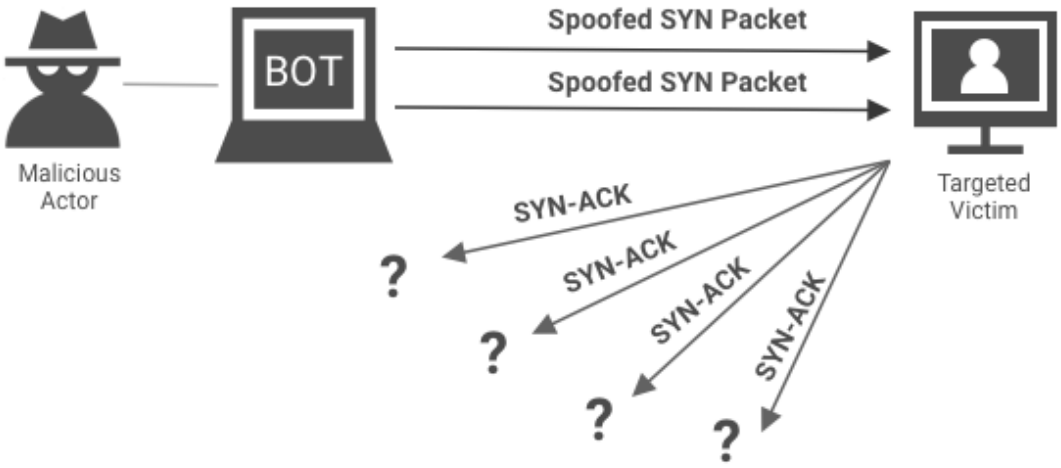
As variantes do *DDoS* podem explorar diferentes camadas do modelo OSI. Exemplos de ataques são o ataque de amplificação de DNS (ver Figura 2.3) e o *SYN Flood* (ver Figura 2.4). Os principais impactos causados são as disrupções de serviços, indisponibilidade dos sistemas e exaustão da rede e do sistema. Camadas de proteção devem estar dispostas na rede, assim como o monitoramento contínuo dela, para evitar que haja espaço para a aplicação dessa técnica de invasão.

Figura 2.3: Ataque de Amplificação de DNS



Fonte: (ONELOGIN, 2024)

Figura 2.4: SYN Flood



Fonte: (ONELOGIN, 2024)

Um dos casos mais conhecidos de *DDoS* foi o ataque à Google em 2017 (ZDNET, 2020). Nesse ataque, foram combinados diversos protocolos para inundar a rede,

através de *DNS Flood* e *SMTP Flood*, com o objetivo de limitar ao máximo o uso dos serviços disponibilizados. Estimou-se um tamanho de 2,54 Tbps, o maior tráfego percebido por um ataque de *DDoS* até então e, devido à complexidade, os impactos financeiros se tornaram extremamente difíceis de serem avaliados. Portanto, deve haver preparação perante a hostilidade existente na rede, a fim de atenuar os riscos e os impactos técnicos e financeiros causados por *DDoS*.

2.2 Métodos Probabilísticos

Métodos probabilísticos, ou também conhecidos como métodos estatísticos, se referem aos procedimentos matemáticos aplicados para obtenção de novos conhecimentos a partir de probabilidades já existentes. Em geral, se baseiam em dados assumidos como verdades, de forma a permitir a inferência de novas informações. Esses métodos servem como auxílio na predição de acontecimentos, com base na probabilidade de eventos adjacentes acontecerem, permitindo assim uma melhor previsão do evento principal a ser descoberto, de forma mais simplificada. Portanto, os métodos probabilísticos ajudam na compreensão do mundo, das relações entre eventos e de seus possíveis resultados.

Podemos encontrar na literatura diversas técnicas matemáticas e computacionais que são aplicadas com foco cibernético, tais como Modelos de *Markov* (KHARCHENKO et al., 2022) e Redes Neurais (ASHIKU; DAGLI, 2020). Neste trabalho, será abordado o método de Monte Carlo e do Teorema de Bayes, a fim de compreender suas definições e campos de atuação, assim também entender suas contribuições no âmbito de cibersegurança ao cenário internacional.

2.2.1 Teorema de Bayes

O Teorema de Bayes (TB), também conhecido como Regra de Bayes, é uma importante ferramenta nas áreas de probabilidade e estatística. Foi desenvolvido e nomeado em homenagem ao seu criador, Thomas Bayes, que propôs uma fórmula para determinar probabilidade condicional, ou seja, a priori de um evento, com base em novas informações.

Esse teorema é descrito como a probabilidade posterior de um evento acontecer, dado o conhecimento inicial das evidências e de suas chances serem verdadeiras. De

forma mais simplificada, tendo a informação da probabilidade de B ocorrer dado que A aconteceu e com entendimento dos eventos de A e B acontecerem independentemente, o teorema permite ajustar as expectativas sobre o evento principal. Portanto, para a execução da fórmula do Teorema de Bayes é necessária a obtenção das probabilidades necessárias, expressas na Equação 2.1, onde:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{2.1}$$

- $P(A|B)$ é a probabilidade de A ocorrer dado que B ocorreu (probabilidade condicional)
- $P(B|A)$ é a probabilidade de B ocorrer dado que A ocorreu (probabilidade condicional inversa)
- $P(A)$ é a probabilidade de A ocorrer (probabilidade a priori)
- $P(B)$ é a probabilidade de B ocorrer (probabilidade marginal)

O Teorema de Bayes trabalha em problemas de inferência e permite a tomada de decisões sob incertezas. É um método analítico, ou seja, utiliza de uma estrutura lógica para obter deduções e atualizar probabilidades, e não requer simulações. O TB é sensível a suposições sobre conhecimentos prévios dos eventos adjacentes, tornando-se inviável para conclusões em eventos complexos pois é dificultada a relação entre múltiplos fatores.

A formulação simples do teorema não implica na sua superficialidade. Pelo contrário, tornou-se amplamente usado, visto que a barreira de compreensão é baixa e os resultados convincentes. A sua aplicação compreende diversos campos, como os campos da medicina e o da segurança cibernética. No campo da medicina, um dos exemplos é o uso das abordagens bayesianas para interpretar os resultados de um teste diagnóstico e de um ensaio clínico na enfermagem clínica e a aplicação da prática baseada em evidências (MAI et al., 2023). No campo da segurança cibernética, também podemos entender como a ideia do teorema pode auxiliar o administrador da rede na melhor tomada de decisões, de maneira otimizada (KANDOUSSI et al., 2024). Portanto, o Teorema de Bayes é capaz de gerar informações concisas e precisas em cenários incertos, de forma a aprofundar as informações presentes e auxiliar nas tomadas de decisões futuras.

2.2.2 Método de Monte Carlo

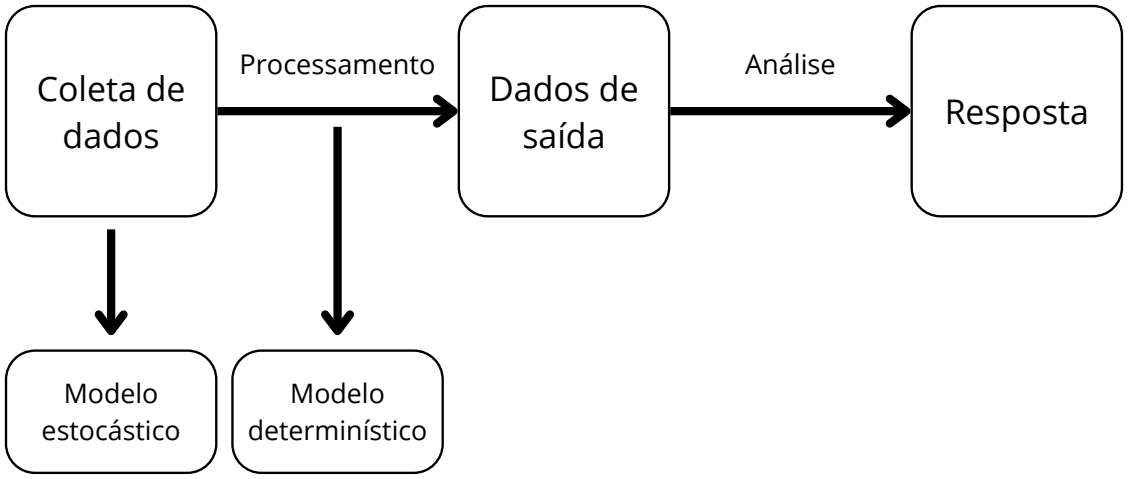
O Método de Monte Carlo (MC) foi proposto durante a Segunda Guerra Mundial por John von Neumann e Stanislaw Ulan com o objetivo de auxiliar nas tomadas de decisões em condições incertas. O nome do método foi dado em homenagem a um casino em Mônaco com este mesmo nome, na qual o elemento "chance"era o centro do modelo, assim como nas jogatinas (*e.g.*, a roleta). Com este método criado, diversos avanços no conhecimento ocorreram, de forma que auxiliam positivamente nosso cotidiano, tais como na Dinâmica Molecular (HAMMOND; LESTER; REYNOLDS, 1994) e na Modelagem Financeira (BOYLE, 1977).

Por definição, o MC é uma técnica matemática que prevê possíveis resultados a partir de eventos incertos, ou seja, na qual não há a convicção e sim a chance de acontecimento. Este método, que é um algoritmo computacional, se baseia no uso de amostras aleatórias repetidas vezes para obter a probabilidade aproximada de ocorrência de um evento. Em suma, é uma investida na simulação dos eventos do passado e na permissão de aplicação dos resultados em situações no futuro. Para a utilização dessa técnica, o algoritmo é baseado em 3 passos, ao menos, indicados abaixo e na Figura 2.5:

1. Definir um modelo preditivo, que aborde todas as variáveis dependentes e independentes do sistema a ser predito, assim como definir os dados de entrada do sistema.
2. Utilizar de dados coletados para especificar as distribuições probabilísticas e seus pesos a serem utilizados nos critérios de avaliação do algoritmo.
3. Executar simulações repetidas vezes, gerando assim valores diversos de variáveis dependentes. Executar até possuir informações suficientes coletadas dessa amostra, permitindo assim uma conclusão.

Pela versatilidade do método, é permitido que seja utilizado largamente nos muitos setores existentes. Como, por exemplo, no setor financeiro - atuando como modelo na tomada de decisões em investimentos (TOBISOVA; SENOVA; ROZENBERG, 2022) -, na agricultura - permitindo a avaliação da viabilidade econômica de sistemas de cultivo (KADIGI et al., 2020) - e na meteorologia - possibilitando tomadas de ações com maior precisão a partir de uma análise das situações do clima e de fenômenos naturais do local (SABERI et al., 2021)-, ou seja, adequando-se a cenários reais e também ao auxílio de Inteligência Artificial (IA) (ALANAZI et al., 2020) para alcançar resultados conclusivos, satisfatórios e qualitativos em diversos problemas.

Figura 2.5: Modelo do Método de Monte Carlo



Fonte: Resultados Originais da Pesquisa

Por fim, esse método é robusto, numérico e de simulações: robusto, pois permite adequação em diversas situações e problemas, do mais simples ao mais complexo, e uma maior confiança no resultado gerado; numérico, pois utiliza de métodos computacionais de aproximação matemática para estimar problemas de difíceis resoluções ao ser humano, além de permitir otimização da solução a partir de um maior conhecimento prévio do problema através dos dados coletados; por fim, de simulações, pois utiliza da modelagem para copiar a realidade e executa múltiplas vezes os códigos, de forma a explorar os diversos cenários possíveis, ao custo de ser potencialmente caro computacionalmente. Portanto, MC se mostra integral na sua proposta, atuando matematicamente para inferir resultados consistentes com base em conhecimentos prévios de eventos complexos e em simulações.

3 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; RAHMAN; ZHANG, 2020). A escassez de recursos e a falta de profissionais especializados intensificam esses desafios, especialmente para pequenas e médias empresas (FRANCO; GRANVILLE; STILLER, 2023). 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; KOWALSKI; ØVERBY, 2021).

Focado nos aspectos econômicos da cibersegurança, (GORDON; LOEB; ZHOU, 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. 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; 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.

Tabela 3.1: 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
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; APRIYANA, 2019)	Predição de riscos e vulnerabilidades usando dados de redes sociais	Não	Não	Sim	Não
SecRiskAI (FRANCO et al., 2022)	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

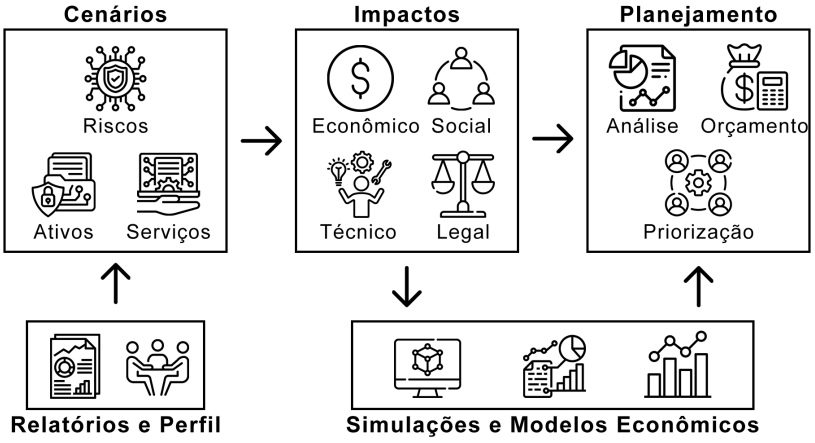
Ainda usando IA, a ferramenta SecRiskAI (FRANCO et al., 2022) determina o nível de exposição de uma empresa a ciberataques. Sua análise de riscos busca correlacionar características da empresa, como, por exemplo, setor de atuação, número de funcionários, e vulnerabilidades conhecidas. A ferramenta não 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 3.1 resume os trabalhos relacionados encontrados na literatura. As soluções atuais focam na análise de riscos e vulnerabilidades utilizando métricas como Common Vulnerabilities and Exposures e Exploit Prediction Scoring System. 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 uti-

lizando dados coletados de redes sociais, incidentes e de interações com especialistas, a literatura ainda carece de trabalhos que utilizem dados de relatórios estatísticos reais de impactos econômicos de cibersegurança.

Com o mesmo viés de trabalhar com cibersegurança e medidas de proteção, o projeto GT-IMPACTO, criado em 2024 e financiado pela Rede Nacional de Ensino e Pesquisa (RNP) através da Chamada Pública de Pesquisa, Desenvolvimento e Inovação da RNP junto ao programa Hackers do Bem (NOBRE et al., 2024). A ideia do projeto (Figura 3.1) é criar uma plataforma que permita a capacitação em cibersegurança, através da modelagem e de simulações de aspectos econômicos. Esta plataforma então permite a criação de cenários de riscos personalizados, a compreensão de riscos e impactos econômicos, planejamento orçamentário e a definição de proteções com alto custo-benefício. Portanto, a geração e o fornecimento de informações relevantes servirão como uma base no planejamento e no investimento em cibersegurança, de maneira a garantir um futuro digital mais seguro.

Figura 3.1: Fluxo do Projeto GT-IMPACTO



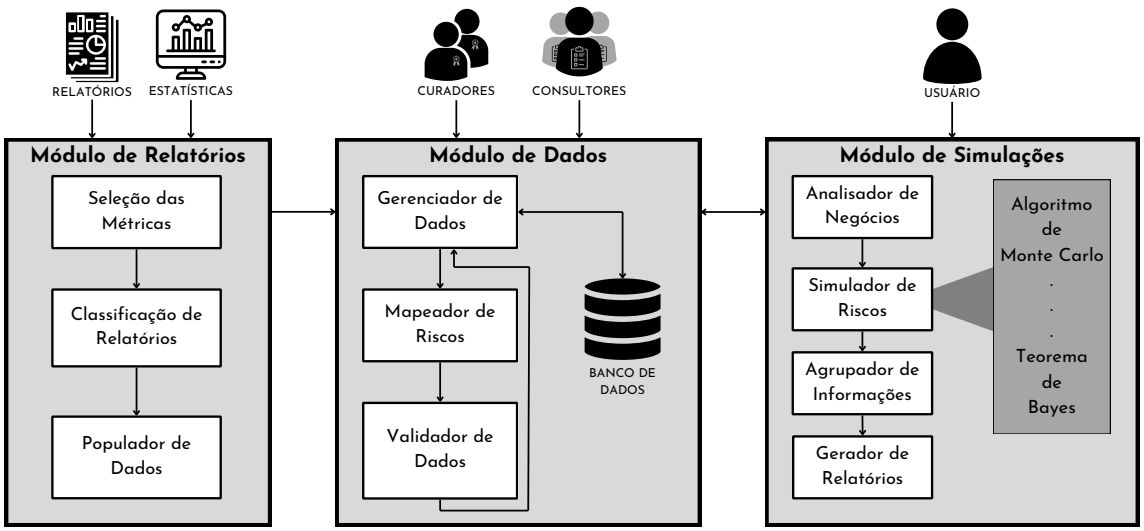
Fonte: (NOBRE et al., 2024)

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; LAGERSTRöM, 2022) para inferir potenciais riscos e impactos econômicos de ciberataques em empresas.

4 SOLUÇÃO *SIM-CIBER*

Este trabalho 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 abordagem da solução proposta é apresentada na Figura 4.1. 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 deste capítulo.

Figura 4.1: Abordagem da Solução *SIM-Ciber*



Fonte: Resultados Originais da Pesquisa

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 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 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 dos ciberataques. As simulações buscam, mediante filtros baseados nas características da empresa, os dados e informações mais relevantes existentes no *Banco de Dados* que se adequem ao perfil da empresa, executando algoritmos e técnicas probabilísticas, como, por exemplo, 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.

4.1 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 as métricas definidas na Tabela 4.1, permitindo assim que as ECs sejam classificadas por *Notas*. Estas

Tabela 4.1: 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

métricas foram definidas previamente através de revisão da literatura e da análise das principais características de distinções das empresas.

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. Essa verificação se baseia na comparação dos valores designados para as ECs através das IAs Generativas contrastando com a situação real da empresa. Nessa etapa, é necessária muita atenção, pois influencia diretamente como as ECs serão processadas pela abordagem e como os dados obtidos serão utilizados na avaliação da requisição do usuário no *Módulo de Simulações*. Tendo essa mentalidade em vista, 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 4.2.

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. Os pesos foram definidos através de uma análise estatística e empírica, na qual foram priorizadas as métricas mais relevantes, tais como a reputação da EC e a metodologia de pesquisa, para ajustar a classificação da EC. Assim, cada EC recebe uma *Nota*, calculada pela média ponderada das métricas e seus respectivos pesos, conforme descrito na Equação 4.2. 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 alta reputação e que sejam muito relevantes para o cenário em que a empresa do usuário está inserida. As ECs são classificadas com base nas *Notas* computadas, sendo definidas como: pouco relevante (TP$<$3), relevante (3$\leq$TP$\leq$5) e muito relevante (TP$>$5)

Tabela 4.2: 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

$$TS = \frac{Rep + Per + Cob + Esc + Abr + Met}{3} \tag{4.1}$$

$$TP = \frac{(Rep*P_{Rep})+(Per*P_{Per})+(Cob*P_{Cob})+(Esc*P_{Esc})+(Abr*P_{Abr})+(Met*P_{Met})}{P_{Rep}+P_{Per}+P_{Cob}+P_{Esc}+P_{Abr}+P_{Met}} * 3 \tag{4.2}$$

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*.

4.2 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

Tabela 4.3: 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 por Ano	(SONICWALL, 2023)

Tabela 4.4: 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)

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 4.4 e 4.3 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 fornece dados para as etapas seguintes.

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 ou 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 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; BOER; WIJK, 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*.

4.3 Módulo de Simulações

Este módulo é responsável por receber a requisição do usuário, analisar a empresa e realizar simulações para ser 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 certeza 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 (*cf.* Tabela 5.1). 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); já o Teorema de Bayes gera 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). 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 permitir uma estimativa dos custos econômicos da empresa, caso o ataque ocorra com sucesso. Por fim, as informações geradas são enviadas adiante na solução, mas também são retornadas ao *Gerenciador de Dados* para serem armazenadas separadamente dos demais dados e disponibilizadas para usos futuros de outros usuários.

O *Agrupador de Informações* é responsável por receber as informações geradas e organizá-las para serem redirecionadas para o *Gerador de Relatórios*. Por último, 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 e os impactos técnicos (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; SILVA; ITZAZELAIA, 2023) e IA Generativa (*e.g.*, ChatGPT e Gemini).

5 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 5.1 e 5.2.

Um protótipo do *SIM-Ciber* foi implementado utilizando *Python 3.11.9*, juntamente com as bibliotecas *Pandas 2.2.1*, *Numpy 1.26.4* e *ReportLab 4.0*. 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 (NUNES, 2024), assim como os relatórios estatísticos utilizados neste trabalho.

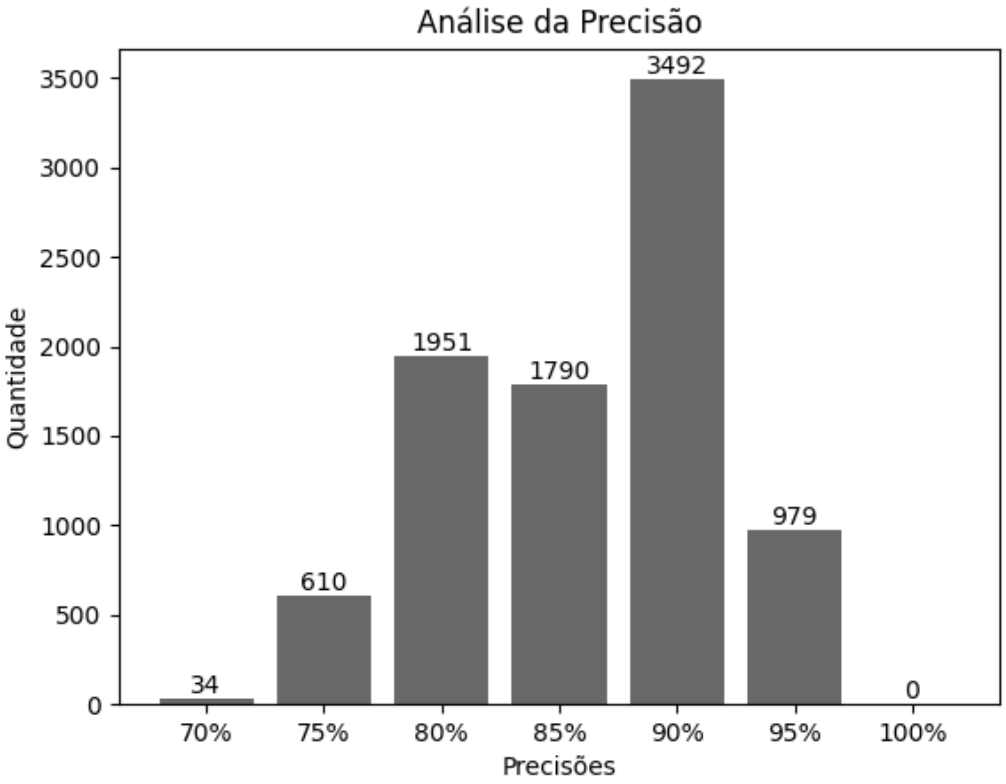
5.1 Classificação dos Relatórios

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 4.1 e 4.2 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 confiabilidade 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 5.1). Portanto, as *Notas* de ECs que estavam nos limites dos Totais (TS e TP) foram alteradas.

Figura 5.1: Análise das Precisões dos Pesos



Fonte: Resultados Originais da Pesquisa

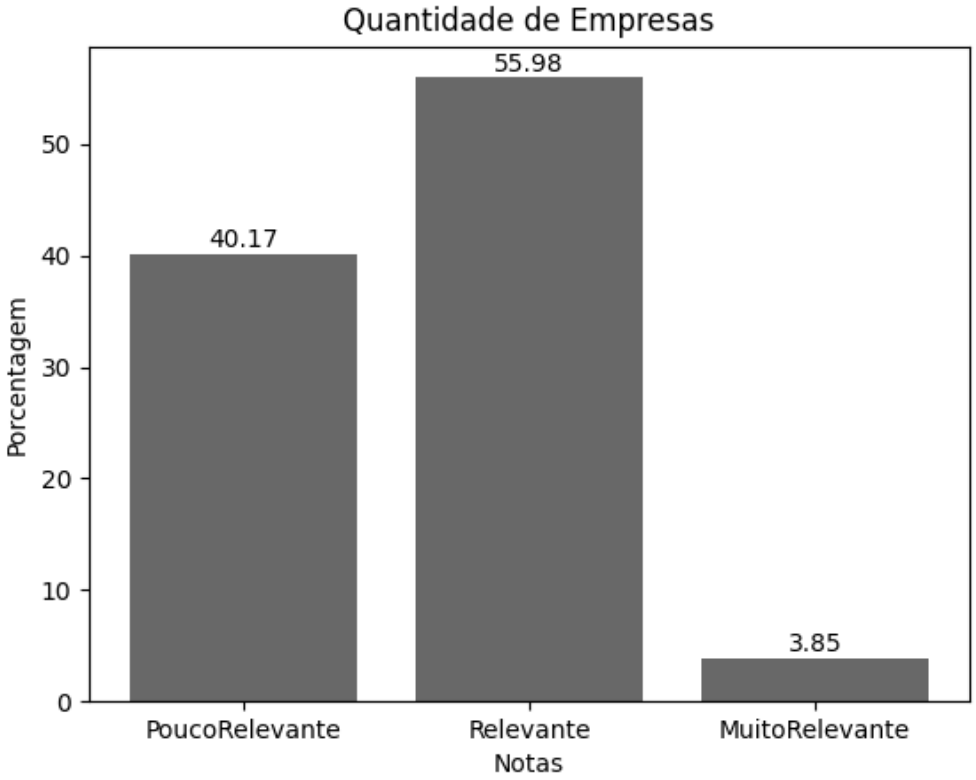
De todos os pesos possíveis, foi selecionado um dos conjuntos de pesos (10, 5, 8, 6, 5 e 10, para cada métrica, respectivamente) com precisão de 95% para realizar trinta rodadas de testes e verificar o comportamento das *Notas* de forma mais específica. Em cada rodada, foram gerados dez mil ECs com métricas 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 5.2;
- 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 (*e.g.*, *Relevante* $\rightarrow$ *Pouco Relevante* e *Muito Relevante* $\rightarrow$ *Relevante*). Esta obser-

vação pôde ser verificada na Figura 5.3, onde "P"significa *Pouco Relevante*; "R", *Relevante* e "M", *Muito Relevante*. A junção das letras indica as transições de *Notas* das ECs (*e.g.*, "PR"= *Pouco Relevante* → *Relevante*);

- A percepção de transição foi ligeiramente suave, como demonstrado nas Figuras 5.5 e 5.4, onde é identificado que a maioria manteve suas *Notas* após a aplicação dos pesos nas métricas e poucos trocaram de *Notas*, indicados por "Falsos Positivos"(*e.g.*, *Notas* melhoraram) e "Falsos Negativos"(*e.g.*, *Notas* pioraram).

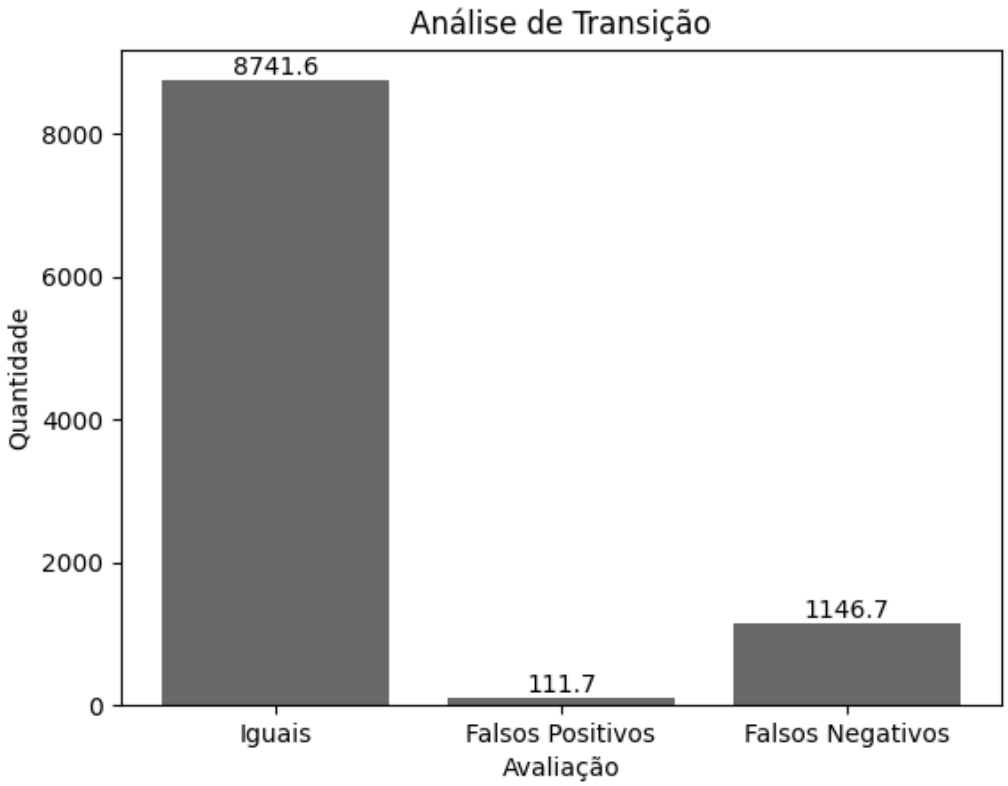
Figura 5.2: Média de ECs por Notas



Fonte: Resultados Originais da Pesquisa

Após os testes realizados, foi decidida 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*.

Figura 5.3: Média das Transições de Notas



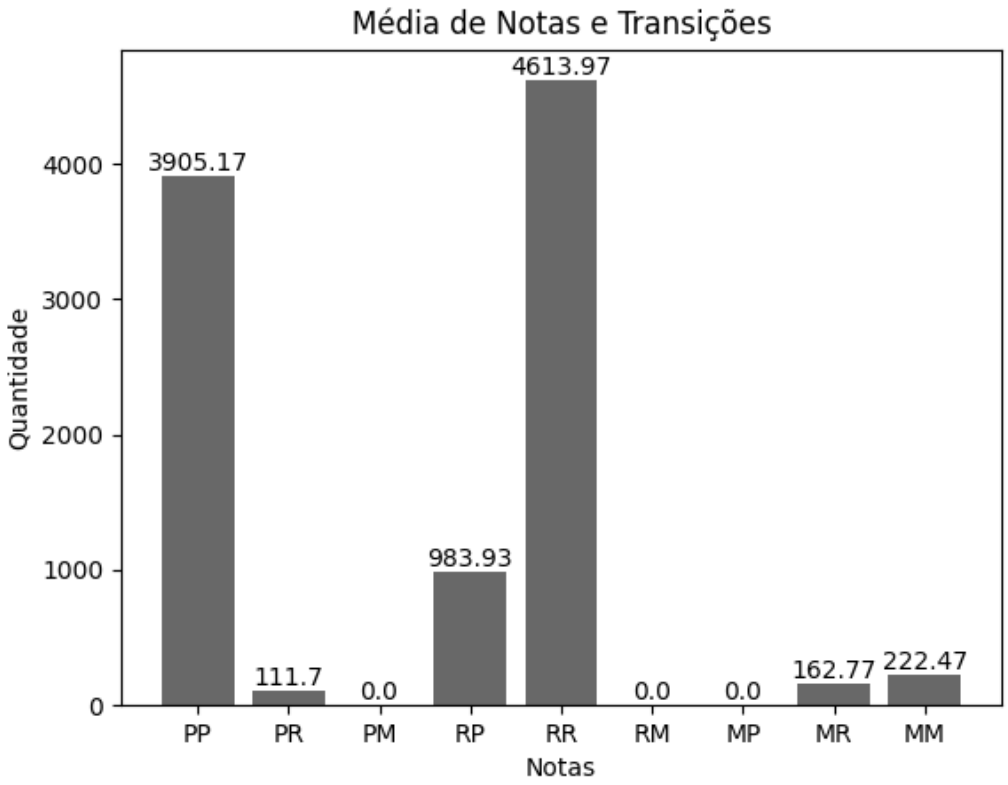
Fonte: Resultados Originais da Pesquisa

5.2 Simulação de Riscos e Impactos

Com os dados prontos, classificados e adicionados ao Banco de Dados, foram realizadas as simulações de riscos e verificada a sua precisão. 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 seção explica como as requisições estão formatadas e também avalia os resultados gerados pelas simulações.

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 536 dados tangíveis (Tabela 4.3) e não tangíveis (Tabela 4.4) oriundos de 51 ECs e relatórios diferentes. Para isso, foi criado um formato padrão de requisições, tornando viável a análise do cenário e a aplicação das simulações.

Figura 5.4: Análise das Transições de Notas



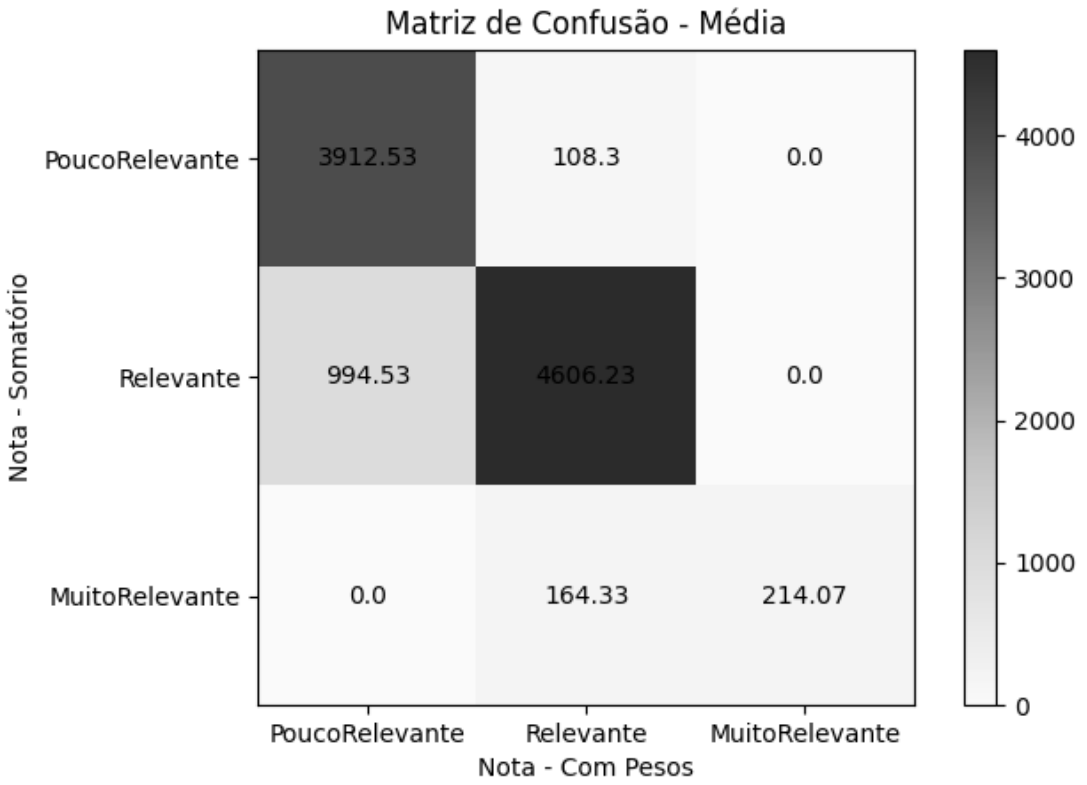
Fonte: Resultados Originais da Pesquisa

5.2.1 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 5.1. 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.

Com o modelo de requisições padronizado, foram gerados 315 exemplos de requisições, com a variação dos diferentes cenários suportados pelo *SIM-Ciber* (exceto as informações extras), servindo como base para as simulações e suas avaliações. Assim,

Figura 5.5: Matriz de Confusão das Transições de Notas



Fonte: Resultados Originais da Pesquisa

Tabela 5.1: 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

a avaliação das simulações também pôde ocorrer e ajudar na análise da qualidade das informações expressas no relatório final.

5.2.2 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 5.2. 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 5.2: 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

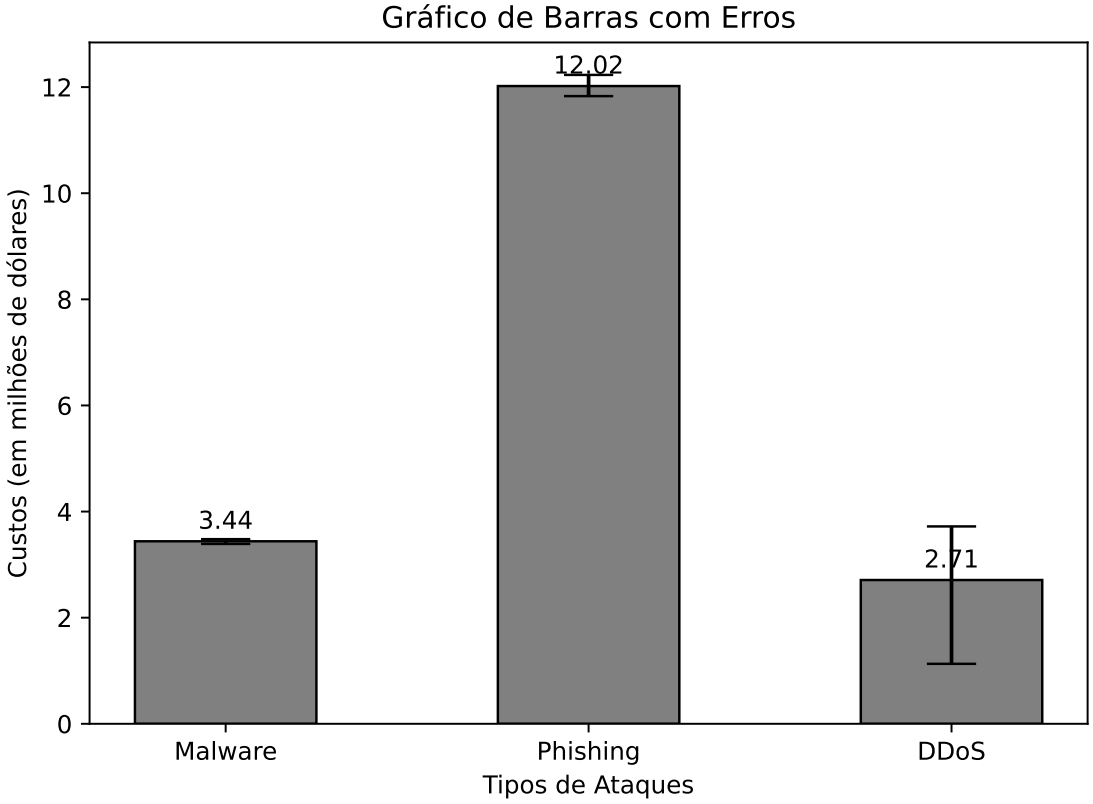
Após simular todas as rodadas com Monte Carlo, é encontrado o custo médio, assim como os custos mínimos e máximos que a empresa teria naquelas condições informados na requisição. Para uma análise comportamental da solução, para cada uma das requisições foram efetuadas cem rodadas de testes, com o objetivo 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).

Analisando as informações por tipo de ataque, demonstrado pela Figura 5.6, 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 5.6: Custos Médios por Tipo de Ataque

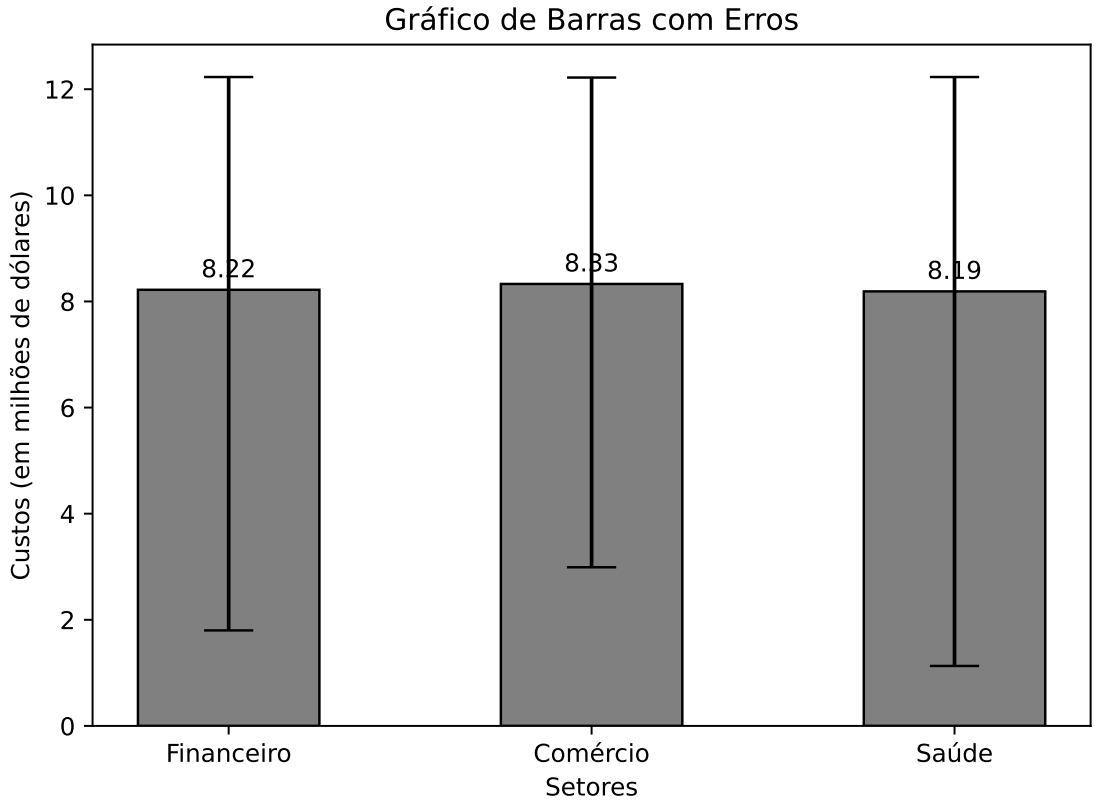


Fonte: Resultados Originais da Pesquisa

Analisando as informações por setor, demonstrado pela Figura 5.7, 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 5.8, 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

Figura 5.7: Custos Médios por Setor



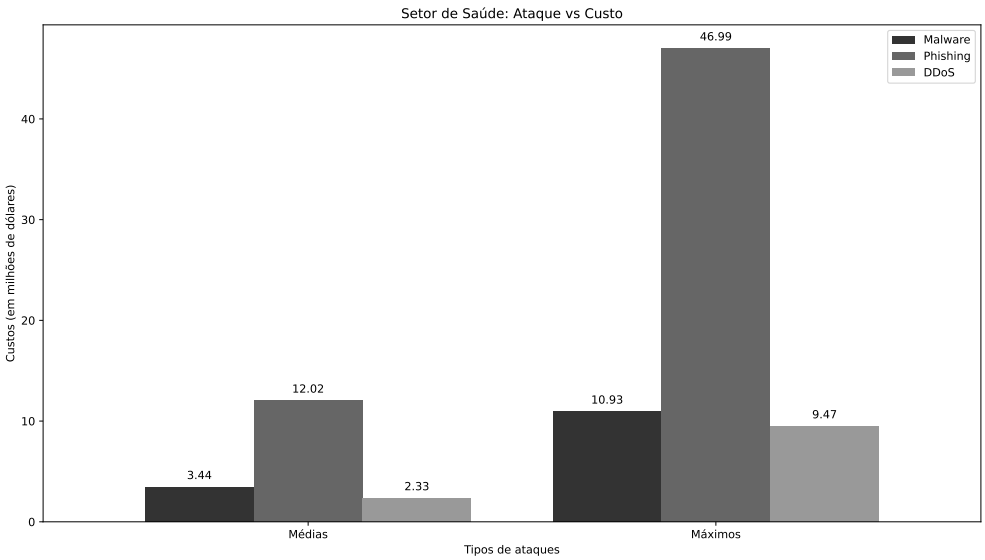
Fonte: Resultados Originais da Pesquisa

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.

5.3 Discussão

Com uma gama de relatórios e estatísticas disponíveis na Internet provenientes dos mais diversos tipos de empresas de setores distintos e com diferentes níveis de confiabilidade, a busca por dados relevantes à proposta do trabalho poderia ser exaustante. A quantidade de dados coletados foi suficiente para se adequar à abordagem da solução, para a geração de resultados e para a lógica implementada no cálculo dos impactos econômicos em cada setor da indústria. A respeito dos dados existentes, muitos outros ainda poderiam ser coletados (como, por exemplo, os tamanhos das empresas e seus ativos) de forma a auxiliar na precisão das informações gerados pela solução *SIM-Ciber*, porém demandaria maior força operacional, tempo despendido na procura e na particularidade,

Figura 5.8: Custos por Tipo de Ataque no Setor de Saúde



Fonte: Resultados Originais da Pesquisa

com um possível risco de pouca variação dos dados encontrados e ao fato de nunca chegar à totalidade de dados alcançáveis na indústria.

A solução proposta neste trabalho introduziu a classificação das Empresas de Consultoria por meio da utilização da Fórmula Total com Peso (Equação 4.2) e da graduação por *Notas*. No entanto, essa classificação pode variar conforme os pesos atribuídos no cálculo utilizando o TP. A decisão de aplicar pesos com 95% de precisão resultou nas *Notas* utilizadas no *SIM-Ciber*, visando contemplar os resultados gerados. Alternativamente, outros pesos poderiam ser aplicados, o que modificaria as avaliações das ECs e, consequentemente, os dados fornecidos por meio das requisições. Um estudo adicional poderia ser realizado para verificar os múltiplos resultados obtidos com diferentes conjuntos de pesos, assim como para analisar a influência das variações de precisão nos pesos e seu impacto nas *Notas* das ECs. Portanto, a solução apresentada considera um ajuste mínimo (diferença de 5% nas Notas ao comparar as Equações 4.1 e 4.2), ou seja, apenas uma pequena correção na classificação dos relatórios das ECs que estavam próximos aos limites das *Notas*.

A decisão dos pesos igualmente pode ter a participação de árbitros. Esse agente ficaria responsável pelos julgamento dos pesos aplicados na Fórmula TP e da conferência dos resultados das *Notas*, de forma a certificar a correta classificação das ECs, com base na realidade. Assim, é gerada mais uma etapa de validação da abordagem, garantindo uma

melhor identificação da relevância das ECs e, consequentemente, auxiliando no processo de seleção dos dados, a partir da requisição do usuário no *Módulo de Simulações*.

Pela perspectiva dos ciberataques existentes na atualidade, a variedade de métodos existentes para invasão (BIJU; GOPAL; PRAKASH, 2019) podem colaborar na estimativa dos possíveis impactos técnicos e econômicos, assim como no planejamento de métodos efetivos de defesa. Neste trabalho, a escolha de *malware*, *phishing* e *DDoS* se deram a respeito de suas reputações perante o âmbito cibernético (MANDIANT, 2024)(PROOFPOINT, 2023)(ARELION, 2024), pela quantidade de ataques realizados ao redor do mundo (SONICWALL, 2023)(BARRACUDA, 2023)(NEXUSGUARD, 2024) e pelos seus diferentes comportamentos no modo operante durante uma invasão (ver Seção 2.1). O vislumbre desses três tipos de ataques satisfez o objetivo proposto no trabalho, de forma a fornecer resultados conclusivos e satisfatórios, ajudando na compreensão dos ciberataques e seus impactos.

A existência de diversos setores na indústria permite entender pelo lado de fora como o mundo corporativo está organizado, porém, cada setor possui sua particularidade e seu nível de seriedade na tomada de decisões de segurança. Para evitar a abrangência de setores e gerar resultados generalistas, a decisão foi de escolher três setores grandes no topo de interesse dos atacantes cibernéticos (ver Figura 5.9) e que tivesse dados possíveis a serem coletados nos relatórios e nas estatísticas. Portanto, a decisão foi acertada e propícia ao trabalho proposto, a qual permitiu a obtenção de diversos dados relevantes para a solução *SIM-Ciber*.

Figura 5.9: Setores da Indústria e suas Parcelas de Ataques por Ano

Share of attacks by industry 2019–2023

Industry	2023	2022	2021	2020	2019
Manufacturing	25.7%	24.8	23.2	17.7	8
Finance and insurance	18.2%	18.9	22.4	23	17
Professional, business and consumer services	15.4%	14.6	12.7	8.7	10
Energy	11.1%	10.7	8.2	11.1	6
Retail and wholesale	10.7%	8.7	7.3	10.2	16
Healthcare	6.3%	5.8	5.1	6.6	3
Government	4.3%	4.8	2.8	7.9	8
Transportation	4.3%	3.9	4	5.1	13
Education	2.8%	7.3	2.8	4	8
Media and telecommunications	1.2%	0.5	2.5	5.7	10

Finance = Financeiro | Retail = Comércio | Healthcare = Saúde
Fonte: (IBM, 2024b)

Para o cálculo dos potenciais custos a partir de cada tipo ataque, esse trabalho utilizou apenas o algoritmo de Monte Carlo como método probabilístico para a obtenção desse valor. Essa decisão foi tomada devido ao funcionamento do algoritmo, que pode ser conferido no Capítulo 2, pois permite atuar em cenários complexos, ou seja, com mais de dois eventos interferindo no cenário e no resultado final, nas mesmas situações em que o *SIM-Ciber* estava disposto e onde foi aplicado. Já o Teorema de Bayes serviu como modelo de organização dos dados coletados, permitindo a execução do MC a partir das tabelas inseridas no Banco de Dados. Assim, constando o algoritmo de Monte Carlo em muitos trabalhos na literatura, tornou-se viável para a sua aplicação na solução proposta.

Por fim, os resultados gerados através das simulações utilizando MC basearam-se nos impactos técnicos dos ataques simulados provenientes das requisições dos usuários (ver Tabela 5.2). Como os ciberataques são utilizados em momentos específicos da invasão e que também podem atuar em conjunto, conclui-se que os custos envolvidos no ataque são a soma dos custos de cada impacto, cada qual na sua chance de ocorrência, em unidade com a análise dos outros tipos de ataques e seus impactos. Os relatórios e estatísticas disponíveis a partir das ECs já tratam de fornecer dados relevantes desses custos pelo ponto de vista das consequências geradas, considerando as possibilidades de exposição da infraestrutura de cibersegurança das empresas. Portanto, os valores coletados e as informações geradas pelas simulações são condizentes com a realidade enfrentada pelas empresas em seus setores da indústria.

6 CONCLUSÕES

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* pode ser facilmente ampliada com a finalização do relatório final a ser gerado na saída do *Módulo de Simulações*, de forma a possuir um artefato relevante ao usuário que fez a requisição. Assim também, *(i)* uma coleta mais abrangente e detalhada de dados sobre as empresas examinadas (por exemplo, número de funcionários e ativos), *(ii)* a inclusão de dados de setores e tipos de ataques diferentes e *(iii)* o estudo dos métodos de defesa e suas efetividades contra os ciberataques que proporcionem maior precisão nos dados a serem simulados e gerados. Combinando todas as ideias e o empreendedorismo em mente, a criação de uma plataforma confiável e orgânica a partir da cooperação na produção de dados recentes nas empresas pode ser um caminho seguro de pesquisa e atuação.

Avaliações mais robustas podem ser desenvolvidas utilizando modelos baseados em IA para quantificar riscos e impactos com diferentes níveis de granularidade. Além disso, interfaces intuitivas podem ser desenvolvidas e expandidas para que o SIM-Ciber seja utilizado por usuários com diferentes níveis técnicos, tal como a proposta do GT-

IMPACTO (NOBRE et al., 2024), fomentando estudos mais precisos na área de cibersegurança e aspectos econômicos, e também contribuir para a aproximação da pesquisa e da indústria. Assim, o assunto que foi tratado neste trabalho é relevante o suficiente para auxiliar as empresas na compreensão dos riscos e impactos gerados pelos ciberataques, entretanto permite diversas outras ideias de crescimento na pesquisa, em prol de maior conhecimento cibernético e de viabilidade de decisões de segurança.

REFERÊNCIAS

AHMED, M. et al. MITRE ATTCK-Driven Cyber Risk Assessment. In: **17th International Conference on Availability, Reliability and Security (ARES)**. New York, NY, USA: Association for Computing Machinery, 2022. ISBN 9781450396707.

AKAMAI. **What Is Spear Phishing?** 2024. <https://www.akamai.com/glossary/what-is-spear-phishing>.

ALABDAN, R. Phishing attacks survey: Types, vectors, and technical approaches. **Future internet**, MDPI, v. 12, n. 10, p. 168, 2020.

ALANAZI, Y. et al. Ai-based monte carlo event generator for electron-proton scattering. **arXiv preprint arXiv:2008.03151**, 2020.

ALAWIDA, M. et al. A deeper look into cybersecurity issues in the wake of covid-19: A survey. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, v. 34, n. 10, p. 8176–8206, 2022.

ARELION. **DDoS Threat Landscape Report**. 2024. <https://www2.arelion.com/l/71232/2024-04-19/cjrfz6/71232/1713531613deuszPTU/Arelion_DDoS_Threat_Landscape_Report_2024_final.pdf>.

ASHIKU, L.; DAGLI, C. Agent based cybersecurity model for business entity risk assessment. In: IEEE. **2020 IEEE International Symposium on Systems Engineering (ISSE)**. [S.l.], 2020. p. 1–6.

BARRACUDA. **2023 Spear-Phishing Trends**. 2023. <https://assets.barracuda.com/assets/docs/dms/2023-spear-phishing-trends.pdf>.

BBC. **Massive Ransomware Infection Hits Computers in 99 Countries**. 2017. <https://www.bbc.com/news/technology-39901382>.

BELLAMY, L.; HUTCHINSON, D.; WELLS, J. User perceptions and acceptance of benevolent worms–a matter of fear? In: IEEE. **6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)**. [S.l.], 2007. p. 29–36.

BERGER, C. E.; BOER, H. H. de; WIJK, M. van. Use of bayes' theorem in data analysis and interpretation. In: **Statistics and probability in forensic anthropology**. [S.l.]: Elsevier, 2020. p. 125–135.

BIJU, J. M.; GOPAL, N.; PRAKASH, A. J. Cyber attacks and its different types. **International Research Journal of Engineering and Technology**, v. 6, n. 3, p. 4849–4852, 2019.

BOYLE, P. P. Options: A monte carlo approach. **Journal of Financial Economics**, v. 4, n. 3, p. 323–338, 1977. ISSN 0304-405X. Available from Internet: <https://www.sciencedirect.com/science/article/pii/0304405X77900058>.

CHOCKALINGAM, S. et al. Bayesian network models in cyber security: A systematic review. In: SPRINGER. **22nd Nordic Conference**. Tartu, Estonia, 2017. p. 105–122.

ENGSTRöM, V.; LAGERSTRöM, R. Two Decades of Cyberattack Simulations: A Systematic Literature Review. **Computers Security**, v. 116, p. 102681, 2022. ISSN 0167-4048. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S0167404822000797>.

FAGHANI, M. R.; NUGYEN, U. T. Modeling the propagation of trojan malware in online social networks. **arXiv preprint arXiv:1708.00969**, 2017.

FERREIRA, L.; SILVA, D. C.; ITZAZELAIA, M. U. Recommender Systems in Cybersecurity. **Knowledge and Information Systems**, Springer, v. 65, n. 12, p. 5523–5559, 2023.

FORTINET. **Retail Cybersecurity Statistics Not To Be Ignored**. 2021. <https://www.fortinet.com/solutions/industries/retail/retail-cybersecurity-statistics>.

FRANCO, M. F.; GRANVILLE, L. Z.; STILLER, B. CyberTEA: a Technical and Economic Approach for Cybersecurity Planning and Investment. In: **36th IEEE/IFIP Network Operations and Management Symposium (NOMS 2023)**. Miami, USA: [s.n.], 2023. p. 1–6.

FRANCO, M. F. et al. Rcvar: an economic approach to estimate cyberattacks costs using data from industry reports. **Computers & Security**, Elsevier, p. 103737, 2024.

FRANCO, M. F.; LACERDA, F. M.; STILLER, B. A Framework for the Planning and Management of Cybersecurity Projects in Small and Medium-sized Enterprises. **Revista de Gestão e Projetos**, UNINOVE, v. 13, n. 3, p. 1–25, nov 2022.

FRANCO, M. F. et al. Secriskai: A machine learning-based approach for cybersecurity risk prediction in businesses. In: IEEE. **2022 IEEE 24th Conference on Business Informatics (CBI)**. Amsterdam, Netherlands, 2022. v. 1, p. 1–10.

GANDOTRA, E.; BANSAL, D.; SOFAT, S. Malware analysis and classification: A survey. **Journal of Information Security**, Scientific Research Publishing, v. 2014, 2014.

GORDON, L. A.; LOEB, M. P.; ZHOU, L. Information Segmentation and Investing in Cybersecurity. **Journal of Information Security**, v. 12, p. 115–136, January 2021.

GORE, R.; PADILLA, J.; DIALLO, S. Markov chain modeling of cyber threats. **The Journal of Defense Modeling and Simulation**, v. 14, n. 3, p. 233–244, 2017.

HAMMOND, B.; LESTER, W.; REYNOLDS, P. **Monte Carlo Methods in Ab Initio Quantum Chemistry**. World Scientific, 1994. (Lecture and Course Notes In Chemistry Series). ISBN 9789810203214. Available from Internet: <https://books.google.com.br/books?id=aydqDQAAQBAJ>.

HAVAKHOR, T.; RAHMAN, M. S.; ZHANG, T. Cybersecurity investments and the cost of capital. **SSRN Electronic Journal**, p. 1–48, 2020.

HUANG, K. et al. **The Devastating Business Impacts of a Cyber Breach**. 2023. <https://hbr.org/2023/05/the-devastating-business-impacts-of-a-cyber-breach>.

IBM. **Cost of a Data Breach Report 2023**. 2023. <https://www.ibm.com/downloads/cas/E3G5JMBP>.

IBM. **O que é Whale Phishing?** 2024. <https://www.ibm.com/br-pt/topics/whale-phishing>.

IBM. **X-Force Threat Intelligence Index 2024**. 2024. <https://www.ibm.com/downloads/cas/46WOEPKL>.

JACOBS, J. et al. Enhancing Vulnerability Prioritization: Data-Driven Exploit Predictions with Community-Driven Insights. In: IEEE. **IEEE European Symposium on Security and Privacy Workshops (EuroS&PW 2023)**. Delft, Netherlands, 2023. p. 194–206.

JAWAD, A.; JASKOLKA, J. Modeling and Simulation Approaches for Cybersecurity Impact Analysis: State-of-the-Art. In: **Annual Modeling and Simulation Conference (ANNSIM)**. Fairfax, USA: [s.n.], 2021. p. 1–12.

KADIGI, I. L. et al. An economic comparison between alternative rice farming systems in tanzania using a monte carlo simulation approach. **Sustainability**, MDPI, v. 12, n. 16, p. 6528, 2020.

KANDOUSSI, E. M. et al. Enhancing cloud security: Harnessing bayesian game theory for a dynamic defense mechanism. **Cluster Computing**, Springer, p. 1–18, 2024.

KASPERSKY. **O que é o Ransomware WannaCry?** 2024. <https://www.kaspersky.com.br/resource-center/threats/ransomware-wannacry>.

KAVAK, H. et al. Simulation for Cybersecurity: State of the Art and Future Directions. **Journal of Cybersecurity**, Oxford University Press, v. 7, n. 1, p. tyab005, 2021.

KHARCHENKO, V. et al. Combining markov and semi-markov modelling for assessing availability and cybersecurity of cloud and iot systems. **Cryptography**, MDPI, v. 6, n. 3, p. 44, 2022.

KIA, A. N. et al. A cyber risk prediction model using common vulnerabilities and exposures. **Expert Systems with Applications**, v. 237, p. 121599, 2024. ISSN 0957-4174.

KIANPOUR, M.; KOWALSKI, S. J.; ØVERBY, H. Systematically understanding cybersecurity economics: A survey. **Sustainability**, MDPI, v. 13, n. 24, p. 13677, 2021.

LI, Q. et al. A comprehensive survey on ddos defense systems: New trends and challenges. **Computer Networks**, Elsevier, p. 109895, 2023.

MAI, H. T. et al. An introduction to bayes' theorem and examples of its application to a diagnostic test and a clinical trial. **Nurse Researcher**, RCN Publishing Company Limited, v. 31, n. 3, 2023.

MANDIANT. **M-Trends 2024 Special Report**. 2024. <https://services.google.com/fh/files/misc/m-trends-2024.pdf>.

MICROSOFT. **DDoS Attack Trends and Insights**. 2022. <https://www.microsoft.com/en-us/security/blog/2023/02/21/2022-in-review-ddos-attack-trends-and-insights/>.

NEXUSGUARD. **DDoS Trend Report 2024**. 2024. <https://www.nexusguard.com/threat-report/ddos-trend-report-2024>.

NOBRE, J. C. et al. **Projeto GT-IMPACTO**. 2024. <https://www.inf.ufrgs.br/gt-impacto/>.

NUNES, J. **Projeto TCC**. 2024. <https://github.com/JoaoDaviMNunes/Projeto_TCC>.

ONELOGIN. **What is a DDoS Attack?** 2024. <https://www.onelogin.com/learn/ddos-attack>.

OZ, H. et al. A survey on ransomware: Evolution, taxonomy, and defense solutions. **ACM Computing Surveys (CSUR)**, ACM New York, NY, v. 54, n. 11s, p. 1–37, 2022.

PROOFPOINT. **2023 State of the Phish**. 2023. <https://www.proofpoint.com/us/resources/webinars/2023-state-phish>.

ROLDáN-MOLINA, G. et al. A Comparison of Cybersecurity Risk Analysis Tools. **Procedia Computer Science**, Elsevier, v. 121, p. 568–575, 2017.

SABERI, N. et al. The use of a monte carlo markov chain method for snow-depth retrievals: A case study based on airborne microwave observations and emission modeling experiments of tundra snow. **IEEE Transactions on Geoscience and Remote Sensing**, 2021.

SNIDER, K. L. et al. Cyberattacks, Cyber Threats, and Attitudes Toward Cybersecurity Policies. **Journal of Cybersecurity**, Oxford University Press, v. 7, n. 1, p. tyab019, 2021.

SONICWALL. **2023 SonicWall Cyber Threat Report**. 2023. <https://www.sonicwall.com/resources/white-papers/2023-sonicwall-cyber-threat-report/>.

SOPHOS. **The State of Ransomware 2021**. 2021. <https://assets.sophos.com/X24WTUEQ/at/k4qjqs73jk9256hffhqsmf/sophos-state-of-ransomware-2021-wp.pdf?cmp=120469>.

SOPHOS. **The State of Ransomware in Financial Services 2023**. 2023. <https://news.sophos.com/en-us/2023/07/13/the-state-of-ransomware-in-financial-services-2023/>.

SUBROTO, A.; APRIYANA, A. Cyber risk prediction through social media big data analytics and statistical machine learning. **Journal of Big Data**, Elsevier, v. 6, n. 50, p. 1–19, 2019.

TOBISOVA, A.; SENOVA, A.; ROZENBERG, R. Model for sustainable financial planning and investment financing using monte carlo method. **sustainability**, MDPI, v. 14, n. 14, p. 8785, 2022.

UNIVERSITY, I. **Real Phishing Emails**. 2024. <https://phishing.iu.edu/stories/index.html>.

VARONIS. **Motivações de um Hacker: Bandeira Vermelha e Prevenção**. 2021. <https://www.varonis.com/pt-br/blog/motivacoes-de-um-hacker-bandeira-vermelha-e-prevencao/#:~:text=Motiva%C3%A7%C3%B5es%20de%20um%20hacker%3A%20Tend%C3%AAncias,grande%20variedade%20de%20outros%20motivos>.

VERIZON. **2023 Data Breach Investigations Report**. 2023. <https://www.verizon.com/business/en-gb/resources/reports/dbir/>.

WORLD, S. **Hedge Fund Closes Down After Cyber Attack**. 2020. <https://www.secureworld.io/industry-news/hedge-fund-closes-after-bec-cyber-attac>.

YAMIN, M. M.; KATT, B. Modeling and Executing Cyber Security Exercise Scenarios in Cyber Ranges. **Computers Security**, v. 116, p. 102635, 2022. ISSN 0167-4048. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S0167404822000347>.

ZDNET. **Google Says It Mitigated a 2.54 Tbps DDoS Attack in 2017, Largest Known to Date**. 2020. <https://www.zdnet.com/article/google-says-it-mitigated-a-2-54-tbps-ddos-attack-in-2017-largest-known-to-date/>.

ZIMPERIUM. **2023 Global Mobile Threat Report**. 2023. <https://www.zimperium.com/global-mobile-threat-report/>.

APÊNDICE A — ARTIGOS CIENTÍFICOS

Com base na solução proposta e dispondo de colaborações, um artigo foi desenvolvido visando o contexto deste trabalho. O artigo científico foi submetido para um evento nacional de grande expressividade visando coletar respostas da comunidade acadêmica. Essa submissão permite validação da metodologia e dos resultados gerados nesse trabalho de Bacharelado, assim como permite melhorias com base nas revisões técnicas a serem recebidas.

A.1 Artigo Aceito – SBSeg 2024

J. D. M. Nunes, M. F. Franco, J. C. Nobre, L. R. Soares, E. J. Scheid, G. Kozenieski, H. Lindemann, L. Z. Granville: SIM-Ciber: Uma Solução Baseada em Simulações Probabilísticas para Quantificação de Riscos e Impactos de Ciberataques Utilizando Relatórios Estatísticos; 24º Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg - 2024), São José dos Campos, São Paulo, Brasil, Setembro 2024.

- **Título:** *SIM-Ciber: Uma Solução Baseada em Simulações Probabilísticas para Quantificação de Riscos e Impactos de Ciberataques Utilizando Relatórios Estatísticos*
- **Contribuição:** Com base em dados coletados de empresas de consultorias, uma solução é proposta para quantificação de riscos e impactos de ciberataques, utilizando métodos probabilísticos (como, por exemplo, o Método de Monte Carlo) a partir da relação entre a localização e o setor da indústria da empresa com os tipos de ciberataques abordados.
- **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 viabili-

dade 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.

- **Status:** Aceito para publicação.
- **Qualis:** A4
- **Conferência:** 24º Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg - 2024)
- **Data:** 16 de setembro – 19 de setembro, 2024
- **Local:** São José dos Campos, São Paulo, Brasil
- **Digital Object Identifier (DOI):** A ser publicado.