Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Ciência da Computação

Métricas de privacidade

Revisão da literatura e aplicação em software científico na área de segurança de redes de computadores

Jessica Yumi Nakano Sato

Monografia Final

mac 499 — Trabalho de Formatura Supervisionado

Supervisor: Prof. Dr. Daniel Macêdo Batista

São Paulo
2023

Agradecimentos

Primeiramente gostaria de agradecer minha mãe que me apoiou durante todo o curso possibilitando que eu me dedicasse completamente aos meus estudos. Quero agradecer também minhas amigas que, não só permitiram que eu desfrutasse uma ótima experiência na faculdade, mas também me ajudaram incontáveis vezes durante todos esses anos. Por fim gostaria de agradecer ao corpo docente e ao instituto pela oportunidade e por desempenharem um papel tão importante na minha formação. Em especial gostaria de agradecer ao professor Daniel que tem me orientado desde 2022 no projeto de iniciação científica e também me orientou durante este trabalho de conclusão de curso, e agradecer ao InterSCity e ao CNPq pelas oportunidades de pesquisa, eventos e financiamento.

Resumo

Jessica Yumi Nakano Sato. **Métricas de privacidade: *Revisão da literatura e aplicação em software científico na área de segurança de redes de computadores***. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

Apesar de regulamentações recentes como a Lei Geral de Proteção de Dados (LGPD) e a *General Data Protection Regulation* (GDPR) terem exigido que desenvolvedores de software passem a se preocupar de forma mais severa com privacidade, recomendações relacionadas a esse assunto existem há muitos anos, por exemplo desde a proposta da P3P (*Platform for Privacy Preferences Project*) para a Internet em 2007. Além disso, embora exista um entendimento, relativamente antigo, de que sistemas computacionais precisam garantir a privacidade do usuário, tem sido difícil encontrar trabalhos que avaliem essas regulamentações em um sistema existente, principalmente porque uma métrica de privacidade pode variar a depender do domínio da aplicação (Por exemplo, a privacidade de um algoritmo de roteamento pode ser calculada como $1/n$, onde n é a quantidade de roteadores por onde um fluxo de rede passa, enquanto que a privacidade de um sistema de aprendizado de máquina para detecção de intrusão pode ser calculada como $a/b - 1$ onde a é o tamanho total do fluxo e b é o tamanho das informações armazenadas para treinamento). Nesse trabalho é apresentado um estudo sobre privacidade e as métricas existentes para medi-la, focando na sua aplicação em software científico na área de segurança de redes de computadores. Além disso, com base nesse estudo, a metodologia Secure Tropos é adaptada para mensurar a privacidade de sistemas de detecção de intrusão. Foi possível concluir que ainda é necessário um maior aprofundamento no estudo de privacidade na área de segurança de redes de computadores por parte dos desenvolvedores de software.

Palavras-chave: Segurança de redes. Detecção de intrusão. IDS. Métricas de privacidade. LGPD. GDPR.

Abstract

Jessica Yumi Nakano Sato. **Privacy Metrics:** ***Literature review and application in scientific software in the field of computer networks security***. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

Although recent regulations such as the *Lei Geral de Proteção de Dados* (LGPD) and General Data Protection Regulation (GDPR) have required software developers to become more concerned about privacy, recommendations related to this subject have existed for many years, for example since the P3P (Platform for Privacy Preferences Project) proposal for the Internet in 2007. Furthermore, although there is a relatively old understanding that computer systems need to guarantee user privacy, it has been difficult to find works that evaluate these regulations in an existing system, mainly because a privacy metric can vary depending on the application domain (For example, the privacy of a routing algorithm can be calculated as $1/n$, where n is the number of routers a network flow passes through, while the privacy of a machine learning system for intrusion detection can be calculated as $a/b - 1$ where a is the total size of the flow and b is the size of the information stored for training). This work presents a study on privacy and the existing metrics to measure it, focusing on its application in scientific software in the field of computer networks security. Moreover, based on this study, the Secure Tropos methodology is adapted to measure the privacy of intrusion detection systems. It was possible to conclude that software developers still need greater depth in the study of privacy in the area of computer network security.

Keywords: Network security. Intrusion detection. IDS. Privacy metrics. LGPD. GDPR.

Lista de abreviaturas

LGPD	Lei Geral de Proteção de Dados
GDPR	*General Data Protection Regulation*
IDS	*Intrusion Detection System*
UE	União Europeia
PbD	*Privacy by Design*
PIA	*Privacy Impact Assessment*

Lista de figuras

Lista de tabelas

Sumário

Apêndices

Introdução

Privacidade na Internet tem sido um assunto recorrente tanto na literatura científica quanto em literatura para o público geral, inclusive em mídias sociais online (Silva e França, 2023). Essa preocupação com Segurança, Privacidade e Ética são essenciais hoje nas telecomunicações, devido à proliferação de ataques DDoS massivos e à necessidade de adequação às recentes regulamentações relacionadas a proteção de dados pessoais, como a LGPD (Lei Geral de Proteção de Dados Pessoais) no Brasil e a GDPR (*General Data Protection Regulation*) na Europa, apesar de já ter sido tratada em normas específicas em outros países como nos Estados Unidos Iwaya *et al.*, 2023. Outro motivo que tem aumentado as discussões em torno de privacidade na Internet é o aumento da utilização de técnicas de inteligência artificial, principalmente aprendizado de máquina, para classificação de dados. Por exemplo, em um sistema de detecção de intrusão (IDS, do inglês *Intrusion Detection System*), é necessário treinar o classificador com fluxos que sejam o mais próximo possível do real, para que ele tenha uma boa acurácia. Caso a detecção seja feita em um ambiente não confiável, por exemplo fora das premissas do proprietário dos dados, isso pode levar ao vazamento de informações sigilosas Mosaiyebzadeh *et al.*, 2023.

Do ponto de vista de quem possui os dados, seria importante que ele(a) tivesse uma gama de opções de ferramentas que garantisse a privacidade, por exemplo antes desses dados serem enviados para uma etapa de treinamento, supondo um sistema baseado em aprendizado de máquina. Porém, mesmo que houvesse tal gama disponível, a tomada de decisão sobre qual delas usar dependeria de alguma comparação entre as opções, o que justifica a importância de uma métrica padronizada para julgar a garantia de privacidade dessas opções[1]. Entretanto, ao se buscar por métricas de privacidade na literatura, muito se encontra a respeito de métricas do ponto de vista da percepção que o usuário tem de um sistema, ou métricas bem específicas do domínio de aplicação. Por exemplo, em um algoritmo de roteamento ad-hoc, a métrica de privacidade pode ser calculada como $1/n$, onde n é a quantidade de roteadores por onde um fluxo de rede passa. Porém, carece-se da aplicação dessas métricas gerais para atestar o aumento da privacidade em algum sistema que se proponha a isso. Isso ocorre principalmente pela dificuldade de se ter uma definição técnica absoluta sobre privacidade, fazendo com que a definição de uma métrica para esse conceito seja tão desafiadora Mendes e Vilela, 2017. Considerando a literatura científica, atestar o nível de privacidade de software científico resultante de pesquisas em redes de computadores é algo extremamente relevante.

Neste TCC, o principal objetivo foi estudar sobre privacidade e as métricas existentes

[1] Inclusive, nos artigos 35 e 36 da GDPR exige-se que, em operações onde haja riscos à privacidade, esses sejam mensurados mesmo antes deles ocorrerem Tikkinen-Piri *et al.*, 2018

para medi-la em sistemas de software resultantes de pesquisas científicas em redes de computadores. O conhecimento das métricas existentes foi aplicado para propor uma forma de medir o nível de privacidade de tais sistemas, principalmente daqueles que afirmam que melhoram a privacidade de algum cenário. O estudo sobre privacidade, suas métricas existentes e a busca pelos sistemas de software ideais para o projeto foi feito considerando a metodologia a seguir.

Metodologia do Trabalho

Para a execução deste trabalho os seguintes passos fora seguidos:

1. Revisão da bibliografia preliminar
2. Busca e revisão de literatura geral
3. Modelagem do sistema
4. Análise do modelo criado
5. Teorização da métrica
6. Busca e revisão de literatura específica

No passo 1, foi estudada a viabilidade da proposta observando um conjunto de 4 trabalhos principais. A partir desses trabalhos foi observado que seria possível criar uma métrica de privacidade no contexto de redes de computadores e, mais especificamente, em sistemas de detecção de intrusão.

Foi dado início então ao passo 2, em que 3 textos sobre privacidade em geral, 4 textos sobre métricas de privacidade, 3 textos sobre análise de sistemas, 4 textos sobre privacidade em aprendizado de máquina e 9 textos sobre IDSs foram lidos. Uma síntese dos trabalhos lidos pode ser vista na Tabela B.1. Esses textos foram importantes para embasar mais profundamente as motivações do trabalho e exemplificar modelos de como a privacidade é analisada e medida em outros contextos. Entretanto em nenhum dos trabalhos encontrados havia a proposta ou execução de uma métrica de privacidade para IDSs.

Nesses dois primeiros passos foram usadas como principais fontes de busca as seguintes bases de dados de artigos científicos: Google Scholar, IEEEXplore, ACM Portal e a biblioteca online SOL da Sociedade Brasileira de Computação. Material a respeito do assunto também foi encontrado no formato de vídeo, por isso a plataforma de compartilhamento de vídeo YouTube também foi considerada como fonte de informação (nesse caso, a autoria do conteúdo em vídeo foi levado em conta ao selecionar o conteúdo como relevante).

O passo 3, embora não estivesse no cronograma inicial do trabalho se mostrou necessário ao estudarmos métricas de privacidade em outros contextos. Foi observado que é preciso fazer a análise do sistema, especificando primeiro os riscos existentes e o que se deseja proteger. Além disso, a privacidade por si só possui diversos aspectos fazendo com que a análise também seja útil para separar quais aspectos são mais relevantes para o contexto em questão. Só então, após essa análise, foi possível pensar sobre como metrificar o sistema.

A modelagem do sistema portanto foi escolhida como uma forma de fazer essa análise de forma gráfica ao invés de textual. Esse método era mais familiar, além de ter se provado útil em outro trabalho estudado Kioskli *et al.*, 2022. O modelo foi várias vezes modificado e analisado (passo 4) uma vez que trabalhos anteriores não foram encontrados. O resultado gerado foram 3 diagramas que descrevem as principais restrições e riscos de um detector de intrusão baseado em aprendizado de máquina genérico (Figuras 3.1, 3.2 e 3.3).

Utilizando esse modelo e a bibliografia previamente encontrada foi possível "planejar" uma métrica (passo 5): cada aspecto da privacidade mapeado tem uma métrica relacionada, esses aspectos são medidos independentemente e então uma média entre eles é feita. Neste trabalho, porém, apenas foram sugeridas algumas métricas para cada aspecto e seu agrupamento não foi consolidado. Acredita-se que é necessário um estudo mais profundo sobre como unir as métricas e quais métricas usar. Um diagrama foi gerado (Figura 3.4) e dois IDSs (Shi *et al.*, 2021; Elias *et al.*, 2022) foram analisados seguindo a metodologia proposta.

Além disso, fazer um levantamento mais rígido de quais aspectos de privacidade devem ser considerado também se mostrou necessário ao considerarmos um sistema real. O passo 6, portanto diz respeito a busca por essas referências.

Organização da Monografia

O restante desta monografia está organizado da seguinte forma: O Capítulo 1 explica conceitos considerados básicos para o entendimento do tema tratado. No Capítulo 2 é feita a revisão das principais literaturas e outros trabalhos de apoio usados. A modelagem de um IDS é feita no Capítulo 3 bem como o planejamento da métrica. A aplicação da metodologia proposta é encontrada no Capítulo 4. Por fim o Capítulo 5 conclui o TCC. Os Apêndices A e B são apresentados no final desta monografia com métricas de preocupação de privacidade e com uma tabela sintetizando os trabalhos estudados.

Capítulo 1

Conceitos Básicos

Para a compreensão da contribuição do trabalho e de algumas decisões de projeto que foram tomadas, é importante revisar alguns conceitos básicos como: privacidade, regulamentações e sistema de detecção de intrusão.

1.1 Privacidade e Regulamentações Relacionadas

O conceito de privacidade já passou por diversas definições, sendo modificado e atualizado de acordo com o contexto e/ou momento social. Uma das primeiras utilizações do termo foi em 1890 em que ele é definido como o direito de ficar sozinho Warren e Brandeis, 1890. Neste caso, não estamos tratando de dados ou informações, mas sim de um estado físico do qual a pessoa teria direito. Em 1967, surge pela primeira vez uma definição que abrange dados pessoais: a habilidade de um indivíduo controlar os termos sob os quais suas informações pessoais são adquiridas e usadas Westin, 1967.

A partir dessa definição e com o reconhecimento de que cada vez mais dados pessoais seriam usados para fins comerciais, diretrizes regendo a manipulação dos dados começaram a ser propostas em diversas regiões do mundo. Nos anos 70, a Alemanha e a Suécia adotaram leis de proteção de dados e o Governo dos Estados Unidos formulou um conjunto de regras (FIPs, *Fair Information Practices*) que as organizações deveriam seguir para garantir a privacidade dos usuários. Essas práticas foram então usadas como base para a criação da lei de privacidade americana em 1973. Esses mesmos princípios foram utilizados em outros países como base de suas próprias legislações Tikkinen-Piri *et al.*, 2018.

Na Europa, diversas diretrizes foram criadas para cada país, dificultando o transporte e uso de dados dentro da União Europeia (UE). Em 1990, iniciou-se uma tentativa de uniformizar a proteção de dados na UE e, em 1995, foi adotado a "*European Commission's DIR95*", que, após diversas atualizações, seria substituída pela *General Data Protection Regulation* (GDPR) em Maio de 2018. Alguns meses depois, o Brasil também teria uma legislação específica para proteção de dados e privacidade, a *Lei Geral de Proteção de Dados Pessoais* (LGPD).

Ao mesmo tempo, definições, tecnologias e conceitos envolvendo privacidade também iam se desenvolvendo. Na área de segurança, a privacidade foi definida dentro do pilar de

Confidencialidade, garantido que os usuários controlem as autorizações sobre os seus dados e que apenas as pessoas autorizadas tenham de fato acesso a eles Stallings e Brown, 2021. Sobre os avanços de técnicas e conceitos, em 1995, no Canadá, os princípios de *Privacy By Design* (PbD) e *Privacy Enhancing Technologies* (PET) foram desenvolvidos Hustinx, 2010. PbD é um conjunto de princípios que devem ser seguidos para que o desenvolvimento de um projeto seja orientado a privacidade, dando ênfase que, para a privacidade ser plena, ela deve estar presente durante todas as etapas do projeto e manipulação dos dados. Já PETs, são tecnologias desenvolvidas para aumentar a privacidade de um conjunto de dados, sendo mais comumente usadas em dados que virão a ser públicos. Ambas ideias foram essenciais para o desenvolvimento de políticas e técnicas que garantissem a privacidade no meio organizacional. Em 2004, a definição de privacidade foi estendida entendendo a ligação profunda entre a privacidade e o contexto em que está inserida Nissenbaum, 2004. Em 2013, Densmore, 2013 divide a definição de privacidade em 4 categorias: de informação, física, espacial e de comunicação. Privacidade de informação é definida como um conjunto de regras que controla a coleta e uso das informações pessoais; privacidade física (corporal) como o conjunto de regras que protege o corpo físico; privacidade espacial (territorial) como um conjunto de regras que protege o ambiente em que se encontra o corpo físico; e, por fim, privacidade de comunicação como um conjunto de regras que protege a comunicação e seus meios de correspondência. Neste trabalho, estaremos considerando privacidade como privacidade de informação, assim, nos preocuparemos com os dados, sua sensibilidade, seus princípios e as regras vigentes sobre eles.

Hoje, mais do que definições sobre privacidade, buscam-se também formas de se garanti-la. Existem duas principais abordagens: a primeira, foca em um aspecto mais técnico querendo assegurar acima de tudo que os dados não entreguem informações sigilosas. Neste caso as técnicas empregadas seriam PETs, criptografia e/ou uso de dados artificiais. A segunda abordagem é por políticas de controle, como as leis antes citadas, e possuem mais foco no usuário Wagner e Eckhoff, 2018. Essas duas abordagens não são completamente excludentes entre si, já que no primeiro caso é necessário conhecimento e concordância com as leis e, no segundo, que ainda se tenha alguma maneira técnica de se comprovar o cumprimento da lei; sua separação portanto está no enfoque que se dá a cada aspecto. Neste trabalho nos dedicaremos a pesquisar métricas de privacidade a partir da primeira abordagem, ou seja, focaremos nos dados e nas técnicas usadas sobre eles. Um estudo breve sobre métricas considerando a segunda abordagem, mais especificamente, focando no usuário, foi feita e se encontra no Apêndice A.

Assim, aqui a privacidade será entendida como proteção aos dados sensíveis. Uma vez que a classificação dos dados entre sensíveis ou não ainda não está plenamente desenvolvida, o enfoque do trabalho será em torno de analisar a quantidade mínima de dados necessários para o sistema. Essa ideia vai ao encontro dos princípios de necessidade e adequação defendidos pela LGPD, GDPR e PbD.

1.2 IDS

Um IDS (*Intrusion Detection System*), ilustrado na Figura 1.1, é um sistema capaz de detectar se um pacote que está adentrando uma rede é malicioso, podendo ler tanto o cabeçalho do pacote quanto seu conteúdo. Na literatura Kurose e Rossa, 2009 são descritos

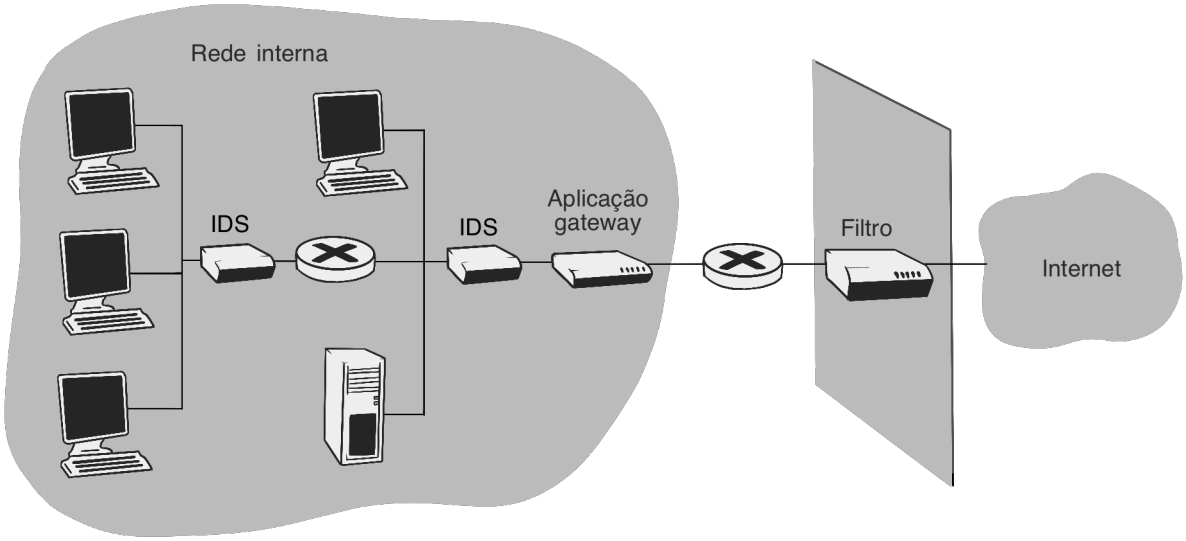

Figura 1.1: *Ilustração IDS (baseado na figura 8.37 de Kurose e Rossa, 2009)*

dois principais tipos de IDSs: 1- baseados em assinatura, em que se compara diretamente trechos dos pacotes com assinaturas de ataques conhecido. Para isso é necessário haver um banco de dados com essas assinaturas, assim, a principal dificuldade desse tipo de detectores é como manter o banco atualizado de forma a não ser vulnerável a novos ataques. 2- baseado em anomalias, em que é reconhecido se o fluxo dos pacotes é anômalo. A maioria dos detectores da segunda categoria utiliza técnicas de aprendizado de máquina supervisionado fazendo com que o IDS funcione como um classificador. Dessa forma, ele é treinado para detectar se um pacote, ou sequência deles, é estranho(a) podendo trazer risco a rede. Esse treinamento é feito utilizando um conjunto de dados em que diversos pacotes comuns de redes de computadores já estão classificados como maliciosos, ou não, podendo inclusive especificar a qual tipo de ataque aquele pacote se refere.

Assim temos que o processo de treinamento desses IDSs segue uma rotina similar a sistemas que utilizam aprendizado de máquina supervisionado. Primeiro deve-se escolher e treinar o modelo da rede detectora sendo que esse modelo corresponde a um algoritmo de aprendizado que dita como será feita a classificação. Nessa etapa, são usados dados gerais ao problema (no nosso caso, estes seriam pacotes de rede de computadores que são marcados como maliciosos ou não). Para comprovar sua eficiência uma técnica muito comum é dividir o conjunto de dados em treino e testes, assim, após o modelo ser escolhido e treinado ele pode ser testado com um conjunto de dados "novos", comprovando sua eficiência. Depois disso, é feito o treinamento dos parâmetros do modelo para utilização dentro de uma rede local específica utilizando pacotes classificados dessa própria rede. Ou seja, teremos que o modelo, escolhido na etapa anterior, deve ter seus pesos (coeficientes) ajustados para reconhecer o ambiente específico da rede local em que o detector será colocado e aprender sua "normalidade". Por fim, classificamos os novos dados entrantes dessa rede local utilizando o detector. O treinamento dos parâmetros, e até a escolha do modelo, podem ser feitos frequentemente, possibilitando que novas entradas sejam utilizadas no treinamento, aumentando sua precisão e atualidade Arbex *et al.*, 2021.

A implementação de um IDS é um tema muito recorrente na literatura de segurança

de redes. Grandes esforços são feitos para aumentar a precisão com que esses sistemas reconhecem uma anomalia, aumentando, consequentemente, a segurança da rede local. Dessa forma, vale ressaltar que a importância principal dos IDSs é garantir a segurança da rede, sem que haja, a principio, qualquer preocupação com a privacidade.

Porém, visto a quantidade de informações que os IDSs têm acesso Elias *et al.*, 2022, novos esforços em melhorar a privacidade desses sistemas estão sendo estudados, principalmente considerando um cenário em que o treinamento seja realizado fora das premissas da organização, algo comum em um cenário de computação em nuvem. Existe uma questão entre segurança e privacidade discutida em relação aos IDSs: ao aumentar a segurança da rede local instalando-se um IDS, diminui-se a privacidade do acesso a rede que será constantemente monitorado por esse sistema terceiro. Porém, ao aumentar a privacidade, por exemplo reduzindo o acesso às informações originais dos pacotes, a eficiência e precisão do detector podem diminuir, afetando a segurança.

Capítulo 2

Revisão da Literatura

Neste capítulo iremos resumir e analisar trabalhos relacionados a métricas de privacidade. O estudo dessas métricas no contexto de redes de computadores e, mais especificamente, no de IDSs ainda está muito limitado. Por isso, optou-se por estudar artigos e palestras que tratem a privacidade num contexto mais amplo ou num contexto específico completamente diferente do nosso mas cujas metodologias de cálculo e análise possam ser usadas como base em nosso cenário. Aqui cinco trabalhos foram resumidos individualmente em 5 seções (2.1, 2.2, 2.3, 2.4, 2.5) e outros dez trabalhos foram compilados na última seção desse capítulo (2.6). Uma síntese dos trabalhos revisados pode ser vista na Tabela B.1.

2.1 Developing a Structured Metric to Measure Privacy Risk in Privacy Impact Assessments

O artigo Agarwal, 2016 se encontra no contexto de uma avaliação de impacto de privacidade (PIA, do inglês *Privacy Impact Assessments*). PIA, é um processo que ajuda organizações a identificarem e gerenciarem os riscos de privacidade decorrentes de um novo projeto ou política implantada. Embora esse não seja o cenário em que este trabalho esteja inserido, é de extrema relevância analisar projetos que visem medir a privacidade em uma cadeia de produção sob legislações que regulam o tratamento de dados pessoais. Em especial, a PIA é prevista pela GDPR como uma medida de proteção proativa que cumpre com as especificações obrigatórias da lei Tikkinen-Piri *et al.*, 2018. Por esse motivo, formas de calcular a privacidade nessa área, principalmente empresarial, se encontram mais desenvolvidas e suas considerações podem ser usadas como base em outros contextos.

A autora afirma que uma avaliação de impacto de privacidade deve ser fácil e rápida de se realizar, pois deve ser feita continuamente ao longo de todo o processo de desenvolvimento de um sistema. Entretanto, a maioria das avaliações resultam em longos relatórios que são complexos de entender e comparar. Dessa forma, ela procura estudar formas já existentes para medir a privacidade durante uma análise PIA e propor uma nova métrica quantitativa para suprir as deficiências das demais métricas. As seguintes métricas são apresentadas e suas principais desvantagens são discutidas no artigo:

- GS1 PIA Tool (*GS1* 2012): classificação muito genérica e semiquantitativa

- iPIA Tool (Oetzel e Spiekermann, 2014): não classifica risco, e a demanda de proteção é qualitativa
- SPIA Tool (*SPIA* s.d.): difícil distinguir o critério de pontuação

Analisando-se essas métricas, percebeu-se que era necessário uma métrica mais quantitativa que pudesse ser usada como meio de comparação. A autora então propõe uma métrica semiquantitativa capaz de comparar duas versões de um mesmo projeto em relação aos riscos de privacidade.

A metodologia apresentada é:

1. identificação do risco: qualquer cenário em que se está contra a lei (ou colocando sua integridade em risco) configura um cenário de risco a privacidade
2. modelagem do risco: análise qualitativa do risco encontrado, simplifica e abstrai o cenário, descrevendo as possibilidades de perda, dano ou destruição de um recurso por conta da exploração a vulnerabilidade resultante desse risco
3. avaliação do risco: tentativa de quantificar e pontuar um risco. No artigo a técnica usada é analisar o impacto e probabilidade do dano modelado ocorrer. São descritas um total de 13 formas de impacto (por exemplo: identificação, exclusão, exposição). Para calcular a pontuação de impacto, atribui-se um valor binário para a existência de cada um deles e então tem-se uma proporção de quantos impactos são encontrados no sistema. Para a medida de probabilidade existem "parâmetros" (por exemplo: pessoas envolvidas, quantidade de dados, valor dos dados), para cada um deles damos uma nota do quão problemático/relevante ele é para o risco encontrado e então somamos esses valores. Cada parâmetro tem um valor máximo, assim a nota de probabilidade é a razão entre as somas dos valores dados e a soma dos valores máximos possíveis (pré definido)

A medida resultante pode ser usada como guia para calcular o PIA de um sistema, entretanto, é importante ressaltar que ela é relativa, ou seja, não pode ser utilizada para comparar projetos diferentes, mas funciona muito bem para comparação de versões e a progressão de um mesmo projeto.

O estudo desse artigo é relevante pois apresenta uma forma de quantificar características qualitativas durante a análise de risco de privacidade de um sistema. Isso é feito de forma que a "medida" calculada é baseada em problemas pré definidos e sua pontuação é dada de forma a analisar a presença e possibilidade de ocorrer cada problema devido a vulnerabilidade gerada pelo risco. Essa metodologia poderia então ser usada no contexto de redes.

Outro aspecto importante é a discussão sobre a efetividade dessa forma de cálculo. Quando os desenvolvedores não entendem a necessidade e a forma que a privacidade é medida, existe uma maior chance de se deixar de lado essa avaliação. Dessa forma, é importante que, ao tratarmos de uma métrica de privacidade, os pontos de risco sejam claramente explicados e as características medidas sejam facilmente interpretadas e observadas. No caso desse trabalho, tal esclarecimento foi feito de forma que a metodologia seja dividida em passos lógicos e diretos e, na etapa de avaliação de risco, as categorias dos riscos e suas escalas de pontuação já estejam pré-definidas.

Entretanto, a relatividade dessa avaliação é uma falha quando trazida para nosso contexto uma vez que desejamos ser capazes de comparar diferentes IDSs. A relatividade da métrica é proveniente principalmente de como as categorias de risco são medidas. A pontuação embora possua uma escala bem definida (com máximo e mínimo) não possui uma regra clara em como incluir ou retirar pontos ficando a cargo do desenvolvedor, de forma pessoal, atribuir a pontuação. Mesmo assim, ela continua sendo uma forma quantitativa de se analisar a privacidade, passo importante para a confecção de uma métrica de privacidade.

2.2 Styx: Design and Evaluation of a New Privacy Risk Communication Method for Smartphones

A proposta do artigo Bal *et al.*, 2014 é apresentar o estudo feito sobre privacidade em aplicativos de celular e como essa informação é transmitida ao usuário. Os autores argumentam que os usuários não têm meios fáceis de saberem quais os dados que os aplicativos coletam e quais as informações que podem ser retiradas deles. Pensando nisso, os autores desenvolveram uma aplicação que consegue: informar os usuários de forma clara e simples as informações que estão sendo coletadas, quais os possíveis riscos de privacidade que o aplicativo traz, as inferências que podem ser feitas sobre o usuário e um sumário simples de ser interpretado sobre as questões de privacidade levantadas. O principal objetivo é criar uma aplicação capaz de comunicar ao usuário os riscos de privacidade dos aplicativos de celular de forma fácil e que seja entendida por todos. As diretrizes do projeto seguidas foram:

- Evitar o uso de jargões de privacidade
- Comunicar a existência de riscos
- Filtrar as informações e alertar os usuários sobre potenciais riscos
- Minimizar distrações
- Não obscurecer o fluxo de informação
- Prover oportunidades de educação do usuário
- Prover sumários claros para os usuários
- Considerar exoinformação [1]

Nesse caso temos que a "medida" de privacidade calculada foi feita pensando em quais informações são coletadas pelo aplicativo e quais inferências podem ser feitas a partir dela. Não está claramente especificado como a métrica é calculada, mas os valores finais são exibidos aos usuários, utilizando-se de gráficos, figuras e relatório conciso. É importante ressaltar também que esse projeto não foi colocado em prática, embora tenham ocorrido alguns casos de teste, os autores não conseguiram colocar o projeto em produção. Sendo assim, o foco do artigo está principalmente no design da aplicação.

[1] Exoinformação se refere às novas informações que podem ser inferidas a partir dos dados coletados

O professor Kai Rannenberg, segundo autor do artigo, participou da palestra *SECOMU: Falta de privacidade e controle comportamental numa economia de vigilância* 2021, realizada em 2021 no Congresso da Sociedade Brasileira de Computação (CSBC). O autor falou sobre o estudo desse sistema e apresentou um caso de uso em que ele foi usado para detectar que haviam aplicativos de lanterna captando dados sobre geolocalização do usuário, o que, para as finalidades do aplicativo, são coletas incoerentes e podem significar um possível risco à privacidade.

Esse estudo contribuiu para este trabalho uma vez que explora exaustivamente como o usuário pode se informar e, principalmente, comparar diferentes aplicativos em seu celular. De forma semelhante, este trabalho tem por objetivo ajudar os pesquisadores e usuários de IDS a analisar e comparar de forma fácil e clara diferentes trabalhos propostos em termos de privacidade.

Em específico, o estudo faz uso de figuras e gráficos que, em conjunto com pequenas descrições com palavras comumente usadas, torna a compreensão do risco geral (ou segurança em determinados casos) mais óbvia ao usuário. Essas práticas foram também levadas em consideração durante o desenvolvimento deste projeto.

Entretanto pela diferença de contexto dos sistemas e aparelhos utilizados e dos dados disponíveis não foi possível utilizar diretamente a forma representativa que os autores usaram nesse trabalho, sendo necessário procurar outras representações gráficas que sigam as diretrizes de design e sejam tão facilmente interpretadas quanto as apresentadas.

2.3 Privacidade e Monitoramento: Uma perspectiva LGPD e GDPR - RoadSec

Durante a palestra Monteiro, 2020, a consultora Alessandra Monteiro traz alguns conceitos sobre privacidade, sua relação com as leis e como o projeto de um sistema que possui contato com uma grande quantidade de dados deve ser feito para garantir a privacidade dos usuários e ser concordante com lei. Dessa forma existem algumas boas práticas que devem guiar a fase de desenvolvimento (a palestrante ressalta que elas devem ser seguidas durante toda a fase de construção do projeto, como uma medida preventiva).

O primeiro conjunto de princípios discutido é o *Privacy by Design* (privacidade desde sua concepção) Hustinx, 2010:

1. Proatividade e não reatividade: deve-se propor medidas a se garantir a privacidade antes de um problema de fato ocorrer
2. Embarcada no design: a arquitetura e modelo do projeto devem garantir a privacidade dos dados usados
3. Segurança fim a fim: a informação, durante toda sua vida (coleta, uso e descarte), deve ser protegida
4. Respeito pela privacidade do usuário: a garantia de privacidade deve ser concordante ao interesse dos usuários

5. Privacidade como configuração padrão: a configuração padrão do sistema deve garantir privacidade ao usuário
6. Funcionalidade completa: deve-se possibilitar o uso do sistema sem que a privacidade do usuário seja prejudicada
7. Visibilidade e transparência: deve-se possibilitar que o usuário garanta que os princípios estão sendo seguidos

Além disso foram expostos rapidamente alguns princípios sobre *Security by Design* que também contribuem em nível de privacidade:

1. Minimizar a superfície de área de ataque
2. Estabelecimento de padrões
3. Princípio do menor privilégio: deve-se atribuir "perfis" aos usuários do sistema de forma que o mínimo de privilégio sobre o sistema seja dado a cada um deles
4. Princípio da defesa por profundidade: segurança distribuída por todas as camadas do sistema
5. Falhar com segurança: não expor seus dados em caso de falha
6. Não confiar nos Serviços: inspecionar pacotes e bibliotecas usadas, gerenciar versões e privilégios e fazer manutenção do sistema
7. Separação de deveres: a mesma pessoa não deve ter acesso a todas informações e etapas do desenvolvimento pois isso impacta nas camadas de segurança
8. Evitar segurança por obscuridade
9. Mantenha a segurança simples

Diferente do anterior, o conceito de *Security by Design* ainda não possui um conjunto fundamental básico de princípios, mas os pontos apresentados são os principais defendidos. Seguindo esses dois conceitos, temos os princípios jurídicos da LGPD que disciplinam o tratamento de dados dentro da lei nacional.

1. Finalidade: haver uma finalidade específica de coleta e tratamento das informações
2. Adequação: os dados coletados se relacionam às finalidades
3. Necessidade: somente tratar dos dados que são imprescindíveis para a finalidade
4. Livre acesso: o usuário tem acesso ao dado que ele forneceu, sabe o que foi coletado e por quanto tempo será usado
5. Qualidade dos dados: direito do usuário à correção de dados incorretos/incompletos dentro das necessidades e finalidades
6. Transparência: o usuário sabe como e por quem será realizado o tratamento dos dados

7. Segurança: utilizar medidas técnicas e administrativas para garantir a proteção dos dados [2]
8. Prevenção: reiteração do princípio anterior, prevenir ocorrências de danos
9. Não discriminação: a utilização dos dados não deve ser para fins discriminatórios ou ilícitos
10. Responsabilização e prestação de contas: comprovar e mostrar a eficácia de que os princípios estão sendo seguidos

Existe uma forte correlação entre a LGPD e os princípios PbD Castro *et al.*, 2022, e, embora esses conceitos sejam antigos (Privacy by Design foi primeiramente citado em 1995, por exemplo), eles ainda são vagos para os desenvolvedores Rocha e E. Canedo, 2023. Formas práticas e diretas de se traduzir o que é sugerido pelos especialistas legislativos em funcionalidade do sistema ainda não são claras, e muito menos saber quantificar a efetividade e o cumprimento de tais regras dentro do projeto. Esses princípios entretanto são importantes pois descrevem pontos qualitativos a serem analisados para entender o nível de privacidade de um sistema. Dessa forma, é necessário conhecer e estudar esses conceitos para que os progressos feitos neste trabalho reflitam os pontos discutidos, principalmente no que diz respeito ao último princípio da LGPD (Responsabilização e Prestação de contas).

2.4 Technical Privacy Metrics: A Systematic Survey

Em Wagner e Eckhoff, 2018 os autores analisam métricas disponíveis para medir a eficiência de PETs e definem alguns conceitos:

- Violação de privacidade é definida como o uso de informações pessoais de forma a descumprir a norma legal.
- Uma métrica m, é uma medida para distância entre dois elementos de um conjunto X e deve:
  - ser não-negativa: $m(x, y) \geq 0, \forall x, y \in X$
  - ser simétrica: $m(x, y) = m(y, x), \forall x, y \in X$
  - seguir a desigualdade triangular: $m(x, z) \leq m(x, y) + m(y, z), \forall x, y, z \in X$
  - seguir o princípio da identidade dos indiscerníveis: $m(x, x) = 0, \forall x \in X$
- Domínios de privacidade são áreas em que PETs podem ser aplicadas. Alguns domínios comentados no artigo foram: sistemas de comunicação, banco de dados, serviços baseados em localização, métricas inteligentes, redes sociais e privacidade de genoma.

Métricas de privacidade compartilham certas características em comum:

[2] Essas medidas são entendidas como técnicas contemporâneas de segurança aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão. https://www.conjur.com.br/dl/artigo-marcio-pestana-lgpd.pdf

1. **Objetivos do adversário**: ao medir o nível de privacidade que um PET garante ao sistema é importante considerar um adversário específico e quais seus objetivos ao atacar o sistema.
2. **Capacidades do adversário**: é necessário fazer um modelo do adversário. Características como: local/global, ativo/passivo, interno/externo, estático/adaptativo, conhecimento prévio e recursos são importantes nessa modelagem do adversário.
3. **Origens do dado**: qual dado deve ser protegido e como um suposto adversário poderia tomar posse dele. Exemplos de dados: publicados, observáveis, de objetivos repropostos, outros.
4. **Entrada para cálculo da métrica**: diferentes métricas podem usar diferentes informações para fazer o cálculo do valor da privacidade. Os dados de entrada podem ser: estimação do adversário, recursos do adversário, resultados verdadeiros (o quanto da estimação do adversário se concretizou) e/ou informação prévia.
5. **Saída**: o tipo de propriedade da privacidade que aquela métrica está medindo. É importante ter esse conceito bem categorizado pois ele demonstra como uma única métrica é incapaz de capturar todo o conceito de privacidade.
    - **Incerteza**: alta incerteza dos dados corresponde a uma alta privacidade, pois o adversário teria dificuldade em se aproveitar das informações
    - **Ganho ou perda de informação**: quanta informação o adversário consegue ganhar ao ter acesso aos dados.
    - **Similaridade de dados**: quão bem é possível adquirir informações "privadas" observando informações públicas (mais atrelado ao tipo de dado do que ao adversário)
    - **Indistinguibilidade**: analisar se duas saídas são indistinguíveis. A privacidade é maior quando não é possível saber se duas informações pertencem ao mesmo usuário. Geralmente a saída da métrica é binária.
    - **Probabilidade de sucesso do adversário**: dado um número de ataques, qual a proporção que teve sucesso.
    - **Erro**: quão correta a estimação do adversário é sobre os dados (se a estimação feita tem poucos erros e a distância do "real" é curta, ou seja, se aproxima do real, baixa privacidade)
    - **Tempo**: supondo que o sucesso do adversário é inevitável, calcula quanto tempo demora para conseguir os dados. Alto gasto de tempo corresponde a alta privacidade
    - **Acurácia ou precisão**: quantificar o quão precisa é a estimação do adversário, sem considerar a corretude da estimação.

Para cada um dos tipos de saída apresentados acima, os autores compilaram uma série de métricas que visam medir qual o nível de cumprimento dessa propriedade em um conjunto de dados. Inclusive muitas dessas métricas estão associadas a PETs utilizadas no conjunto. Por exemplo, a técnica do k-Anonimato, em que retiramos informações do

conjunto de dados de forma que sempre haverá no mínimo k entradas com a mesma identidade, fazendo com que esses k elementos estejam anônimos dentro desse grupo, pode ser usada como métrica de similaridade, em que k será o valor da métrica. Assim, quanto maior o k, mais privado estão os dados dos indivíduos dentro desse conjunto.

Todas as métricas apresentadas possuem uma equação de como calculá-las e uma pequena descrição explicando qual aspecto de privacidade essa métrica está medindo. Essa descrição também vem muitas vezes acompanhada de vantagens e desvantagens de se usar essa métrica, se ela foi derivada de outra métrica e, se sim, qual foi e quais os novos benefícios que esse outro cálculo traz.

Por fim, os autores também discutem, como escolher as melhores métricas, ressaltando que nenhuma delas é capaz de medir a privacidade geral do sistema, apenas aspectos dela. Dessa forma, é relevante antes de tudo, saber qual aspecto é o alvo da saída da métrica desejada, os recursos que possuem e contra qual tipo de adversário o sistema está tentando se proteger. Não só isso, mas também é desejável que, ao se calcular a métrica, se utilize uma implementação já conhecida da qual já se tenha provado sua eficiência e corretude. Essas métricas, embora descritas de forma matemática, possuem parâmetros cujos valores podem não ser tão facilmente definidos ou que necessitam de testes empíricos para serem descobertos (por exemplo, probabilidade ou tempo de sucesso do atacante), e portanto faz-se necessário um estudo prévio de como conseguir esses valores para trazerem melhores resultados na hora de se calcular a métrica.

Esse artigo é relevante para este trabalho pois constitui uma importante base para entender as principais métricas de privacidade utilizadas em conjunto de dados e que podem ser utilizadas em um IDS. Ele descreve aspectos relevantes sobre a privacidade além de mostrar como cada um deles pode ser calculado de diferentes maneiras. Além disso, as definições e discussões que permeiam a utilização dessas métricas foram levadas em consideração durante este trabalho.

Porém, em relação ao trabalho aqui desenvolvido, alguns pontos estão faltando. Por exemplo, não houve qualquer menção sobre propósito ou re-propósito *(re-purpose)* dos dados, esse conceito é de extrema importância quando falamos de privacidade, uma vez que nas principais legislações sobre privacidade (LGPD e GDPR) existe um grande apelo em garantir que os dados coletados/tratados sejam exclusivos para as finalidades propostas e sua necessidade claramente explicada ao usuário. Uma forma de se calcular a garantia dessa cláusula entretanto não foi discutida. Além disso, como comentado anteriormente, conseguir alguns dos valores para calcular a métrica pode ser trabalhoso. Isso é um empecilho para o uso da métrica, já que, como os próprios autores sugeriram, é necessário usar as métricas em conjunto, fazendo com que a parte prévia ao cálculo possa ser complexa e, dessa forma, desincentivando a utilização da métrica.

Ademais, embora o artigo não trate sobre análise de privacidade sobre um sistema de forma direta, chegou-se a conclusão que ela é necessária. Quais são os riscos, quais aspectos da privacidade devem ser mantidos, quais os recursos e dados atrelados ao sistema e a quem interessa os resultados da métrica de privacidade, são pontos que devem ser primeiramente mapeados para que então possam ser escolhidas as métricas a serem usadas.

2.5 The Supply Chain of a Living Lab: Modelling Security, Privacy, and Vulnerability Issues alongside with their Impact and Potential Mitigation Strategies

No trabalho Kioskli *et al.*, 2022, os autores estudam a esfera de um Laboratório Vivo (*Living Lab*) examinando a segurança e privacidade desse contexto. O artigo é iniciado discutindo os principais desafios, ataques e riscos que permeiam o ambiente médico, além de descrever a cadeia de produção e agentes envolvidos. Essa análise entretanto não será discutida uma vez que está fora do escopo desta pesquisa. A modelagem de privacidade e as técnicas envolvidas para tal, no entanto, são de grande valor e foram utilizadas como principal base para este projeto.

Em um *Living Lab* há a integração de pesquisa e inovação em contextos práticos de uma clínica médica. Nesses espaços há uma grande troca de informação entre os vários níveis organizacionais: visitantes, pacientes, médicos, pesquisadores, trabalhadores gerais e trabalhadores terceiros. Dessa forma há uma grande circulação de informações pessoais que pode trazer riscos à privacidade dos usuários. Por essa razão foi feita uma elicitação das restrições, modelagem dos riscos e um mapeamento dos dados seguindo as diretrizes da metodologia Secure Tropos Mouratidis e Giorgini, 2007, utilizando a ferramenta SecTro Pavlidis e Islam, 2011.

Secure Tropos deriva da metodologia Tropos (Giunchiglia *et al.*, 2003; Giorgini *et al.*, 2004; *Tropos Project* s.d.) dando ênfase maior aos aspectos de segurança. Tropos é baseado em duas principais ideias: agentes e seus objetivos. Dessa forma, a metodologia irá auxiliar a criação e análise de um sistema desde as fases de concepção até a implementação de fato. Isso é vantajoso para o desenvolvimento do sistema uma vez que permite ter uma maior abstração de diversas etapas do processo, como também unificar o levantamento de restrições, planejamento de mecanismos e elementos sociais, como os objetivos dos stakeholders, as regras e diretrizes do contexto e os agentes envolvidos.

Assim, Secure Tropos é uma metodologia de desenvolvimento de software concordante com a segurança, que permite uma analise completa do desenvolvimento do sistema, combinando elementos de Engenharia de Software, como os conceitos de agente e objetivos, que permitem um planejamento geral do sistema, junto com os elementos de Engenharia de Segurança, como riscos, restrições e mecanismos de segurança *Secure Tropos Project* s.d. Além disso, por seu aspecto social, permitindo a análise da interação dos agentes e dos recursos, é possível analisar de forma sociotécnica os riscos e restrições de privacidade desde os estágios iniciais, sendo assim, concordante também com os princípios de PbD. Secure Tropos é uma metodologia muito usada na academia para fazer o levantamento de requisitos de privacidade, bem como o planejamento de um sistema E. D. Canedo *et al.*, 2023, isso ocorre pois os vários modelos implementados pela metodologia permitem a representação de diferentes níveis de abstração, as visões, cada uma dando foco a diferentes aspectos do sistema. Abaixo serão apresentados as três visões (Organizacional, de Dados e de PbD) apresentadas no artigo Kioskli *et al.*, 2022.

Visão Organizacional, onde é representada a estrutura organizacional em que os Laboratórios vivos estão inseridos, dando ênfase a seus principais atuadores (agentes), objetivos e recursos utilizados. Nesse contexto é possível abstrair os vários níveis de segurança e privacidade que cada agente deve ter para que se diminua os riscos de uso dos recursos ao mesmo tempo que torna viável as operações dos agentes.

Visão de Dados, onde são representadas as ações executadas sobre os recursos. Nesse caso, embora essa análise pareça ser repetitiva, ela tem por objetivo enfatizar como e o que será feito sobre cada recurso sendo portanto possível esclarecer a necessidade do recurso para o agente executar a ação. Ter clareza sobre esses aspectos é essencial para a próxima visão.

Visão de *Privacy by Design (PbD)*, onde são modelados os principais riscos e restrições sobre/dos recursos. Essa visão é importante para guiar o desenvolvimento do sistema, uma vez que segue os princípios de PbD. Ela também ajuda a identificar quais medidas devem ser tomadas e quais mecanismos devem ser implementados para garantir a privacidade dos usuários relacionados ao laboratório. Nessa etapa final, são mostradas quais são as principais funcionalidades que o sistema/projeto deve ter para que siga as diretrizes do PbD, além de mostrar como vários agentes e recursos podem estar expostos ao mesmo tipo de risco, mas devem ter contra medidas diferentes já que o nível de segurança é diferente.

Essas três visões são geradas de forma sequencial. Para cada visão, um diagrama baseado na visão anterior é criado, podendo ganhar ou perder alguns de seus elementos de forma que os principais aspectos de análise sejam mantidos em foco.

Embora esse artigo não trate de métricas de privacidade, ele foi utilizado como principal base de estudo uma vez que foi capaz de traduzir princípios abstratos do PbD e transformar em funcionalidades técnicas a serem implementadas. Essa mesma abordagem foi usada em nosso contexto de IDSs, porém de maneira inversa. No artigo apresentado a metodologia foi utilizada para desenhar como o sistema do *Living Lab* deve ser para que ele siga as práticas do PbD. Assim, os mecanismos, que futuramente seriam implementados, estariam ligados diretamente a uma medida que garantisse uma restrição. No nosso caso, como queremos analisar IDSs já prontos, os mecanismos já estão implementados e já fazem parte do sistema. Portanto, a proposta é usar os diagramas gerados pela metodologia para associar um mecanismo a uma restrição, destacando como ele está ou não contribuindo para privacidade, e posteriormente medir sua eficiência. Essa rotina será explicada mais detalhadamente no Capítulo 3.

A principal limitação desse artigo em relação ao trabalho é exatamente seu contexto, já que trata de um ambiente médico, com vários agentes, contato direto com o público e frequente movimentação dos dados entre os agentes e atuadores terceiros. No contexto de um IDS, o número de agentes é reduzido para dois (administrador da rede e treinador do modelo de detecção) diminuindo consideravelmente a movimentação e contato com os dados. Além disso, ele não discute sobre métricas de privacidade, mas sobre análise e modelagem do sistema.

2.6 Leituras Adicionais

Nesta sessão serão discutidas algumas ideias encontradas em textos que se relacionam a privacidade e ao projeto de duas formas: aprendizado de máquina e sistemas de detecção de intrusão. Foram lidos no total dez textos sendo três de aprendizado de máquina e seis relacionados com sistemas de detecção e um relacionado a ambos. A revisão dessas literaturas tem como principal objetivo pontuar quais são as principais preocupações e soluções envolvendo privacidade em diferentes contextos.

2.6.1 Aprendizado de Máquina

As áreas de *big data* e aprendizado de máquina têm levantado diversas questões sobre privacidade. Isso ocorre pois para treinar um modelo é necessário uma grande quantidade de dados, esses dados muitas vezes contém informações sensíveis, e, mesmo quando não têm, devido a sua grandeza, é possível utilizá-los como base de inferência e de outras funções contribuindo para o vazamento de informações pessoais.

Pensando nesses problemas observamos duas estratégias. A primeira é sobre geração de dados artificiais Yoon *et al.*, 2020. Nesse caso, o objetivo é aumentar a privacidade de modo que nenhuma informação pessoal seja vazada, uma vez que essas informações não são de pessoas reais. O grande problema, e que o artigo antes citado deseja resolver, é que seria possível identificar quais usuários foram usados para gerar os dados sintéticos, dessa forma, a partir de um dado gerado é possível retirar dados de seu usuário gerador, sendo portanto uma brecha na privacidade. Outro problema é que a geração também é um processo gerado por aprendizado de máquina, sendo portanto, novamente, dependente de dados de treinamento que podem ser escassos e/ou possuir riscos de privacidade.

Outra estratégia é tornando os dados anônimos, assim, mesmo que possuam dados sensíveis o atacante não saberia a quem ele pertence. Um dos métodos para fazer isso na etapa de treinamento é o aprendizado federado Shi *et al.*, 2021; Mosaiyebzadeh *et al.*, 2023. Nele, ao invés de compartilharmos dados entre proprietário dos dados e o treinador, será enviado o modelo e os parâmetros. O treinamento e a validação serão feitos no próprio ambiente do proprietário dos dados fazendo com que os dados permaneçam privados.

Outra proposta de anonimização é remover as informações sensíveis do conjunto de dados. Entretanto, embora pareça uma estratégia direta, ela possui diversas questões: como tirar informações suficiente para que se garanta a privacidade, mas que ainda seja simétrico o suficiente para ser uma fonte confiável de treinamento? Além disso, muitas vezes um único proprietário dos dados não tem dados suficiente para anonimização e precisa unir com outros, mas eles podem não ser confiáveis entre eles, bem como um terceiro elemento "central" pode aumentar o risco também Kim e Chung, 2019. Nesse caso a solução apresentada extrapola mecanismos de proteção de dados e desenvolve uma política que auxiliada/implementada pela tecnologia aumenta a privacidade do sistema como um todo (quem pode acessar o que, em que ordem e os processos que devem ser executados no decorrer da comunicação).

2.6.2 Detecção de Intrusão em Redes

Atualmente, muitos sistemas de detecção vem sendo propostos utilizando técnicas de aprendizado de máquina Cassales *et al.*, 2019; Elias *et al.*, 2022; Arbex *et al.*, 2021; Shi *et al.*, 2021. Fazendo com que eles estejam relacionados com os mesmos problemas antes citados. Deseja-se então encontrar um meio termo em que se garanta a segurança dos usuários (um sistema bem treinado poderá detectar com mais precisão possíveis pacotes maliciosos), ao mesmo tempo que se mantenha a privacidade daqueles que compartilharam seus dados. Esses dados são originados a partir do uso da rede, contendo portanto diversas informações sigilosas Elias *et al.*, 2022.

Embora se defenda que muito dos pacotes transitando sejam criptografados e, portanto, sua informação já esteja protegida, há abordagens de uso de IDS em que essa criptografia é retirada para que a detecção possa ser realizada. Isso viola os princípios de sigilo dos protocolos da Internet e expõe os dados dos usuários para um agente terceiro ("*man in the middle*"). Os trabalhos Sherry *et al.*, 2015; Canard *et al.*, 2017 buscam então criticar tais abordagens e encontrar formas de se manter uma alta precisão de classificação ao mesmo tempo que mantenha a codificação dos pacotes.

Por fim, em Zhang e Zhu, 2018 é discutido um adversário que possui acesso constante às saídas do programa de detecção. Nesse contexto é discutido que a adição e remoção de instâncias de treinamento pode modificar as saídas do programa, revelando informações dessa instância. Dessa forma, assim como brevemente discutindo em Bal *et al.*, 2014 o sistema visa proteger não só as informações diretamente pessoais mas também aquelas que serviriam como base de inferência contribuindo para um vazamento de informação.

Esses textos nos dão um panorama de quais são as principais preocupações dos pesquisadores, em termos de privacidade em IDS, e quais as propostas que estão sendo feitas em relação a isso. Assim como na área de aprendizado de máquina, grande ênfase está sendo dada ao volume de dados de treinamento, sendo várias das estratégias em relação a esses dados. Porém, devido à própria natureza dos dados da rede, também está sendo observado o fluxo de pacotes utilizados no ato da detecção.

Capítulo 3

Metodologia proposta

Os IDSs, junto com diversos outros mecanismos, são usados para proteger redes de computadores que têm acesso à Internet. Ao usar uma rede, todas essas ferramentas são acionadas automaticamente permitindo que o usuário desfrute de alta segurança. Entretanto, muitas vezes, o usuário não sabe que esses mecanismos estão sendo usados e quais dos seus dados estão sendo coletados para que eles funcionem.

Hoje, a relação com os usuários e a segurança da rede é obscura. Se por um lado os mecanismos estão lá para, primeiramente, garantir a segurança e bom uso da Internet, por outro, não existe qualquer forma de consentimento de que o usuário terá seus pacotes rastreados ou a difusão de informação sobre as práticas de segurança aplicadas. Essa estrutura está em prática desde os primórdios da segurança de rede como uma forma de penalizar os adversários maliciosos que desejam atacar a rede.

Nessas condições, neste trabalho foi analisado que aspectos como autorização e consentimento do usuário sob a utilização dos dados seriam inaplicáveis visto que a estrutura da rede não permite. Além disso, a eficiência de pedidos de autorização para aumentar a privacidade dos usuários já vem sendo questionada. Assim tais aspectos e a divulgação de informações, se empregados, devem ser de responsabilidade do agente controlador da rede, sendo o IDSs apenas mais um mecanismo utilizado por ele.

Dessa forma, neste trabalho analisamos a privacidade no que diz respeito aos dados utilizados e a cadeia de funcionamento do IDS. A relação do usuário como agente modificador, ou seja, que seja capaz de modificar aspectos do IDS, não foi considerada, uma vez que retirar ou modificar mecanismos de segurança em prol da privacidade, na verdade, deixaria a rede mais exposta e menos segura, o que foge do objetivo principal dos IDSs. Assim, recomendamos que as técnicas empregadas para proteger a rede sejam divulgadas de forma geral pelos próprios controladores da rede, podendo, inclusive, utilizar este estudo para validar a necessidade do uso de informações e mostrar as medidas tomadas em relação à privacidade dos pacotes usados. Aqui, no entanto, nos basearemos apenas nos dados utilizados pelo IDS.

Ademais, estamos considerando, neste estudo, um adversário que tem como objetivo central atacar o IDS, já que esse sistema, a principio, possui acesso integral aos pacotes entrantes da rede. Dessa forma, consideramos mais privado um IDS que, sob ataque, revele

o mínimo de informação possível.

Adversários internos, que já têm acesso à rede local, ou que desempenham funções que lhes dão acesso a outras informações da rede não foram considerados. O objetivo do trabalho é analisar a privacidade em termos dos dados que se têm acesso exclusivamente pelo IDS.

Os principais riscos portanto estarão associados ao vazamento de informação que o IDS pode prover a um adversário. Então, como mencionado anteriormente, o consentimento do usuário não será amplamente discutido, uma vez que a própria estrutura da rede de Internet atual não permite que o usuário tenha controles decisivos sobre essas informações.

Dessa forma esse estudo auxilia na discussão de qual é a informação mínima necessária (garantindo a máxima privacidade) que não interfira na precisão/acurácia do IDS (garantindo máxima segurança). É importante ressaltar que, como já visto em outros contextos, informação mínima não diz respeito somente à quantidade, mas também a sensibilidade e relevância da informação, aspectos que se provaram complexos de serem medidos.

3.1 Adaptação do Secure Tropos

Utilizamos como base o trabalho Kioskli *et al.*, 2022 e a metodologia Secure Tropos para analisar e metrificar um IDS. A metodologia se baseia principalmente nos agentes e objetivos do sistema, além de, originalmente, ser uma ferramenta para o planejamento e design do projeto, descrevendo, por exemplo, os mecanismos e funcionalidade que o programa deve ter. Em nosso contexto, vamos adaptá-la para analisar um sistema de detecção de intrusão. Dessa forma, retiramos a primeiro visão (Organizacional) uma vez que não há a necessidade em separar os agentes em organizações já que eles são apenas dois (administrador da rede e treinador do modelo). Além disso, na Visão PbD não iremos expor os mecanismos pelos quais uma restrição está sendo cumprida ou as medidas sendo tomadas, já que queremos que esses diagramas sejam genéricos o suficiente para representar qualquer IDS baseado em aprendizado de máquina. Assim, no lugar das medidas e mecanismos que garantem uma restrição, adicionamos um elemento que representa a métrica usada para calcular quanto dessa restrição está sendo respeitada pelo sistema, ou, em outras palavras, quão eficiente estão sendo os mecanismos empregados, mas sem necessidade de sua especificação.

Neste estudo, embora os IDSs já tenham sido construídos e possam não ter sido elaborados seguindo diretrizes do *Privacy by Design*, queremos analisar quão bem eles lidam com os aspectos de risco mapeados. Dessa forma, criamos diagramas que mostram os possíveis pontos de risco e restrições dos recursos para então analisarmos individualmente como esses pontos estão sendo tratados no sistema já pronto.

3.2 Visão de Fluxo de Dados

O primeiro diagrama (Figura 3.1), não faz parte da metodologia do Secure Tropos, ele mostra o fluxo dos dados que são utilizados pelo IDS e/ou que passaram para dentro da rede interna. Em um IDS baseado em aprendizado de máquina, teremos dois momentos: (i)

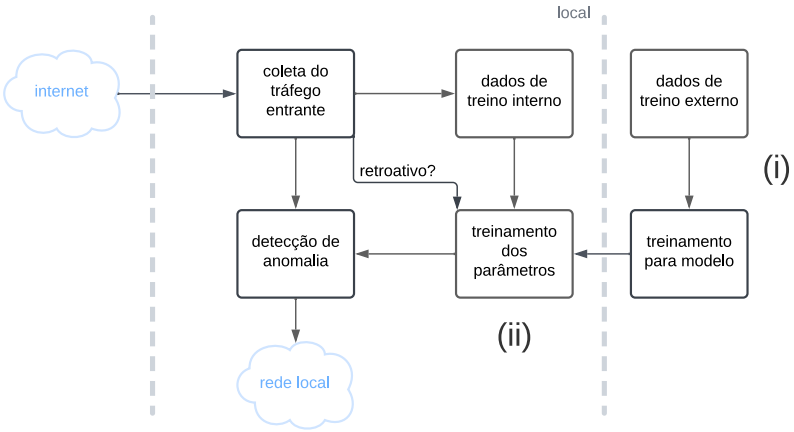

Figura 3.1: *Visão de Fluxo de Dados no IDS*

escolha e treino do modelo de detecção e (ii) treino dos parâmetros do modelo para a rede local. Do lado direito da figura mostramos a etapa (i) em que dados de treino externo (fora da rede local) serão usados para treinar o modelo de detecção de anomalias. Essa etapa é geralmente feita por alguém fora da rede local, por exemplo pelo fornecedor do IDS, e conta com diversas etapas de treinamento e validação, que visam mostrar que o sistema é eficiente em relação a segurança, detectando uma boa porcentagem das anomalias identificadas no conjunto de dados. Os dados de treinamento usados costumam ser públicos, podendo ser reais ou gerados artificialmente. Após o treino do modelo será feita a etapa (ii) para que o sistema possa ser implantado localmente. Assim, dados da rede externa que desejam entrar para rede interna são coletados e usados para treinamento dos parâmetros do modelo. Tal etapa é importante para garantir que os parâmetros do modelo reflitam corretamente a realidade dos dados entrantes da rede local. Após o treinamento, o IDS está pronto para ser usado encontrando possíveis anomalias que desejam adentrar a rede interna. Além disso, embora só seja necessário ambientar o modelo uma vez, pode-se ter regularmente essas duas etapas de treinamento de forma que novas anomalias e rotinas sejam bem detectadas pelo sistema. Nesse caso, os dados entrantes da rede são utilizados de forma retroativa no treinamento.

Ao construir o diagrama da Figura 3.1 percebemos que o IDS lê uma ampla quantidade de pacotes entrantes da rede, possibilitando que diversas informações pessoais e confidencias dos usuários sejam acessadas. Tal situação pode trazer sérios riscos à privacidade do usuário. Pensando nessas questões e nos princípios defendidos pela LGPD em relação ao tratamento de dados as seguintes perguntas surgem:

Todos os dados coletados são de fato necessários ao funcionamento do IDS? O ideal é que apenas os dados usados pelo IDS fossem coletados, e, mais do que isso, que o IDS requisitasse o mínimo de dados possíveis para fazer a classificação, já que, além de dados maliciosos, ele também tem acesso a dados normais.

Como identificar que dentro do IDS, onde estão sendo analisados os dados entrantes, não está havendo uma utilização indevida (*re-purpose*) dos dados? Essa pergunta é mais difícil de responder e foge do escopo do trabalho. Entretanto, algumas medidas para limitar a utilização imprópria dos dados podem ser discutidas, por exemplo di-

minuir os dados acessados pelo IDS, possibilitar a análise do código, ter uma documentação ampla e medidas burocráticas contra o abuso desses dados.

No caso de aprendizagem retroativa, o armazenamento dos dados poderia trazer mais risco à privacidade? Se por um lado o armazenamento de dados pessoais poderia criar mais um ponto de risco à privacidade, por outro existem algumas técnicas de aprendizado online em que os dados ficam apenas provisoriamente armazenados em memória Arbex *et al.*, 2021. Assim, problemas sobre o armazenamento de um grande volume de dados pessoais podem ser evitados.

Qual o impacto da privacidade dos dados de treino externos no sistema final? Embora os dados de treino externo, a princípio, não digam respeito aos usuários da rede local e portanto não impactariam na privacidade dos dados deles, a etapa de treinamento do modelo está dentro da cadeia de funcionamento do IDS e portanto deve ser considerada. Esse estudo e discussão sobre o impacto total desses dados, no entanto, também não serão amplamente tratados nessa pesquisa.

3.3 Visão de Dados

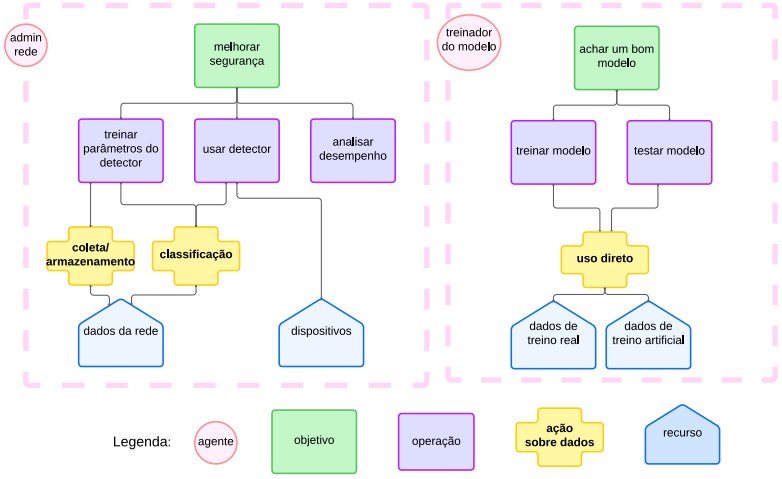

Figura 3.2: *Visão de Dados*

A Figura 3.2 mostra um dos diagramas baseados na metodologia Secure Tropos. Nele evidenciamos agentes, seus objetivos e os recursos disponíveis. Nesta etapa, os elementos e suas descrições podem ser abstratos, traduzindo principalmente a ideia dos objetivos e operações que se deseja alcançar. O ponto principal é mostrar, no formato do Secure Tropos, a utilização dos dados (que são nossa maior preocupação) no sistema descrito na Figura 3.1. Essa fase é importante para construir a base de análise da próxima visão, em que novos elementos serão inseridos. Na legenda da Figura 3.2 embaixo, o agente referencia a pessoa que deseja alcançar um certo objetivo e esse objetivo é alcançado fazendo operações que podem levar à execução de uma ação sobre um recurso. No caso de um IDS, temos que um dos agentes é o administrador da rede (à esquerda da figura 3.2) que, desejando melhorar a segurança, implanta e gerencia o IDS. Para isso ele deve treinar os parâmetros do modelo do IDS para sua rede local, colocá-lo em uso e, posteriormente, fazer

uma análise de desempenho. Assim ele entra em contato com os dados entrantes da rede, tendo de armazená-los para treinamento e usá-los no IDS. Outro recurso acessado pelo administrador são os dispositivos em que o IDS será instalado. À direita da figura temos o agente treinador do modelo, que tem por objetivo achar um bom modelo de detecção de intrusão que seja capaz de detectar com precisão as anomalias. Para isso ele deve treinar e testar o modelo utilizando dados reais e/ou artificiais.

O estudo dos recursos na Figura 3.2 e como eles são utilizados é essencial na análise de privacidade de um sistema. Em especial, no caso dos dados de um IDS, uma parte dos dados utilizados contém informações pessoais dos usuários da rede Elias *et al.*, 2022 e portanto faz-se necessário compreender cada área que tem contato com os recursos para que seja feita a análise do risco relacionado a cada um deles. É importante expor as ações sobre os dados e os objetivos para que possamos identificar possíveis riscos atrelados a esses acontecimentos e, não só isso, mas averiguar que os dados estão sendo estritamente usados para isso.

3.4 Visão de *Privacy by Design* (PbD)

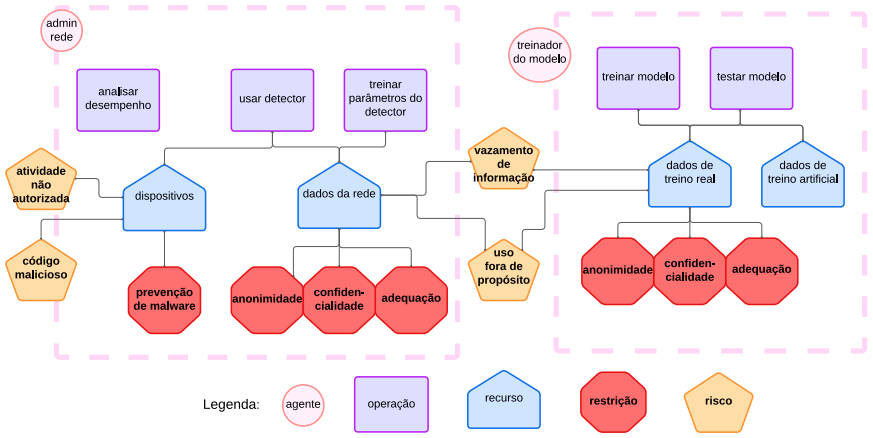

Figura 3.3: *Visão de Privacy by Design*

O diagrama da Figura 3.3 expõe riscos e restrições relacionados a cada recurso. Riscos são acontecimentos desencadeados por ações externas ao sistema trazendo algum efeito prejudicial para os usuários. Restrições são características dos recursos internos do sistema que garantirão que estes não sejam mal utilizados e/ou utilizados fora do cumprimento de alguma política (neste trabalho as políticas principais são a LGPD, GDPR e os princípios de PbD). O não cumprimento de uma restrição gera uma chance maior de um ou mais riscos ocorrerem. Na metodologia de Secure Tropos, as medidas e os mecanismos tomados para garantir que essas restrições sejam cumpridas também devem ser especificados no diagrama, pois ele é responsável por guiar as próximas etapas de elaboração do sistema. No caso de nosso estudo entretanto, tais medidas não precisam ser especificadas. Posteriormente, utilizaremos apenas sua eficiência na métrica.

No diagrama, temos que os dados entrantes da rede e os dados reais de treinamento possuem riscos semelhantes: serem usados fora do propósito a que foram coletados e terem informações pessoais expostas. Alguns riscos e restrições sobre o dispositivo também são

exibidos, embora a análise mais profunda sobre essas característica são mais amplas que a área de pesquisa desse projeto. É importante ressaltar que a segurança do dispositivo em que o sistema se propõe a funcionar é de extrema relevância para a privacidade dos dados uma vez que o uso incorreto pode contribuir para que a privacidade dos dados seja comprometida. As restrições ligadas aos dados querem garantir que a privacidade dos indivíduos, cujas informações estão dentro desses pacotes de rede, seja mantida. Dessa forma os dados devem ser anônimos, confidenciais, conterem apenas as informações adequadas a que serão usadas e serem usadas somente para isso. Essas restrições são comuns ao se tratar de dados pessoais e/ou sensíveis.

Nos dados da rede local, garantir o anonimato de um pacote significa garantir que a identidade do usuário da rede, cujas informações estão contidas neste pacote, não seja descoberta durante sua utilização. Essa identidade pode ser descoberta em diversas partes dos pacotes de rede analisados pelo IDS. Já no caso dos dados de treino, assim como em aprendizado de máquina o anonimato diz respeito a proteger a identidade do proprietário dos dados, informação que pode estar contida tanto no pacote quanto no fluxo de compartilhamento entre treinador e fornecedor.

O conceito de confidencialidade é bem conhecido na área de segurança, sendo uma forma de garantir que apenas as pessoas autorizadas tenham acesso a uma certa área do sistema. Em privacidade, esse conceito é estendido abrangendo também a escolha e transparência sobre quem possui essa autorização. Assim, no caso dos dados da rede local, para garantir a confidencialidade é preciso mecanismos que informem os usuários sobre quem são as pessoas autorizadas e assegurar que apenas essas pessoas autorizadas terão de fato acesso ao IDS e suas informações, garantindo um armazenamento seguro dos dados e a comprovação de que toda a vida dos dados (coleta, uso e descarte) esteja sendo feita de forma segura e privada. No caso dos dados de treino teremos a mesma ideia, porém no contexto de treinador e fornecedor, assim além de assegurar a privacidade ao armazenar, coletar, usar e descartar, é preciso também cuidado no transporte dos dados entre os dois atuadores.

Por fim, a restrição de adequação expressa a quantidade de informação utilizada para que a detecção e o treino sejam realizados. Aqui, numa primeira abordagem, estaremos considerando a quantidade de informação em número absoluto, sem considerar sua sensibilidade. Assim, no caso dos dados da rede local, ao passar pelo detector, quanto do pacote precisa ser lido para que a detecção funcione. Similarmente, queremos saber quanta informação o fornecedor dos dados precisa dar ao treinador para que o treinamento seja executado. Posteriormente é interessante que haja um estudo mais profundo sobre a sensibilidade das informações contidas nos pacotes de rede e como elas impactam na adequação às finalidade do IDS.

Embora exista grande semelhança entre os dados da rede (local) e os dados de treino real (externos), existem duas diferenças consideráveis. A primeira é sobre a forma de entrada dos dados. No caso dos dados usados pelo IDS (dados da rede local), teremos um fluxo contínuo de pacotes. Já no treinamento, é considerado um conjunto estático de dados, uma vez que este treinamento se insere no contexto de aprendizado de máquina supervisionado. A segunda é a gravidade dos riscos e restrições. De um lado, temos os dados pertencentes aos usuários da rede local que prezam por sua segurança e privacidade.

Do outro, temos dados cuja procedência pode ser múltipla, sendo muitas vezes inclusive pública. A privacidade desses dados também deve ser levada em conta no cálculo da privacidade geral do sistema, mas é inegável que tenham um impacto menor já que não comprometem diretamente a privacidade dos usuários do IDS.

Um ponto importante sobre esses diagramas é que, embora tenham sido feitos para metrificar um IDS já existente, eles podem ser úteis para projetar sistemas que desde o início tenham uma preocupação de Privacy by Design, se preocupando para que as métricas sejam boas quando o sistema estiver pronto, já que eles foram baseados em métodos de desenvolvimento de sistema.

3.5 Planejamento da Métrica

Uma vez tendo feito os estudos acima, pontuando as principais restrições que queremos que os recursos tenham para que a privacidade do sistema seja garantida, a métrica geral do sistema será calculada a partir de métricas específicas para cada restrição. Uma visão geral de como funcionaria a métrica pode ser vista na Figura 3.4. Dessa forma, é importante observar que, diferente de outras abordagens como Agarwal, 2016, a nossa métrica é baseada no cumprimento das restrições e não diretamente no risco. Aqui entendemos que o risco é um fator primordialmente externo e que pode ser incentivado e/ou facilitado pelo não cumprimento de uma restrição, aspecto sob controle interno. Por esse motivo, é mais tangível tratar sobre as restrições e observar seu cumprimento no sistema. Assim, consideramos que, ao cumprir completamente todas as restrições, a possibilidade do risco ocorrer e haver uma perda de privacidade do usuário da rede é reduzida.

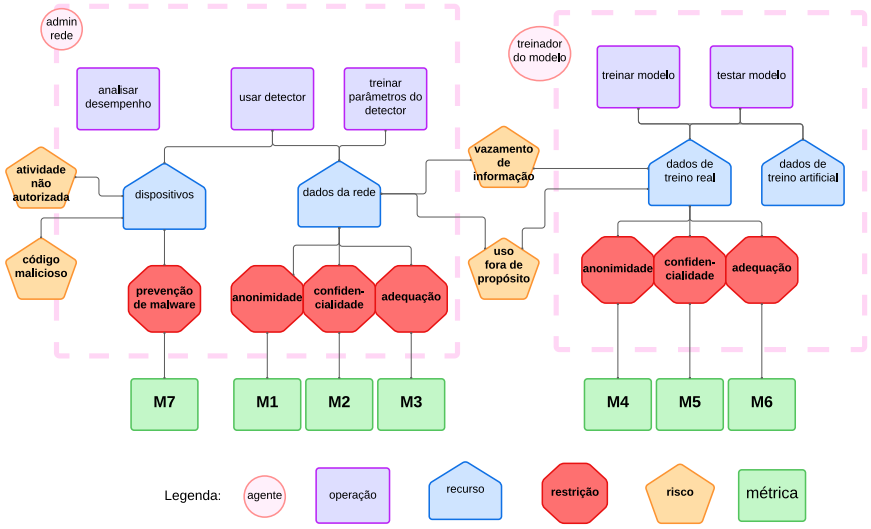

Figura 3.4: *Visão Geral da Métrica*

O passo seguinte portanto é identificar quais métricas podem ser usadas para medir o cumprimento de cada restrição. Uma restrição pode ter mais de uma métrica, sendo necessário um estudo mais aprofundado para definir qual é a melhor para ser usada. O estudo sobre a anonimização de conjunto de dados já está bem avançado Wagner e Eckhoff, 2018, sendo assim, podemos utilizar essas métricas para analisar a anonimidade

dos dados de treinamento (M_4), incorporando esses valores na nota final do sistema. Entretanto, é importante pontuar que muitas dessas métricas são dependentes da forma como a anonimização foi realizada, impedindo que a métrica seja usada em qualquer contexto. Sobre a restrição de prevenção de malware (M_7) é possível utilizar estudos feitos em análise de segurança de sistemas, área bem desenvolvida em engenharia de software, que porém estão fora do contexto desse trabalho. No caso da adequação dos dados da rede (M_3), uma métrica possível seria a porcentagem do pacote a ser lido, permitindo analisar quanta informação é necessária para detecção. Para as outras restrições no entanto ainda é necessário uma busca mais assertiva sobre formas de medir seu cumprimento.

Por fim, tendo as métricas para cada restrição (M_1 - M_7) precisará ser feito um estudo sobre qual a importância real de cada uma delas, para então serem atribuídos pesos que serão utilizados para calcular a métrica final. A métrica final M será uma função

$$M = \sum_{i=1}^{7} p_i M_i$$

em que $\sum_{i=1}^{7} p_i = 1$ e p_i é proporcional à importância que a restrição relacionada a métrica M_i tem na privacidade geral do sistema.

Capítulo 4

Aplicação da Metodologia

Neste capítulo discutiremos como a metodologia pode ser usada para analisar sistemas de detecção de intrusão. Os dois sistemas escolhidos para analise já têm como avanços tecnológicos propostos formas diferentes de melhorar a privacidade do IDS. Nos restringiremos a apresentar os principais aspectos de privacidade atingidos por cada trabalho, listando as restrições que são afetadas. Ainda não há uma métrica para cada restrição, mas quando houver, elas deverão refletir as mudanças que esses IDS implementam em comparação com outros IDS. As análises aqui feitas contribuem para mapear os pontos de melhorias que os sistemas propõem e expor as restrições ainda não respeitadas, direcionando futuros esforços de desenvolvimento e atualização dos respectivos projetos.

4.1 IDS Baseado em Informações das Camadas Inferiores

Em Elias *et al.*, 2022 temos um IDS que se dedica a apenas utilizar as camadas inferiores para fazer a detecção de intrusão. Nesse caso estaremos melhorando a privacidade dos dados que serão analisados pelo IDS antes de entrar na rede local.

Como o detector só precisa ter acesso aos dados das camadas inferiores da Arquitetura da Internet (transporte e rede), o grau de adequação dos dados da rede usados será maior, impactando essa restrição, como apresentado na Figura 4.1. Assim, com a utilidade (acurácia) apresentada no artigo, foi mostrado que apenas esses dados são necessários para a detecção, sendo qualquer outro dado considerado excedente. Ao retirar a camada de aplicação, muito das informações sensíveis são protegidas, entretanto é importante observar que ainda há a necessidade de se usar a camada física, que pode contribuir para o vazamento de informações sensíveis, por exemplo a localização do usuário e sua rotina. Essa informação pode ser deduzida, por exemplo, ao observar o aumento no uso da rede em determinados horários do dia. Sobre os outros aspectos da privacidade, não foram observados mecanismos que garantissem essas restrições.

Esse trabalho, mostra um importante avanço no quesito de adequação uma vez que

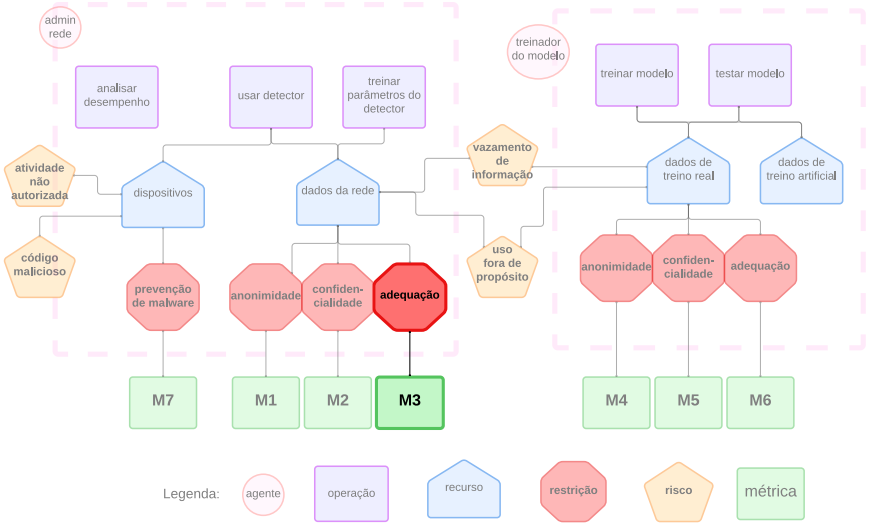

Figura 4.1: *Restrições Afetadas pelo IDS* Elias *et al., 2022*

mostra que, em média, 88,68%[1] menos informação é necessária na detecção mantendo-se um alto nível de acurácia. Essa diminuição de informação necessária foi analisada em termos de tamanho. O estudo sobre a influência da remoção dessas camadas na sensibilidade das informações utilizadas pelo IDS ainda não foi considerado, já que esse cálculo está fora do escopo do projeto devido a sua alta complexidade.

4.2 IDS Baseado em Aprendizado Federado

O segundo IDS que iremos analisar é o Shi *et al.*, 2021, que utiliza a técnica do aprendizado federado para treinar o detector. Nessa técnica, ao invés de se compartilhar os dados entre fornecedor e treinador, apenas o modelo e seus parâmetros serão compartilhados, ou seja, não haverá compartilhamento direto de nenhum dado. Assim o treinamento é feito observando os parâmetros e acurácia dos testes executados pelo fornecedor sem que o treinador tenha que olhar ou transportar os dados. As únicas informações transportadas serão as referentes aos testes e os modelos de treinamento.

Nessas condições, os dados de treino ficam protegidos, e não só o anonimato é garantido, já que o treinador não sabe de quem os dados utilizados são, mas também a confidencialidade já que não haverá risco seja no transporte ou acesso, uma vez que apenas o fornecedor irá manejar os dados. Sobre a adequação, ela é garantida pois apenas o modelo e seus resultados serão fornecidos ao treinador, informações inquestionavelmente necessárias ao treinamento. Tais relações são demonstradas na Figura 4.2

Existem no entanto algumas críticas ao método federado Yin *et al.*, 2021. Ainda não se tem informações suficientes para comprovar que apenas o transporte do modelo garante privacidade máxima aos dados de treino. Está em debate se não seria possível inferir

[1] Esse valor é uma aproximação que considera que o principal protocolo de camada de aplicação que usa TCP é o HTTP e que o principal protocolo da camada de aplicação que usa UDP é o DNS Garsva *et al.*, 2015. Dessa forma, 1380 bytes da camada de aplicação em um pacote HTTP corresponde a 91,15% do pacote total e 512 bytes de um pacote DNS corresponde a 86,20% do pacote total, o que leva à média de 88,68%

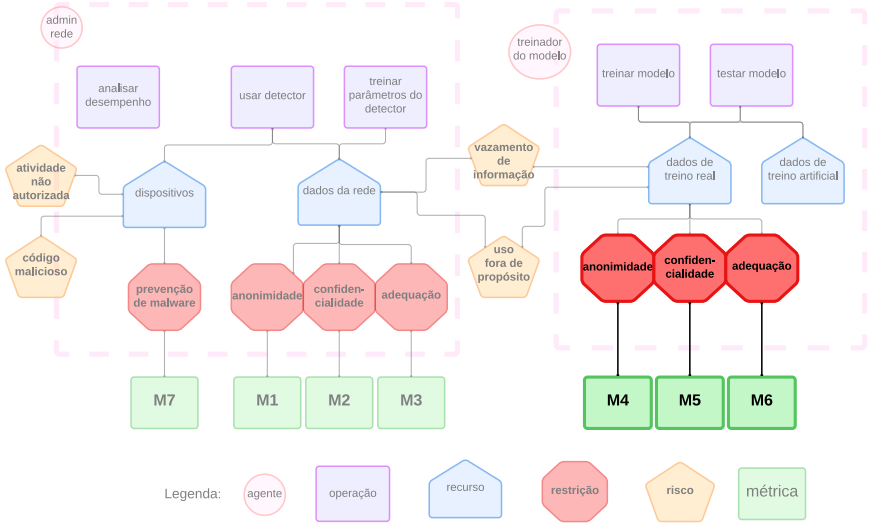

Figura 4.2: *Restrições Afetadas pelo IDS Shi et al., 2021*

informações sobre os dados de treino com as respostas do aprendizado federado, colocando a confidenciabilidade e adequação em risco. Mesmo assim, esta técnica garante que os dados de treino não fiquem diretamente expostos ao treinador, como no caso do treino supervisionado convencional, aumentando a privacidade dos dados.

Nesse sistema, a proposta de se melhorar a privacidade está na etapa e dados de treinamento, no entanto sobre os dados da rede em que será inserido o detector nenhuma melhoria é feita.

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho agrupamos, revisamos e discutimos diversos artigos sobre privacidade, análise de sistemas, aprendizado de máquina e IDSs, transformando esses aprendizados em diagramas e planejando uma métrica de privacidade. Os diagramas construídos contribuem para identificar os principais aspectos de privacidade que devem ser considerados ao desenvolver um IDS. Além disso, eles pontuam em que momento os riscos surgem e quais as restrições que se devem ter sobre os dados para que a ocorrência deles possa ser diminuída. A métrica resultante seria então baseada nesses diagramas e análises, considerando todas as restrições pontuadas. Este trabalho difere dos outros textos estudados em dois principais aspectos: contexto e abrangência do sistema. Não foram encontrados outros trabalhos que tratem de analisar um sistema de segurança de rede com foco na privacidade. Os artigos sobre IDSs, embora tivessem propostas que focavam em aumentar a privacidade, não apresentavam maneiras de se analisar e/ou medir essas contribuições. Além disso, embora já houvessem trabalhos que tratassem de medir a privacidade dos dados usados para treinamento de sistemas baseados em aprendizado de máquina, poucas referências que inserissem esse estudo no produto final do sistema (não apenas considerando seu treinamento) foram encontrados.

Assim, a principal contribuição deste trabalho foi explorar uma área ainda pouco desenvolvida em segurança e privacidade da rede, unindo visões da legislação, metodologias da engenharia de software e técnicas de privacidade e produzindo uma análise de privacidade de um IDS bem como um planejamento de métrica para medi-la. Essa metodologia foi utilizada para analisar duas propostas de IDS que se dedicam a aumentar privacidade. Reconhecemos que, por seu caráter fortemente interdisciplinar, o nosso panorama focado em segurança e computação foi limitante ao desenvolvimento e reconhecimento de certos aspectos de privacidade Rocha e E. Canedo, 2023. Por isso, apontamos como principal trabalho futuro um estudo mais aprofundado sobre elicitação de restrições de privacidade Ferrão, 2022, para que haja máxima concordância com as legislações (área de ensino do direito) e de utilidade do sistema (área da engenharia de software).

Além disso, é necessário ainda buscar formas de estimar quanto de cada restrição é cumprida pelo sistema. De acordo com essa medida, um valor geral para o sistema pode ser calculado. Para isso, no entanto, também será necessário saber quão importante é cada restrição no sistema geral e achar pesos que balanceiem bem essa relação. Por fim,

é necessário que mais sistemas sejam analisados e metrificados pela metodologia aqui proposta e seus resultados observados por desenvolvedores, possibilitando que a questão entre segurança e privacidade seja bem balanceada e que ao tentar maximizar a pontuação dada por essa métrica, um sistema bom para segurança e privacidade dos usuários seja de fato alcançado.

Disseminação dos Resultados e Dúvidas

Resultados parciais deste trabalho foram apresentados no Workshop Interscity 2023 [1]e no Workshop de Trabalhos de Iniciação Científica e de Graduação (WTICG) do Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg) 2023, sendo, nesse segundo, publicado nos anais da conferência e referenciado da seguinte forma:

- SATO, J. Y. N. ; BATISTA, D. M. . Modelagem das Áreas de Risco de Sistemas de Detecção de Intrusão para Cálculo de Métricas de Privacidade. In: Workshop de Trabalhos de Iniciação Científica e de Graduação (WTICG) do SBSeg, 2023. Anais dos Workshops do Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg), 2023.

Nesta sessão traremos algumas dúvidas e sugestões que surgiram durante essas apresentações:

- *Privacidade Diferencial*

  Foi perguntado se o trabalho teria alguma relação com ou se poderia ser utilizada privacidade diferencial Dwork, 2006. Privacidade diferencial, foi primeiramente proposta no contexto de banco de dados estatísticos Adam *et al.*, 2011, querendo reconhecer quão impactante é a entrada de um elemento nas respostas do sistema. Assim, ao ser realizada uma consulta, por exemplo, queremos saber quão diferente serão as respostas devolvidas no caso de um conjunto com o elemento e outro sem. Quanto maior a diferença, menos privado será o sistema já que ele expõe a existência ou não desse elemento. Uma forma de se melhorar a privacidade nesse sentido seria adicionar ruído às respostas, de forma que a diferença no conjunto usado não seja tão fortemente percebida.

  No nosso caso entretanto, ao pesquisar sobre o assunto, não encontramos formas imediatas de se usar essa técnica uma vez que estamos lidando principalmente com fluxo contínuo de dados (no IDS) ou conjuntos já prontos (durante o treinamento) tornando a análise dada pela privacidade diferencial ineficiente.

- *Qual atacante?*

  Como descrito no Capítulo 3, estamos lidando com um atacante externo que tem como principal objetivo retirar as informações do IDS, sem informações prévias. Nesse caso, portanto, a ênfase da análise é a quantidade e sensitividade dos dados utilizados pelo IDS, uma vez que, sob ataque, esses dados estariam expostos.

[1] InterSCity é um projeto de pesquisa colaborativo financiado pelas agências de fomento CAPES, CNPq e FAPESP – https://interscity.org/

- *Influência do usuário da rede*

  Outra questão levantada foi a influência dos usuários. Esta dúvida está intimamente ligada ao fato de que na legislação a privacidade é considerada como o poder do usuário em reger as regras sob as quais seus dados serão tratados, e, portanto, espera-se que existam formulários de autorização e mecanismo para autenticar, corrigir e limpar os dados que o sistema irá usar. Porém, no contexto de redes, não existe infraestrutura para tal, sendo inviável para um IDS pedir autorização por cada pacote que irá analisar, por exemplo. Além disso, sobre autorização, existem trabalhos que mostram que os usuários muitas vezes não leem, não entendem e não verificam esses formulários, tornando-os ineficientes e diminuindo a utilidade do sistema Assim, este trabalho foi regido pela ideia da informação mínima, ou seja, estaremos analisando sobretudo o princípio de necessidade e adequação dos dados utilizados.

- *Criptografia*

  Foi apresentado o ponto de que muitos dos pacotes hoje transportados pela rede estão criptografados e então já estariam protegidos. Já foi mostrado que pode haver vazamento de dados mesmo quando eles estão criptografados Wang *et al.*, 2015; Atkinson *et al.*, 2018. Nesses casos, é possível identificar quais os aplicativos de celular estão sendo usados pelos usuários da rede, trazendo risco a sua privacidade.

  Além disso, alguns trabalhos apresentados no capítulo 2 discutem como o IDS pode agir como um "*man in the middle*", retirando a codificação do pacote para fazer a detecção. Dessa forma, mesmo que criptografados, queremos saber sobre os dados que o IDS está utilizando e analisar quais os riscos eles podem trazer ao usuário da rede.

- *Privacidade x Segurança*

  Outro ponto levantando foi que a detecção de intrusão já é uma tarefa difícil por si só. Dessa forma, ao considerar também as restrições de privacidade, elas poderiam se tornar um fator limitante do detector diminuindo sua acurácia e funcionando a favor do atacante. Neste trabalho nos propomos a analisar IDS que já se dispõem a aumentar a privacidade, dessa forma, esse estudo pode na verdade ser usado a favor dos desenvolvedores e usuários que podem verificar se as mudanças feitas em prol da privacidade que debilitam em algum grau a segurança, de fato são vantajosas para a rede. Entretanto, é inegável que existe uma questão entre a usabilidade e segurança do sistema e a privacidade dos dados usados. Aqui está em foco a privacidade, mas incentivamos que, em paralelo, as medidas de acurácia e precisão do detector também sejam levadas em consideração. Essas medidas são em sua maioria as que regem o desenvolvimento do IDS e portanto são facilmente encontradas nas propostas de cada sistema.

- *Influência do usuário da métrica (administrador da rede, desenvolvedor)*

  Por fim, dois pontos sobre os usuários da métrica foram comentados. Primeiro, sobre a confiança desse usuário. Estamos considerando que ele é confiável e utilizará o sistema da forma que ele foi proposto. Essa consideração foi feita pois, caso contrário, não seria possível ter qualquer nível de privacidade sobre o sistema, uma vez que ele

necessita ser implantando e mantido por algum agente. O segundo ponto foi que as restrições escolhidas e as métricas usadas para medi-las poderiam vir dos próprios administradores e desenvolvedores. A princípio essa seria uma boa solução, uma vez que, no contexto de segurança de redes, pouca informação sobre requisitos de privacidade e suas métricas estão disponíveis. Porém, para que essa métrica seja usada para comparação, ela deve ter requisitos e métricas (específicas para cada um deles) constantes para todos os sistemas, portanto é necessário fixar quais seriam esses elementos a fim de produzir uma nota passível de comparação.

Agradecemos a oportunidade de ter participado desses eventos, à atenção e às perguntas durante as apresentações

Apêndice A

Métricas de Preocupação com a Privacidade

Ao se buscar artigos sobre métricas de privacidade, muitos trabalhos tratam de métricas de preocupação com a privacidade Preibusch, 2013; Martin e Nissenbaum, 2017; Braunstein *et al.*, 2011; Knijnenburg *et al.*, 2017. Nesse caso, a área de pesquisa se torna fortemente interdisciplinar, se relacionando principalmente com a psicologia e o direito. O foco, deixa de ser o sistema ou projeto, mas passa a ser a percepção do usuário sobre eles. Assim, para que a nota de privacidade seja alta, não basta ter funcionalidades eficientes que garantam a privacidade, mas que essas sejam conhecidas pelo usuário.

Temos então que a métrica passa a ser dependente do contexto a que está inserida, das informações passadas ao usuário e da opinião do usuário sobre ambos esses aspectos. Fazer essa medição torna-se especificamente complexo ao se considerar o conceito de Paradoxo da Privacidade. Esse paradoxo descreve o comportamento incompatível dos usuários em dizer que são preocupados com a privacidade porém têm atitudes que colocam ela em risco no dia-a-dia, mostrando que essa preocupação está mais no discurso pontual sobre o assunto do que em práticas diárias. Isso faz com que a métrica de percepção seja mais estrita do que a realidade, uma vez que ao serem perguntados sobre a privacidade, sua preocupação durante a resposta será maior do que ao se utilizar o sistema.

Dessa forma, alguns trabalhos focam em entender qual é a melhor maneira de abordar os entrevistados, de forma que as respostas traduzam fielmente sua percepção diária sobre o sistema. Alguns deles inclusive estudam o impacto da palavra "privacidade" nessas métricas, daí sua interdisciplinaridade. Assim, essas pesquisas não estão relacionadas com a eficiência das funcionalidades do sistema em si, mas sim de sua relevância no ponto de vista do usuário.

Sendo assim, embora possuam o mesmo nome ("métricas de privacidade"), essas métricas relacionadas a percepção do usuário estão pouco relacionadas com os intuitos desta pesquisa. O principal objetivo deste trabalho é que as métricas sejam um auxílio aos desenvolvedores, sendo possível analisar e comparar diferentes sistemas em termos de privacidade, permitindo que sejam identificadas mudanças possíveis ao sistema. No caso das métricas de preocupação, por outro lado, o parecer resultante é pouco proveitoso em

termos de análise e atualização de tecnologias do sistema.

Entretanto, o conhecimento sobre esse tipo de métrica permite entender a importância de comunicar ao usuário sobre os mecanismos que garantem a privacidade e os intuitos e procedimentos que o recurso dado por ele serão submetidos. Esses princípios estão inclusive presentes nas legislações que regem o tratamento de dados pessoais (GDPR e LGPD), na medida que elas defendem que as pessoas têm direito em saber como e para que seus dados serão utilizados e ter ferramentas que permitem validar esse uso.

Apêndice B

Tabela de Síntese dos Trabalhos

Na Tabela B.1 os trabalhos estão divididos de três maneiras: colunas, linhas e cores. As colunas tratam dos temas encontrados em cada artigo: os trabalhos marcados como "Legislação" discutem as principais legislações, diretrizes e definições de privacidade. Já os marcados como "Análise" se referem àqueles que estudam maneiras de se analisar sistemas de software com foco na privacidade. Em "Métricas" teremos trabalhos cujos objetivos principais são criar, discutir e/ou utilizar alguma métrica de privacidade. Os trabalhos assinalados na coluna "IDS" tratam de propor, descrever e/ou testar detectores de intrusão de rede. Por fim, os marcados como "Aprendizado de máquina" debatem e/ou propõem soluções em relação à privacidade e seus riscos no contexto de aprendizado de máquina. Já as linhas exibem o passo em que o trabalho foi encontrado e lido: "Preliminar" no passo 1 (Revisão da bibliografia preliminar), "Literatura geral" no passo 2 (Busca e revisão de literatura geral) e "Específica" no passo 6 (Busca e revisão de literatura específica). Por fim as cores simbolizam onde e como os artigos podem ser encontrados nesta monografia: verde (Trabalhos principais) e rosa (Trabalhos revisados), no Capítulo 2, tendo os primeiros uma seção dedicada para cada trabalho e os segundos duas subseções resumindo os principais temas abordados sobre eles, e azul (Apêndice) cuja discussão dos trabalhos se encontra no Apêndice A.

	Trabalho	Legislação	Análise	Métricas	IDS	Aprendizado de máquina
Preliminar	Agarwal, 2016	X		X		
	Bal *et al.*, 2014			X		
	Monteiro, 2020	X				
	Wagner e Eckhoff, 2018			X		
Literatura geral	Tikkinen-Piri *et al.*, 2018	X				
	Kurose e Rossa, 2009				X	
	Kioskli *et al.*, 2022		X			
	Mouratidis e Giorgini, 2007		X			
	Pavlidis e Islam, 2011		X			
	Stallings e Brown, 2021	X				
	Yoon *et al.*, 2020					X
	Mosaiyebzadeh *et al.*, 2023					X
	Shi *et al.*, 2021				X	X
	Kim e Chung, 2019					X
	Cassales *et al.*, 2019				X	
	Elias *et al.*, 2022				X	
	Arbex *et al.*, 2021				X	
	Sherry *et al.*, 2015				X	
	Canard *et al.*, 2017				X	
	Cassales *et al.*, 2019				X	
	Zhang e Zhu, 2018				X	
	Preibusch, 2013			X		
	Martin e Nissenbaum, 2017	X		X		
	Braunstein *et al.*, 2011			X		
	Knijnenburg *et al.*, 2017			X		
Específica	Dwork, 2006			X		
	E. D. Canedo *et al.*, 2023	X	X			
	Castro *et al.*, 2022	X	X			
	Rocha e E. Canedo, 2023	X	X			
	Mendes e Vilela, 2017	X		X		X

Trabalhos principais Trabalhos revisados Apêndice

Tabela B.1: *Síntese de Trabalhos Estudados*

Referências

[Adam *et al.* 2011] Nabil Adam, Haibing Lu, Jaideep Vaidya e Basit Shafiq. "Statistical databases". In: *Encyclopedia of Cryptography and Security*. Ed. por Henk C. A. van Tilborg e Sushil Jajodia. Boston, MA: Springer US, 2011, pp. 1256–1260. isbn: 978-1-4419-5906-5. doi: 10.1007/978-1-4419-5906-5_767. url: https://doi.org/10.1007/978-1-4419-5906-5_767 (citado na pg. 34).

[Agarwal 2016] Sushant Agarwal. "Developing a Structured Metric to Measure Privacy Risk in Privacy Impact Assessments". In: *Privacy and Identity Management. Time for a Revolution? : 10th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2.2 International Summer School, Edinburgh, UK, August 16-21, 2015, Revised Selected Papers*. Ed. por David Aspinall, Jan Camenisch, Marit Hansen, Simone Fischer-Hübner e Charles Raab. Vol. AICT-476. IFIP Advances in Information and Communication Technology. Springer International Publishing, 2016, pp. 141–155. doi: 10.1007/978-3-319-41763-9_10. url: https://inria.hal.science/hal-01619743 (citado nas pgs. 9, 27, 40).

[Arbex *et al.* 2021] Gustavo Vitral Arbex, Kétly Gonçalves Machado, Michele Nogueira, Daniel M. Batista e Roberto Hirata. "IoT DDoS Detection Based on Stream Learning". In: *12th NoF*. 2021. doi: 10.1109/NoF52522.2021.9609940 (citado nas pgs. 7, 20, 24, 40).

[Atkinson *et al.* 2018] John S. Atkinson, John E. Mitchell, Miguel Rio e George Matich. "Your wifi is leaking: what do your mobile apps gossip about you?" *Future Generation Computer Systems* 80 (2018), pp. 546–557. issn: 0167-739X. doi: https://doi.org/10.1016/j.future.2016.05.030. url: https://www.sciencedirect.com/science/article/pii/S0167739X16301480 (citado na pg. 35).

[Bal *et al.* 2014] Gökhan Bal, Kai Rannenberg e Jason Hong. "Styx: Design and Evaluation of a New Privacy Risk Communication Method for Smartphones". In: *29th IFIP International Information Security Conference (SEC)*. Ed. por Nora Cuppens-Boulahia, Frédéric Cuppens, Sushil Jajodia, Anas Abou El Kalam e Thierry Sans. Vol. AICT-428. ICT Systems Security and Privacy Protection. Part 3: Mobile Security. Marrakech, Morocco: Springer, jun. de 2014, pp. 113–126. doi: 10.1007/978-3-642-55415-5_10. url: https://inria.hal.science/hal-01370359 (citado nas pgs. 11, 20, 40).

[Braunstein *et al.* 2011] Alex Braunstein, Laura Granka e Jessica Staddon. "Indirect content privacy surveys: measuring privacy without asking about it". In: SOUPS '11. Pittsburgh, Pennsylvania: Association for Computing Machinery, 2011. isbn: 9781450309110. doi: 10.1145/2078827.2078847. url: https://doi.org/10.1145/2078827.2078847 (citado nas pgs. 37, 40).

[Canard *et al.* 2017] Sébastien Canard, Aïda Diop, Nizar Kheir, Marie Paindavoine e Mohamed Sabt. "Blindids: market-compliant and privacy-friendly intrusion detection system over encrypted traffic". In: *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*. ASIA CCS '17. Abu Dhabi, United Arab Emirates: Association for Computing Machinery, 2017, pp. 561–574. isbn: 9781450349444. doi: 10.1145/3052973.3053013. url: https://doi.org/10.1145/3052973.3053013 (citado nas pgs. 20, 40).

[E. D. Canedo *et al.* 2023] Edna Dias Canedo *et al.* "Privacy requirements elicitation: a systematic literature review and perception analysis of it practitioners". *Requirements Engineering* 28.2 (jun. de 2023), pp. 177–194. issn: 1432-010X. doi: 10.1007/s00766-022-00382-8. url: https://doi.org/10.1007/s00766-022-00382-8 (citado nas pgs. 17, 40).

[Cassales *et al.* 2019] Guilherme Weigert Cassales, Hermes Senger, Elaine Ribeiro de Faria e Albert Bifet. "Idsa-iot: an intrusion detection system architecture for iot networks". In: *2019 IEEE Symposium on Computers and Communications (ISCC)*. 2019, pp. 1–7. doi: 10.1109/ISCC47284.2019.8969609 (citado nas pgs. 20, 40).

[Castro *et al.* 2022] Evandro Thalles Vale de Castro, Geovana R. S. Silva e Edna Dias Canedo. "Ensuring privacy in the application of the brazilian general data protection law (lgpd)". In: *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. SAC '22. Virtual Event: Association for Computing Machinery, 2022, pp. 1228–1235. isbn: 9781450387132. doi: 10.1145/3477314.3507023. url: https://doi.org/10.1145/3477314.3507023 (citado nas pgs. 14, 40).

[*SECOMU: Falta de privacidade e controle comportamental numa economia de vigilância* 2021] *SECOMU: Falta de privacidade e controle comportamental numa economia de vigilância*. https://youtu.be/MOLY8cTcxRs?t=3488. Acessado em 11 de Novembro de 2022. CSBC, 2021 (citado na pg. 12).

[Densmore 2013] R. Densmore. *Privacy Program Management: Tools for Managing Privacy Within Your Organization*. International Association of Privacy Professionals, 2013 (citado na pg. 6).

[Dwork 2006] Cynthia Dwork. "Differential privacy". In: *Automata, Languages and Programming*. Ed. por Michele Bugliesi, Bart Preneel, Vladimiro Sassone e Ingo Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–12. isbn: 978-3-540-35908-1 (citado nas pgs. 34, 40).

[Elias *et al.* 2022] Erik Miguel de Elias *et al.* "A Hybrid CNN-LSTM Model for IIoT Edge Privacy-Aware Intrusion Detection". In: *14th IEEE LATINCOM*. 2022. doi: 10.1109/LATINCOM56090.2022.10000468 (citado nas pgs. 3, 8, 20, 25, 29, 30, 40).

[Ferrão 2022] Sâmmara Éllen Renner Ferrão. "Proposta de uma taxonomia de requisitos de privacidade baseada na lgpd e iso/iec 29100: aplicação prática no open banking brasil" (2022) (citado na pg. 33).

[Garsva *et al.* 2015] Eimantas Garsva, Nerijus Paulauskas e Gediminas Grazulevicius. "Packet size distribution tendencies in computer network flows". In: *2015 Open Conference of Electrical, Electronic and Information Sciences (eStream)*. 2015, pp. 1–6. doi: 10.1109/eStream.2015.7119483 (citado na pg. 30).

[Giorgini *et al.* 2004] Paolo Giorgini, Manuel Kolp, John Mylopoulos e Marco Pistore. "The tropos methodology". In: *Methodologies and Software Engineering for Agent Systems: The Agent-Oriented Software Engineering Handbook*. Ed. por Federico Bergenti, Marie-Pierre Gleizes e Franco Zambonelli. Boston, MA: Springer US, 2004, pp. 89–106. isbn: 978-1-4020-8058-6. doi: 10.1007/1-4020-8058-1_7. url: https://doi.org/10.1007/1-4020-8058-1_7 (citado na pg. 17).

[Giunchiglia *et al.* 2003] Fausto Giunchiglia, John Mylopoulos e Anna Perini. "The tropos software development methodology: processes, models and diagrams". In: *Agent-Oriented Software Engineering III*. Ed. por Fausto Giunchiglia, James Odell e Gerhard Weiss. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 162–173. isbn: 978-3-540-36540-2 (citado na pg. 17).

[*GS1* 2012] *GS1. GS1 EPC/RFID Privacy Impact Assessment Tool*. 2012 (citado na pg. 9).

[Hustinx 2010] Peter Hustinx. "Privacy by design: delivering the promises". *Identity in the Information Society* 3.2 (ago. de 2010), pp. 253–255. issn: 1876-0678. doi: 10.1007/s12394-010-0061-z. url: https://doi.org/10.1007/s12394-010-0061-z (citado nas pgs. 6, 12).

[Iwaya *et al.* 2023] Leonardo Horn Iwaya, Muhammad Ali Babar e Awais Rashid. "Privacy Engineering in the Wild: Understanding the Practitioners' Mindset, Organisational Aspects, and Current Practices". *IEEE Transactions on Software Engineering* (2023), pp. 1–26. doi: 10.1109/TSE.2023.3290237 (citado na pg. 1).

[Kim e Chung 2019] Soohyung Kim e Yon Dohn Chung. "An anonymization protocol for continuous and dynamic privacy-preserving data collection". *Future Generation Computer Systems* 93 (2019), pp. 1065–1073. issn: 0167-739X. doi: https://doi.org/10.1016/j.future.2017.09.009. url: https://www.sciencedirect.com/science/article/pii/S0167739X17319520 (citado nas pgs. 19, 40).

[Kioskli *et al.* 2022] Kitty Kioskli, Daniele Dellagiacoma, Theofanis Fotis e Haralambos Mouratidis. "The Supply Chain of a Living Lab: Modelling Security, Privacy, and Vulnerability Issues alongside with their Impact and Potential Mitigation Strategies". *JoWUA* 13.2 (jun. de 2022), pp. 147–182. URL: https://repository.essex.ac.uk/33323/ (citado nas pgs. 3, 17, 22, 40).

[Knijnenburg *et al.* 2017] Bart Knijnenburg *et al.* "Death to the privacy calculus?" In: fev. de 2017. DOI: http://dx.doi.org/10.2139/ssrn.2923806. URL: https://ssrn.com/abstract=2923806 (citado nas pgs. 37, 40).

[Kurose e Rossa 2009] James F. Kurose e Keith W. Rossa. *Redes de computadores e a internet*. 5ª ed. Pearson Universidades, 2009, pp. 540–544 (citado nas pgs. 6, 7, 40).

[Martin e Nissenbaum 2017] Kirsten Martin e Helen Nissenbaum. "Measuring privacy: an empirical test using context to expose confounding variables". *Columbia Science & Technology Law Review* 18 (jan. de 2017), pp. 176–218 (citado nas pgs. 37, 40).

[Mendes e Vilela 2017] Ricardo Mendes e João P. Vilela. "Privacy-preserving data mining: methods, metrics, and applications". *IEEE Access* 5 (2017), pp. 10562–10582. DOI: 10.1109/ACCESS.2017.2706947 (citado nas pgs. 1, 40).

[Mosaiyebzadeh *et al.* 2023] Fatemeh Mosaiyebzadeh *et al.* "Privacy-Enhancing Technologies in Federated Learning for the Internet of Healthcare Things: A Survey". *Electronics* 12.12 (2023). ISSN: 2079-9292. DOI: 10.3390/electronics12122703. URL: https://www.mdpi.com/2079-9292/12/12/2703 (citado nas pgs. 1, 19, 40).

[Mouratidis e Giorgini 2007] Haris Mouratidis e Paolo Giorgini. "Secure Tropos: A Security-Oriented Extension of the Tropos Methodology". *IJSEKE* 17 (abr. de 2007). DOI: 10.1142/S0218194007003240 (citado nas pgs. 17, 40).

[Nissenbaum 2004] Helen Nissenbaum. "Privacy as contextual integrity". *Wash. L. Rev.* 79 (2004), p. 119 (citado na pg. 6).

[Oetzel e Spiekermann 2014] Marie Caroline Oetzel e Sarah Spiekermann. "A systematic methodology for privacy impact assessments: a design science approach". *European Journal of Information Systems* 23.2 (2014), pp. 126–150. DOI: 10.1057/ejis.2013.18. URL: https://doi.org/10.1057/ejis.2013.18 (citado na pg. 10).

[Pavlidis e Islam 2011] Michalis Pavlidis e Shareeful Islam. "Sectro: a case tool for modelling security in requirements engineering using secure tropos." In: vol. 734. Jan. de 2011, pp. 89–96 (citado nas pgs. 17, 40).

[Preibusch 2013] Sören Preibusch. "Guide to measuring privacy concern: review of survey and observational instruments". *International Journal of Human-Computer Studies* 71.12 (2013), pp. 1133–1143. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2013.09.002. URL: https://www.sciencedirect.com/science/article/pii/S1071581913001183 (citado nas pgs. 37, 40).

[Monteiro 2020] Alessandra Monteiro. *Privacidade e Monitoramento. Uma perspectiva LGPD e GDPR*. https://www.youtube.com/watch?v=-yC-8lP4aFk. Acessado em 11 de Novembro de 2022. Roadsec, 2020 (citado nas pgs. 12, 40).

[Rocha e E. Canedo 2023] Lucas Rocha e Edna Canedo. "Privacy compliance in software development: a guide to implementing the lgpd principles extended abstract – ctdsi/ctccsi 2023". In: *Anais Estendidos do XIX Simpósio Brasileiro de Sistemas de Informação*. Maceió/AL: SBC, 2023, pp. 68–70. doi: 10.5753/sbsi_estendido.2023.229324. url: https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/24587 (citado nas pgs. 14, 33, 40).

[*Secure Tropos Project* s.d.] *Secure Tropos Project*. University of Brighton. url: https://www.brighton.ac.uk/csius/what-we-do/research-projects/secure-tropos.aspx (citado na pg. 17).

[Sherry *et al.* 2015] Justine Sherry, Chang Lan, Raluca Ada Popa e Sylvia Ratnasamy. "Blindbox: deep packet inspection over encrypted traffic". *SIGCOMM Comput. Commun. Rev.* 45.4 (ago. de 2015), pp. 213–226. issn: 0146-4833. doi: 10.1145/2829988.2787502. url: https://doi.org/10.1145/2829988.2787502 (citado nas pgs. 20, 40).

[Shi *et al.* 2021] Jibo Shi, Bin Ge, Yang Liu, Yu Yan e Shuang Li. "Data privacy security guaranteed network intrusion detection system based on federated learning". In: *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 2021, pp. 1–6. doi: 10.1109/INFOCOMWKSHPS51825.2021.9484545 (citado nas pgs. 3, 19, 20, 30, 31, 40).

[Silva e França 2023] Lucas Silva e Rozelma França. "Educação para a Cidadania Digital: Um mapeamento sobre as práticas de ensino para promover a segurança e a privacidade de dados". In: *XXXI WEI*. SBC, 2023, pp. 533–544. doi: 10.5753/wei.2023.230839 (citado na pg. 1).

[*SPIA* s.d.] *SPIA. Introduction to the SPIA Program*. University of Pennsylvania. url: http://www.upenn.edu/computing/security/spia/spia_step_by_step.pdf (citado na pg. 10).

[Stallings e Brown 2021] William Stallings e Lawrie Brown. *Computer Security. Principles and Practice*. 4ª ed. Pearson, 2021 (citado nas pgs. 6, 40).

[Tikkinen-Piri *et al.* 2018] Christina Tikkinen-Piri, Anna Rohunen e Jouni Markkula. "Eu general data protection regulation: changes and implications for personal data collecting companies". *Computer Law & Security Review* 34.1 (2018), pp. 134–153. issn: 0267-3649. doi: https://doi.org/10.1016/j.clsr.2017.05.015. url: https://www.sciencedirect.com/science/article/pii/S0267364917301966 (citado nas pgs. 1, 5, 9, 40).

[*Tropos Project* s.d.] *Tropos Project*. Dipartimento di Ingegneria e Scienzs dell'Informazione - Università degli Studi di Trento. URL: http://www.troposproject.eu/ (citado na pg. 17).

[Wagner e Eckhoff 2018] Isabel Wagner e David Eckhoff. "Technical Privacy Metrics". *ACM Computing Surveys* 51.3 (jun. de 2018), pp. 1–38. DOI: 10.1145/3168389. URL: https://doi.org/10.1145%5C%2F3168389 (citado nas pgs. 6, 14, 27, 40).

[Wang *et al.* 2015] Qinglong Wang, Amir Yahyavi, Bettina Kemme e Wenbo He. "I know what you did on your smartphone: inferring app usage over encrypted data traffic". In: *2015 IEEE Conference on Communications and Network Security (CNS)*. 2015, pp. 433–441. DOI: 10.1109/CNS.2015.7346855 (citado na pg. 35).

[Warren e Brandeis 1890] Samuel D. Warren e Louis D. Brandeis. "The right to privacy". *Harvard Law Review* 4.5 (1890), pp. 193–220. ISSN: 0017811X. URL: http://www.jstor.org/stable/1321160 (acesso em 09/10/2023) (citado na pg. 5).

[Westin 1967] Alan Westin. *Privacy and Freedom*. 1967 (citado na pg. 5).

[Yin *et al.* 2021] Xuefei Yin, Yanming Zhu e Jiankun Hu. "A comprehensive survey of privacy-preserving federated learning: a taxonomy, review, and future directions". *ACM Computing Surveys* 54 (jul. de 2021), pp. 1–36. DOI: 10.1145/3460427 (citado na pg. 30).

[Yoon *et al.* 2020] Jinsung Yoon, Lydia N. Drumright e Mihaela van der Schaar. "Anonymization through data synthesis using generative adversarial networks (ads-gan)". *IEEE Journal of Biomedical and Health Informatics* 24.8 (2020), pp. 2378–2388. DOI: 10.1109/JBHI.2020.2980262 (citado nas pgs. 19, 40).

[Zhang e Zhu 2018] Tao Zhang e Quanyan Zhu. "Distributed privacy-preserving collaborative intrusion detection systems for vanets". *IEEE Transactions on Signal and Information Processing over Networks* 4.1 (2018), pp. 148–161. DOI: 10.1109/TSIPN.2018.2801622 (citado nas pgs. 20, 40).