ANONIMIZAÇÃO DE DADOS PESSOAIS NA GESTÃO DE SEGURANÇA DA INFORMAÇÃO: UMA ANÁLISE A PARTIR DO SECURITY INFORMATION AND EVENT MANAGEMENT (SIEM)

Elivelton de Oliveira Santos[1]; Rosilene Paiva Marinho de Sousa[2]; Luiz Hilário Ferreira Damascena[3]

Resumo

Esta pesquisa tem por objetivo relatar a importância da privacidade de dados no âmbito da segurança da informação e a necessidade de ajustes necessários no *Security Information and Event Management* (SIEM), aplicando meios técnicos razoáveis a fim de anonimizar dados classificados como identificadores indiretos e diretos que são armazenados nestes sistemas utilizando a ferramenta *Graylog* para estar em consonância com a Lei Geral de Proteção de Dados Pessoais (Lei n.º 13.709/2018). O estudo se baseia em pesquisa de abordagem qualitativa, de natureza exploratória e documental que pressupõe a constituição dos resultados com base em leis, regulamentos, portarias, guias invariáveis passíveis de verificação e previsão para o entendimento jurídico e técnico da anonimização à luz da legislação, além disso, utilizando função *pipeline* somada a uma linguagem específica de domínio (DSL), desenvolvida pela *Graylog Inc*., utilizada nas regras para definir a lógica de processamento de eventos, aplicando técnica de mascaramento nos dados antes de serem armazenados. Assim, esta investigação visa resultar na utilização destes dados sem o risco de descoberta dos dados pessoais originais, demonstrando o funcionamento dos códigos desenvolvidos, garantindo que os usuários não sejam personificados/caracterizados, protegendo consequentemente a privacidade e a vida pessoal de uma pessoa física.

Palavras-chave: siem; proteção de dados pessoais; anonimização.

Abstract

This research aims to report the importance of data privacy in the context of information security and the need for necessary adjustments in Security Information and Event Management (SIEM), applying reasonable technical means in order to anonymize data classified as indirect and direct identifiers that are stored in these systems using the Graylog tool to be in line with the General Law for the Protection of Personal Data (Law No. 13,709/2018). The study is based on qualitative research, of an exploratory and documentary nature that presupposes the constitution of the results based on laws, regulations, ordinances, invariable guides that can be verified and predicted for the legal and technical understanding of anonymization in light of the legislation, in addition, using a pipeline function added to a domain-specific language (DSL), developed by Graylog Inc., used in the rules to define the event processing logic, applying a masking technique to the data before being stored. Thus, this research aims to result in the use of this data without the risk of discovery of the original personal data, demonstrating the

[1] Especialista em Segurança da Informação pelo Programa de Pós-graduação Estácio de Sá; Graduado em Gestão da Tecnologia da Informação pela Faculdade São Francisco de Barreiras. E-mail:elivelton.santos@ufob.edu.br

[2] Doutora em Ciência da Informação pela Universidade Federal da Paraíba (UFPB), professora do Programa de Pós-Graduação em Tecnologia para Inovação da Universidade Federal do Oeste da Bahia-PROFNIT/UFOB e do Programa de Pós-Graduação em Gestão nas Organizações Aprendentes da Universidade Federal da Paraíba-PPGOA/UFPB. E-mail:rosilene.sousa@ufob.edu.br.

[3] Mestre em Propriedade Intelectual e Transferência de Tecnologia para Inovação pela Universidade Federal da Bahia-PROFNIT/UFOB, é analista de Tecnologia da Informação e Coordenador da Coordenadoria de Infraestrutura e Segurança da Universidade Federal do Oeste da Bahia-UFOB. E-mail:luiz.damascena@ufob.edu.br.

functioning of the developed codes, ensuring that users are not personified/characterized, consequently protecting the privacy and personal life of an individual.
Keywords: siem; personal data protection; anonymization.

Introdução

A Lei Geral de Proteção de Dados (Lei n.º 13.709/2018) regula o tratamento de dados pessoais no Brasil, incluindo informações que possam identificar direta ou indiretamente uma pessoa natural. Existem informações que são capazes de identificar uma pessoa natural como nome, CPF, endereço, *Media Access Control* - MAC, endereço IP ou hábitos de consumo.

De acordo com Souza (2024, p. 8) no que diz respeito à privacidade, na era digital, é um tema cada vez mais relevante e complexo, especialmente à medida que a tecnologia avança e se torna mais integrada na nossa vida quotidiana. Nessa circunstância, a transformação digital traz novas dimensões às questões de privacidade, e integridade, desafiando as percepções tradicionais e exigindo novas abordagens e quadros jurídicos e éticos.

Dentro desse contexto, o conceito de dado pessoal é o tema central quando se trata de privacidade, visto que, essas informações quando tratadas de forma inadequada, podem ser utilizadas para delinear perfis, direcionando para práticas publicitárias ou para práticas discriminatórias. Diante disso, dado pessoal é toda informação que identifica ou possibilita a identificação de uma pessoa natural, seja de forma direta ou indireta, conforme definido pela Lei n.º 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018).

O dado pessoal, à luz da LGPD, pode ser definida de duas formas, por critérios reducionistas que está vinculado a identificação direta de uma pessoa física como o CPF, RG, endereço, localização de GPS, dados bancários, e-mail, prontuário médico, telefone e mediante critérios expansionistas que se refere a dados que de forma indireta lpodem caracterizar o titular dos dados, como histórico de navegação, nacionalidade e o cookie (BRASIL, 2018). Além disso, a referida lei trata sobre os dados considerados sensíveis aos detentores, que são informações relacionados a questões étnicas, religião, política, vida sexual, entre outras. Por esse motivo, a LGPD estabelece regras para o tratamento adequado desses dados, visando garantir os direitos sobre essas informações, como à privacidade e a autodeterminação informativa.

Nesse cenário que trata o ato de resguardar as informações pessoais, tem-se o processo de anonimização de informações. Para Prata *et al*. (2020, p. 114) o processo de anonimização é um aspecto importante ao nível de garantir a privacidade de cada indivíduo, sem perder a utilidade desses dados, ou seja, é aplicar técnicas que retira determinadas características dos dados, de maneira a obstruir a utilização dessas informações pelos utilizadores. Nesse aspecto,

anonimização é um processo seguro do permanecimento intacto da vida pessoal de cada indivíduo no meio digital, é aplicação da ética por meio de tecnologias que empregam a seguridade dos dados de uma pessoa natural.

Security Information and Event Management (SIEM) usam meios para coletar eventos de diversas fontes de dados, com isso, ele processa esses *logs* de forma proativa para apresentar uma visão de segurança consubstanciada de toda rede interna, tendo uma característica indispensável para uma arquitetura de segurança (Tariq *et al*., 2023, p. 2). Dentro desse contexto, sistemas SIEM são ferramentas com grande capacidade de fornecer uma mitigação de risco apurada, visto que, além de correlacionar informações internas, pode fazer esse tipo de correlação com bases externas, age de maneira proativa aplicando inteligência aos eventos nele armazenado não apenas de característica técnica relacionadas à origem, como também em eventos considerados identificadores diretos e indiretos.

Scholzel *et al*. (2015, apud Neu *et al*., 2019, p. 2) afirmam que ferramentas de gerenciamento de eventos de segurança da informação podem ser utilizadas para facilitar o processo de correlação e retenção de logs. Essas ferramentas têm como objetivo compilar e apresentar informações contidas em eventos de auditoria. Ademais, a implantação de um sistema SIEM possibilita o acompanhamento e a resposta ágil a eventos maliciosos detectados.

Nesse sentido, uma empresa ou instituição ao utilizar um sistema de gerenciamento e correlação de eventos de segurança, conhecido pela nomenclatura em inglês como *Security Information and Event Management* (SIEM), permite a análise refinada de *log*, capacidade de investigação forense, examinar determinados eventos para julgamentos assertivos, aferição de possíveis danos futuros ou danos atuais ocasionados por um ataque. Entretanto, algumas informações nele armazenadas podem ser capazes de caracterizar uma pessoa natural, identificando-a e expondo essas informações a atacantes maliciosos de forma clara.

Nessa conjuntura, questiona-se: Como o SIEM contribui para mitigação de risco e garantia da segurança da informação na anonimização de dados pessoais?

Com isso, o objetivo dessa pesquisa é mitigar esses problemas e adequar o sistema de gerenciamento e correlação de eventos de segurança para estar conforme a Lei Geral de Proteção de Dados (Lei n.º 13.709/2018), utilizando a ferramenta SIEM *Graylog*, aplicando técnica de mascaramento com o intuito de anonimizar determinados dados sensíveis aos usuários, promovendo privacidade e minimizando a identificação de uma pessoa natural identificável.

Dentro desse contexto, a segurança dos dados em redes e sistemas interconectados, sempre será de grande preocupação tanto para pequenas empresas como para médias e grandes

corporações. Com isso, para que o setor da segurança da informação e cibersegurança se desenvolva e possa propagar os seus efeitos é preciso oferecer melhorias, boas práticas, implementação de ferramentas que possam auxiliar na mitigação de riscos aplicando técnica de mascaramento que caracteriza um dado anonimizado. Sistemas de correlação de eventos de segurança, apesar de ser uma ferramenta que emprega uma versatilidade no âmbito da segurança da informação e que consiste em um conjunto de processos para boas práticas de gerenciamento de eventos, ainda é possível o armazenamento de informações que possam personificar um indivíduo.

Diante disso, a adoção de anonimização em sistemas SIEM é justificada pela necessidade de garantir a conformidade legal para proteger a privacidade dos titulares, permitindo o uso seguro e eficiente dos dados para fins de segurança e posterior análise, combatendo riscos de exposição e vazamento de informações sensíveis. Concernente a isso, trata-se de uma prática estratégica e mandatória para organizações que buscam aliar segurança da informação com a confidencialidade, integridade, disponibilidade e o mais importante a privacidade dos dados pessoais.

1 Referencial teórico

1.1 A Lei Geral de Proteção de Dados Pessoais

Os dados pessoais são informações que permitem identificar uma pessoa, seja direta ou indiretamente. No Brasil, a proteção desses dados é regulamentada principalmente pela Lei Geral de Proteção de Dados Pessoais (Lei n.º 13.709/2018) que é uma ferramenta legislativa que estabelece critérios sobre tratamento de dados pessoais, nos meios digitais, por pessoa física ou por pessoa jurídica de direito público, ou privado, visando proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da caracterização da pessoa natural (BRASIL, 2018). Conforme os propósitos gerais dessa lei, essas regras visam garantir que, sempre que associado à produção e à disseminação do conhecimento, o tratamento de dados pessoais seja realizado com segurança jurídica e com respeito aos direitos dos titulares.

Para efeitos dessa lei, o Art. 5º, especifica o que é Dado Pessoal:

> I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;
> II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural [...] (BRASIL, 2018).

No que concerne a dados pessoais, as instituições públicas e privadas concentram uma quantidade expressiva de dados pessoais de toda sua comunidade, aumentando com isso, os desafios na proteção desses dados. No Art. 1º da Lei 13.709/2018 evidencia sobre o tratamento

de dados pessoais, enfatizando os direitos fundamentais de liberdade, integridade e privacidade, além do livre desenvolvimento da pessoa natural nos meios digitais. Complementarmente, em 2022 foi acrescentado o inciso LXXIX ao artigo 5º da Constituição Federal, via Emenda Constitucional de n.º 115/2022, que diz: "é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais." (BRASIL, 2022). Com isso, a Lei Geral de Proteção de Dados altera o comportamento no meio digital e de como devemos nos comportar diante da proteção dos dados das pessoas.

Mediante o desenvolvimento e expansão da *internet*, a velocidade com que a informação é transmitida para qualquer parte do mundo e compartilhada com diversos sistemas, aumentou exponencialmente. À medida que a utilização da tecnologia e da *internet* em quaisquer ambientes cresceram, a sociedade se torna dependente do uso da tecnologia. Diante disso, a segurança desses dados é essencial, de maneira a garantir que os dados pessoais dos titulares sejam devidamente protegidos contra o uso indevido e, para isso, anonimizar os dados é uma maneira aceitável para se garantir diante das ameaças cibernética e respeito a privacidade (AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS, 2023a, p. 4-5).

Concernente a isso, dado anonimizado são aqueles identificados que através da utilização de técnicas de anonimização possam ser anônimos, passível de não identificação de uma pessoa natural. No Art. 5º, inciso III da Lei 13.709/2018, diz: "dado anonimizado: dado relativo ao titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento". Adicionalmente, são informações pessoais tratadas intencionalmente para perder qualquer possibilidade de associação, direta ou indiretamente um indivíduo específico.

Ainda no Art. 5º da Lei 13.709/2018, especificamente no inciso XI, a referida lei específica que deve ser realizado o uso de métodos técnicos adequados e disponíveis no momento do processamento, que tornam um dado incapaz de ser associado, direta ou indiretamente, a uma pessoa. Diante das metodologias existentes, a Autoridade Nacional de Proteção de Dados - ANPD, dentro do âmbito das técnicas de anonimização explana de forma clara uma proposta de processo genérico para cada uma delas existentes, auxiliando e dando apoio técnico à referida lei e, além disso, garantindo a privacidade e proteção dos dados de forma mais ampla para cada processo de modificação realizada (AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS, 2023a, p. 20-26).

Segundo Faleiros Júnior e Martins (2020, p.377-378) a preocupação no contexto da privacidade e proteção dos dados, vão além do contexto econômico, sociais e políticos, haja vista a proteção da pessoa natural é mais importante diante da personificação de seus atributos,

a qual a eventualidade existente na identificação vislumbra um risco a vida íntima e suas particularidades. Nessa perspectiva, impera a visão de centralizar a defesa naquilo que vai ao encontro com dados sensíveis relacionados ao usuário detentor da informação.

Nesse universo da segurança dos dados, mediante anonimização praticando e aplicando o conceito de privacidade, a *General Data Protection Regulation (GDPR)* regulamento da União Europeia que define as regras sobre o tratamento dos dados pessoais abrangendo à privacidade dos dados, envolve o processo de converter dados pessoais em uma forma que não permite a identificação de uma pessoa natural, seja direta ou indiretamente. Com isso, assim como a LGPD, o GDPR não se aplica a dados anônimos, o que significa que esses dados podem ser usados mais livremente. Além disso, ela não fornece um conceito cristalino em relação a anonimização, mas engessa critérios e requisitos em seu Recital 26 do Regulamento (UE) n.º 2016/679 de 2016 que enfatiza sobre a impossibilidade da identificação por meios razoáveis considerando o tempo, tecnologia e custo (EUROPEAN, 2016).

Em geral, a GDPR regula o tratamento e a livre circulação de dados pessoais de uma pessoa singular identificada ou identificável, que consiste na recolha, armazenamento e utilização de informação que identifique ou permita a identificação das pessoas singulares, como o nome, número de identificação, dados de identificação, dados de localização, identificadores eletrônicos ou outros elementos específicos da identidade física, fisiológica, genética, psicológica, econômica, cultural ou social da pessoa singular (EUROPEAN, 2016).

A anonimização pode ser usada para eximir dados do escopo do GDPR, mas é importante garantir que a identificação seja impossível e irreversível, impedindo a designação direta ou indiretamente. A GDPR não define um padrão específico para anonimização, mas estabelece que dados anonimizados não são considerados dados pessoais se não puderem ser relacionados a uma pessoa, por outro lado, é importante mencionar sobre a sobre pseudoanimização que é reversível, que nesse caso não se aplica a garantia de caracterização e continua sendo sujeita tanto a LGPD quando a GDPR.

1.2 *Security information and event management*

O *Security Information and Event Management* (SIEM) é um instrumento de gerenciamento de eventos, formado pela junção de dois termos: *Security Information Management* (SIM) e *Security Event Management* (SEM). O SIM monitora os *logs* em tempo real, correlação dos eventos e as devidas notificações, enquanto o SEM é responsável pelo armazenamento dos dados a longo prazo, faz as devidas análises e gera os relatórios conforme necessidade (Vazão *et al*., 2019, p. 2).

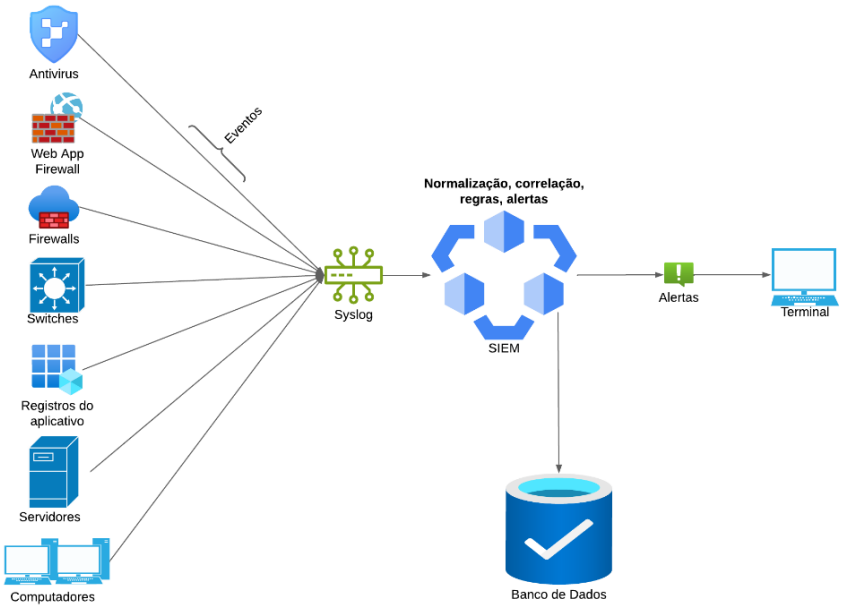
Sistemas de gerenciamento e correlação de eventos de segurança, é uma ferramenta com grande capacidade de classificar, combinar e analisar uma diversidade de eventos gerados por quaisquer ativos de rede. É uma tecnologia que auxilia na guarda dos *logs* gerados pelos ativos e consequentemente dá suporte às equipes de segurança da informação em relação à proteção contra várias ameaças constantemente distribuídas na *internet* e com possibilidades de irradiar na rede interna das organizações.

Nesse contexto, sistemas SIEM avançados nos permitem identificar os principais fluxos dos dados existentes dentro do processo que envolve o gerenciamento de eventos. Dentro desse processo existem as fontes que auxiliam no processamento dos *logs*, a exemplo disso temos a filtragem, agregação, abstração e correlacionamento de eventos, raciocínio e visualização, modelagem de ataques e avaliação de segurança, tomada de decisão e contramedidas que podem ser executadas frente as possíveis ameaças que podem surgir (Kotenko *et al*., 2013, p.358).

Como citado anteriormente, sistemas SIEM tem a capacidade de identificação proativa de um ataque. Para que isso ocorra efetivamente no processo, é necessária a configuração de alerta para cada caso e que ocorra em tempo real, ou seja, no momento do exato do evento malicioso. Com isso, em vias gerais, a arquitetura SIEM é composta por vários itens que fazem parte da sua estrutura. Para Pavlik *et al.*, (2014, p.210-211) a arquitetura SIEM é composta por dispositivos de origem que são fontes de dados que geralmente enviam eventos via protocolo *syslog*, sendo as principais fontes: sistemas operacionais, servidores, aplicativos, dispositivos de redes, *firewalls*, antivírus e sistemas IDS/IPS, coleta de *logs* via método *push* (eventos enviados dos dispositivos de origem) ou método *pull* (onde o SIEM se conecta ao dispositivo), normalização; processo de transformação do evento puro em um formato uniforme e padronizado, mecanismo de regras/correlação onde as correlações ocorrem mediante consultas tanto com informações internas quanto informações consultadas em bases externas, conjunto de regras considerado o cérebro dos sistemas SIEM, capaz de detectar em tempo real potenciais riscos existentes, armazenamento de dados; o ato de guardar os eventos em um banco de dados normalmente *NoSQL* (Banco de Dados não relacional), monitoramento; tanto por envio de alerta quanto na criação de painéis mostrando informações que auxiliam na tomada de decisão.

A figura 1 representa como se comporta uma arquitetura SIEM. Sistemas de segurança, equipamentos de rede, servidores e *endpoints* enviam eventos através do protocolo *syslog* para o SIEM e nele, é realizado a prática de normalização, aplicação de regras, correlação de informações, envio de alerta para os terminais dos profissionais e a todo momento armazenando essas informações em bases de dados não relacionais.

Figura 1 - Arquitetura com SIEM



Fonte: Elaborado pelos autores utilizando a ferramenta Lucidchart (2025).

O SIEM é uma ferramenta essencial para a detecção imediata de ameaças à segurança dos dados, ele monitora continuamente as atividades em ambiente de rede crítica, utilizado como mecanismo de automatização para gerar notificações sobre possíveis violações e atividades suspeitas. Nesse contexto, além de indicar concordância com a Lei 13.709/2018, no que tange a boas práticas, estar também consoante com o Título 45, Subtítulo A, Subcapítulo C, Parte 164, Subparte C, Sessão 164.308 Salvaguardas administrativas do Código de Regulamentações Federais (CRF) dos Estados Unidos que trata sobre as normas de segurança para a proteção de informações eletrônicas de saúde protegidas:

> [4]Uma entidade coberta ou associado de negócios deve, de acordo com § 164.306
> (i) Padrão: Processo de gestão de segurança. Implementar políticas e procedimentos para prevenir, detectar, conter e corrigir violações de segurança.
> (ii) Especificações de implementação: [...]
> (B) Gestão de Riscos (Obrigatório). Implementar medidas de segurança suficientes para reduzir os riscos e vulnerabilidades a um nível razoável e adequado para cumprir §164.306(a) [...]
> (D) Revisão da atividade do sistema de informação (obrigatório). Implementar procedimentos para revisar regularmente registros da atividade do sistema de informações, como registros de auditoria, relatórios de acesso e relatórios de rastreamento de incidentes de segurança. (ESTADOS UNIDOS, 2013)

[4]*(a) A covered entity or business associate must, in accordance with §164.306:*
(i) Standard: Security management process. Implement policies and procedures to prevent, detect, contain, and correct security violations.
(ii) Implementation specifications: [...]
(B) Risk management (Required). Implement security measures sufficient to reduce risks and vulnerabilities to a reasonable and appropriate level to comply with §164.306(a) [...]
(D) Information system activity review (Required). Implement procedures to regularly review records of information system activity, such as audit logs, access reports, and security incident tracking reports.

A partir desse contexto, a ABNT NBR ISO/IEC 27001 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2022a, p. 5) e ABNT NBR ISO/IEC 27002 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2022b, p. 19) são complementares na indicação e importância do uso de sistemas de monitoramento de eventos, expressando a necessária implementação de um sistema de gestão da segurança da informação, baseado na abordagem de riscos do negócio, para estabelecer, implementar, operar, monitorar, produzir inteligência sobre ameaças, manter e melhorar a segurança da informação.

A implementação de medidas de segurança é essencial no meio digital, a concomitância entre as normas regulamentadoras ISO 27001:2022 e ISO 27002:2022, Código de Regulamentações Federais e a Lei Geral de Proteção de Dados Pessoais compreende um controle adequado com técnicas aplicadas sob diversos vieses tecnológicos, demandando uma gestão de risco adotando medidas positivas junto ao tratamento massivo de dados à mercê das ameaças cibernéticas.

Sobre o tema, no Art. 46 da Lei 13.709/2018, diz que os responsáveis pelo tratamento dos dados devem adotar medidas de segurança, técnicas e administrativas para proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas, em geral (BRASIL, 2018). Nesse sentido, um SIEM é uma ferramenta poderosa no que diz respeito a segurança de dados e mitigação de incidentes, além de dar suporte técnico para adequação à LGPD de maneira condizente com o que se pede.

Diante dessa conjuntura de segurança de dados, com a crescente ameaça de ataques cibernéticos, sistemas SIEM são cruciais para proteger os ativos digitais das organizações, esses sistemas automatizam a detecção de padrões suspeitos e geram alerta para a equipe de segurança, permitindo uma resposta rápida a incidentes. Com isso, as implementações de um SIEM melhora a eficiência na gestão de segurança, reduzindo o impacto de violações e melhorando a conformidade com regulamentações jurídicas, aplicando processo ponderando a privacidade e a intimidade das pessoas.

2 Método

O estudo é baseado em pesquisa de abordagem qualitativa, de caráter exploratório e documental que pressupõe sobre a constituição dos resultados regidos por leis invariáveis, normativos, guias que podem ser verificados e previstos para o entendimento legal e técnico da anonimização à luz da legislação, auxiliando na investigação utilizando os dados com relação direta na construção da hipótese referente ao tema em questão, dando sustentação à justificativa de resolução do problema.

Para Sampieri, Collado e Lucio. (2013, p.33) a pesquisa qualitativa "[...] utiliza a coleta de dados sem medição numérica para descobrir ou aprimorar perguntas de pesquisa no processo de interpretação". Nesse mesmo contexto, a pesquisa qualitativa procura principal interpretar o fenômeno que observa, não empregando procedimentos estatísticos, ou não tendo como objetivo principal a abordagem do problema a partir desses procedimentos (Gil, 2019, p.61).

Referente a pesquisa documental, as fontes são documentos no sentido mais amplo, ou seja, não apenas documentos impressos, mas sobretudo todos os outros tipos de documentos, como jornais, fotografias, filmes, gravações sonoras e documentos legais. Nestes casos, o conteúdo dos textos não foi submetido a nenhuma análise, eles permanecem como matéria-prima a partir da qual o pesquisador investigará e analisará. (Severino, 2017, p.131).

Para Severino (2017, p.132), "[...]no que diz respeito a pesquisa exploratória, busca apenas levantar informações sobre um determinado objeto, delimitando assim um campo de trabalho, mapeando as condições de manifestação desse objeto". Nesse mesmo sentido, Marconi e Lakatos (2021, *apud* Severino, 2017, p.135) mencionam que levantado os dados, eles precisam ser articulados de forma lógica com o real, tendo como base uma teoria que lhe dê sustentação, visto que, a ciência é uma modalidade de conhecimento que não se constitui como um simples levantamento de dados.

2.1 Etapas metodológicas

Com a finalidade de solucionar o problema da falta de anonimização e aplicar a técnica de mascaramento na informação é seguido as etapas abaixo:

- Etapa 1: Arquitetado um ambiente para simulação e validação dos códigos desenvolvidos com 3 (três) máquinas virtuais, 2 (duas) utilizando sistema operacional *Ubuntu Server* 24.04.2 LTS e 1 (uma) *appliance* virtual *FortiGate-VM*. Máquina 1; *script bash* para simulação de um ambiente sistêmico utilizando o CPF como usuário e endereço IP aleatórios, Máquina 2; *Firewall Fortinet* para gerar eventos com *Media Access Control* - MAC de um usuário, Máquina 3; ferramenta SIEM *Graylog* para recebimento dos eventos.
- Etapa 2: Análise da Lei Geral de Proteção de Dados (Lei 13.709/2018) e Autoridade Nacional de Proteção de Dados - ANPD para aplicação de técnica conforme Art. 5º, inciso XI da referida lei.
- Etapa 3: Estudo da documentação do Siem *Graylog* para entendimento do funcionamento da ferramenta e da lógica da Linguagem Específica de Domínio - DSL produzida pela *Graylog Inc*.

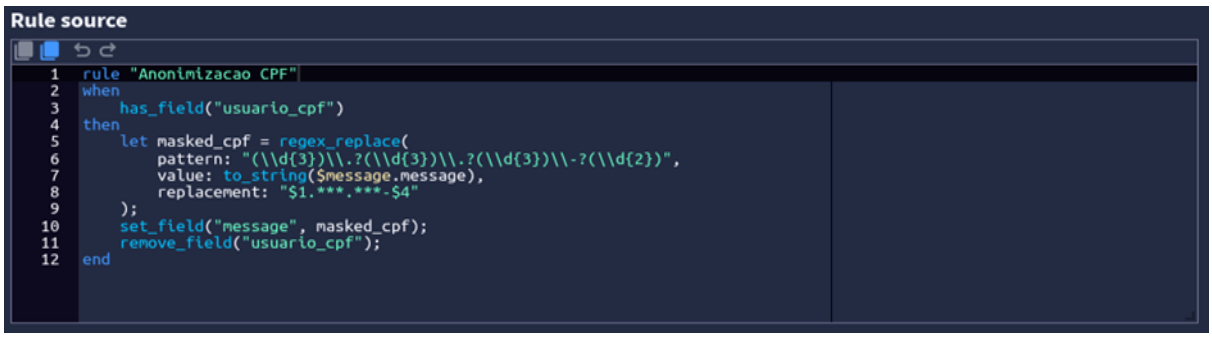
- Etapa 4: Configuração do SIEM *Graylog* para recebimentos dos eventos e extração dos identificadores diretos e indiretos contidos nos *logs* gerados.
- Etapa 5: Desenvolvimento e aplicação do código na função *pipeline* através das *rules* para aplicação da técnica de mascaramento conforme orientação do estudo técnico sobre anonimização de dados na LGPD realizado pela ANPD (AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS, 2023b, p. 20-26).

3 Resultados e discussões

3.1 Anonimização de identificador direto

Um clássico de identificador direto de um titular de dados é o nome completo. Além disso, outro dado comum é o número do Cadastro de Pessoas Físicas (CPF), que, passou a ser reconhecido como o número único para identificação do cidadão nos bancos de dados de serviços públicos que fica estabelecido com a vigência da Lei n.º 14.534/2023 (AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS, 2023a, p. 6-7). Nessa ótica, o Art. 12º da Lei 13.709/2018, específica de maneira evidente que o dado anonimizado para fins de cumprimento da referida lei, não será considerado dado pessoal. Nesse contexto, a técnica de mascaramento é suficientemente compatível com a perspectiva da Lei Geral de Proteção de Dados Pessoais.

Figura 2 - Código para anonimização de CPF



Fonte: Elaborada pelos autores utilizando a ferramenta *Graylog* (2025)

Para gerar o resultado esperado e seguir o que se pede na Lei Geral de Proteção de Dados (Lei 13.709/2018), para auxílio e validação do código é criado uma simulação de uma aplicação via *script bash* gerando eventos com CPF aleatório, além disso, realizado a extração personalizada do campo, para em seguida aplicar técnica de mascaramento com *pipelines* e *rules* utilizando a linguagem específica de domínio desenvolvida pela *Graylog Inc.* especificamente para expressar lógica de processamento dos eventos dento do sistema SIEM *Graylog* (Graylog Inc., 2025). A figura 2 representa o processo para anonimização do CPF:

Seguindo esse raciocínio, a observar a figura 2, na linha 1 a função *rule* determina o nome da regra, nas linhas 2 a 3 verifica se o campo **"*usuario_cpf*"** existe no evento, nas linhas 5 a 9

substitui o segundo grupo e o terceiro grupo de caracteres do CPF. Na linha 10, é criando um novo campo **"*masked_cpf*"** dentro do campo padrão **"*message"*** e consequentemente sobrescrito com a informação CPF mascarado e, logo após, por segurança, remove o campo ***"usuario_cpf"*** (não mascarado) na linha 11 com a função ***"remove_field()"***. Após a aplicação do código, a figura 3 representa o resultado esperado com a técnica de mascaramento para anonimizar o identificador direto CPF visível no *log*.

Figura 3 - Evento com o CPF anonimizado

Fonte: Elaborada pelos autores utilizando a ferramenta *Graylog* (2025)

A anonimização pode ser aplicada por diversas maneiras, conforme expresso no estudo técnico sobre anonimização realizado pela Autoridade Nacional de Proteção de Dados (ANPD) (2023, p. 21), foi aplicado técnica de mascaramento do dado direto CPF anonimizado com o processo acima, substituindo parte dos caracteres por símbolos, dessa maneira, os dados armazenados no sistema *Graylog*, pode-se perceber a conformidade alcançada com a técnica aplicada.

3.2 Anonimização de Identificador Indireto

Dados indiretos, por si só, não tem a capacidade de identificação de um indivíduo, mas um conjunto desses dados associados a informações auxiliares tem total possibilidade de caracterizar, personificar uma pessoa natural (AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS., 2023, p. 7).

Como já mencionado anteriormente, o Art. 5º da Lei 13.709/2018 traz uma definição do que é um dado pessoal, que são informações que possa identificar, caracterizar uma pessoa natural. Para facilitar o entendimento, *National Institute of Standards and Technology (NIST)* tem uma definição, amplamente referenciada, sobre Informações que permitem identificação Pessoal (PII):

> [5]Informações de identificação pessoal; Qualquer representação de informação que permita que a identidade de um indivíduo a quem a informação se aplica seja razoavelmente inferida por meios diretos ou indiretos.

[5] *Personally Identifiable Information; Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means.*
*Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information (*NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*, 2025).*

> Qualquer informação sobre um indivíduo mantida por uma agência, incluindo (1) qualquer informação que possa ser usada para distinguir ou rastrear a identidade de um indivíduo, como nome, número de segurança social, data e local de nascimento, nome de solteira da mãe ou registos biométricos; e (2) qualquer outra informação que esteja vinculada ou possa ser vinculada a um indivíduo, como informações médicas, educacionais, financeiras e de emprego (NIST, 2025. D.p.).

O conceito geral traz de maneira sucinta que quaisquer informações relacionadas a uma pessoa específica podem ser consideradas de natureza "pessoal". Diante disso, sistemas SIEM tem a capacidade de armazenar diversas informações, a exemplo disso, embora o endereço *Media Access Control* (MAC) e o endereço IP sejam, tecnicamente, identificadores de hardware e não contenha informações pessoais diretamente, eles podem ser associados a outros dados para identificar um indivíduo. Desse modo, de posse do endereço MAC e endereço IP, ambos podem ser utilizados para monitorar a localização ou os hábitos de navegação de um usuário conectado à *internet*. Nesse sentido, quando esses endereços são coletados e processados em conjunto com outros dados pelo SIEM, eles passam a ser considerado um dado pessoal sob a Lei 13.709/2018 (LGPD) devido à viabilidade de correlação de eventos que permite a identificação de uma pessoa com essas informações.

Figura 4 - Código para anonimização do endereço MAC

Rule source

```
1  rule "Mascarar endereço MAC"
2  when
3      has_field("srcmac")
4  then
5      let masked_mac = regex_replace(
6          pattern: "([0-9A-Fa-f]{2}:[0-9A-Fa-f]{2}):[0-9A-Fa-f]{2}:[0-9A-Fa-f]{2}:[0-9A-Fa-f]{2}",
7          value: to_string($message.message),
8          replacement: "$1:**:**:**"
9      );
10     set_field("message", masked_mac);
11     remove_field("mastersrcmac");
12     remove_field("srcmac");
13 end
14
```

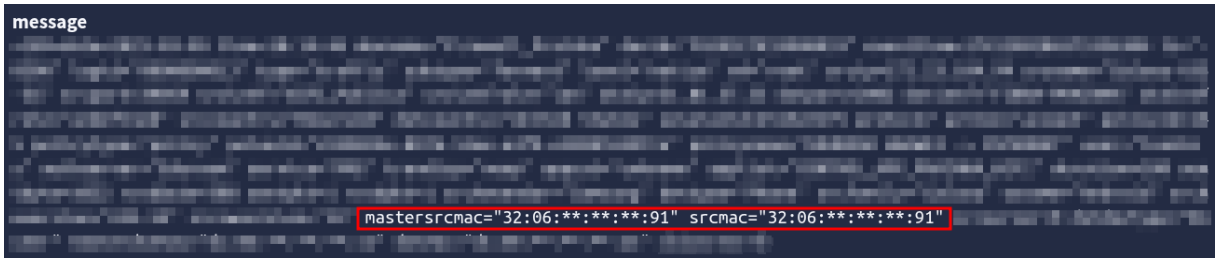
Fonte: Elaborada pelos autor utilizando a ferramenta *Graylog* (2025)

Nesse sentido, para anonimizar dados considerados indiretos, na figura 4, demonstra o código desenvolvido aplicando a técnica de mascaramento do endereço MAC nos eventos gerados pelo *firewall Fortinet*, armazenados no Siem *Graylog,* utilizando a linguagem específica de domínio da *Graylog Inc*.

Na linha 1, a observar a figura 4, a função *rule* determina o nome da regra, nas linhas 2 a 3 verifica se o campo ***"srcmac"*** existe no evento, nas linhas 5 a 9 substitui os últimos caracteres de qualquer endereço MAC existente no evento. Na linha 10, é criando um novo campo denominado ***"masked_mac"*** dentro do campo padrão ***"message"*** consequentemente sobrescrito com a informação do endereço MAC mascarado, e logo após, por segurança, remove os campos que contém o MAC de origem na linha 11 e 12 com a função

"remove_field()". Ademais, na figura 5, mostra o resultado após a aplicação do código acima, que representa a técnica de mascaramento dos campos que contém endereço MAC:

Figura 5 - Evento com o MAC anonimizado



Fonte: Elaborada pelos autores utilizando a ferramenta *Graylog* (2025)

Para gerar o resultado esperado do mascaramento do endereço IP é realizado a uma simulação de aplicação via *script bash* produzindo diversos endereços IP aleatórios, para geração de eventos com o identificador indireto supracitado para teste do código utilizado nas *rules* do *pipeline*. A figura 6 demonstra a aplicação do código de mascaramento conforme medidas razoáveis expressa na Lei Geral de Proteção de Dados Pessoais:

Figura 6 - Código para anonimização do endereço IP

Fonte: Elaborada pelos autores utilizando a ferramenta *Graylog* (2025)

Na linha 1, a observar a figura 6, a função *rule* determina o nome da regra, nas linhas 2 a 3 verifica se o campo ***"usuario_ip"*** existe no evento, nas linhas 5 a 9 substitui os últimos caracteres do endereço IP. Na linha 10, é criado um novo campo denominado ***"masked_ip"*** com a informação dos endereços IP mascarados, e logo após, por segurança, remove o campo original na linha 11 com a função ***"remove_field()"***. Com isso, na figura 7, mostra o resultado após a aplicação da técnica de mascaramento dos campos que contém endereço IP:

Figura 7 - Evento com o endereço IP anonimizado

Fonte: Elaborada pelos autores utilizando a ferramenta *Graylog* (2025)

O mascaramento do endereço IP, no contexto da identificação de uma pessoa natural na *internet*, é essencial para a proteção da privacidade, aumentando a segurança e garantindo o anonimato no acesso livre à informação no meio digital. Sua adoção é recomendada tanto por

usuários comuns quanto para ambientes cooperativos, dando um destaque para ambientes de trabalho remoto, uso em redes públicas e, por se tratar de um identificador indireto capaz de identificação de um indivíduo, pode estar sujeito às sanções da Lei Geral de Proteção de Dados Pessoais.

4 Considerações finais

Diante do exposto, com o advento da *internet* trouxe muitas discussões relacionadas a privacidade das pessoas e segurança dos dados. Destaca-se que ferramentas SIEM são altamente recomendadas no âmbito da segurança da informação e cibersegurança, foi possível mostrar, que sua utilização é considerada com uma prática positiva para ambientes que contém informações sujeita a proteção, tanto na Lei Geral de Proteção Dados (Lei 13.709/2018) quanto em regulamentações e normas internacionais, evidenciando o quanto é indispensável para qualquer negócio. Entretanto, sem os devidos cuidados é possível que essas ferramentas armazenem informações classificadas como identificadores diretos e indiretos de forma clara sem as devidas tratativas.

Avançou-se, em seguida, com a resolução desse problema, aplicando técnica de mascaramento para anonimizar os dados conforme indicado pela Autoridade Nacional de Proteção de Dados de maneira a transformar o dado em uma informação não identificável, utilizando o SIEM *Graylog*, desenvolvendo códigos com linguagem específica de domínio empregada nas *rules* com *pipeline*, perdendo a capacidade de identificar uma pessoa natural e consequentemente estar consoante com a Lei Geral de Proteção Dados (Lei 13.709/2018).

Outrossim, a relevância do tema se torna imprescindível diante do grande volume de informações que sistemas SIEM podem armazenar, além do aprimoramento das técnicas de cruzamento e correlacionamento de dados, potencialmente capaz de caracterizar/personificar uma pessoa natural. Diante disso, com os processos de anonimização aplicados às informações, obtiveram-se conclusões fundamentais extraídas dessa análise, destacando a importância de manter sistemas de gerenciamento de eventos de segurança conforme a legislação vigente.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27001: Segurança da informação, segurança cibernética e proteção à privacidade — Sistemas de gestão da segurança da informação — Requisitos**. 3. ed. Rio de Janeiro: ABNT, 2022a.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27002. Segurança da informação, segurança cibernética e proteção à privacidade - Controles de segurança da informação**. 3ed., Rio de Janeiro: ABNT, 2022c.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Estudo técnico sobre a anonimização de dados na LGPD: A**nálise Jurídica. Brasília: ANPD, 2023b. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd___analise_juridica.pdf. Acesso em: 16 dez. 2025.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Estudo técnico sobre anonimização de dados na LGPD:** uma visão de processo baseado em risco e técnicas computacionais. Brasília: ANPD, 2023a. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd_uma_visao_de_processo_baseado_em_risco_e_tecnicas_computacionais.pdf. Acesso em: 16 dez. 2025.

BRASIL. **Emenda Constitucional Nº 115, de 10 de fevereiro de 2022.** Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais., Diário Oficial da União, p. 2, 2022. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/Emendas/Emc/emc115.htm. Acesso em: 16 dez. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD)**. Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 16 dez. 2025.

ESTADOS UNIDOS. **45 CFR Part 164 Subpart C – Security Standards for the Protection of Electronic Protected Health Information**. Department of Health and Human Services. 2013. Disponível em: https://www.ecfr.gov/current/title-45/part-164/subpart-C . Acesso em: 16 dez., 2025.

EUROPEAN, Union. General Data Protection Regulation. **REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)nGDPR**, *[S. l.]*, 2016. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. Acesso em: 16 dez. 2025.

FALEIROS JÚNIOR, J. L. M.; MARTINS, G. M. Proteção de Dados e Anonimização: Perspectivas à luz da Lei n.º 13.709/2018. **Rei - Revista Estudos Institucionais**, [S. l.], v. 7, n. 1, p. 376–397, 2021. Disponível em: https://doi.org/10.21783/rei.v7i1.476. Acesso em: 16 dez. 2025.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**, 7 ed., São Paulo: Atlas, Grupo GEN, 2019.

GRAYLOG, INC. **Pipeline Rule Logic**. 2025. Disponível em: https://go2docs.graylog.org/current/making_sense_of_your_log_data/rules.html#Rule. Acesso em: 2 mar. 2025.

KOTENKO, I. *et al*. Design and implementation of a hybrid ontological-relational data repository for SIEM systems. **Future internet**, v. 5, n. 3, p. 355-375, 2013. Disponível em: https://doi.org/10.3390/fi5030355. Acesso em: 16 dez. 2025.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Glossário: PII**. Gaithersburg: NIST-U.S. Department of Commerce, 2025. Disponível em: https://csrc.nist.gov/glossary/term/PII . Acesso em: 17 dez. 2025.

NEU, C. et al. Extração e gerenciamento de incidentes em SIEM. *In*: ESCOLA REGIONAL DE REDES DE COMPUTADORES (ERRC), 17, 2019, Alegrete. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019, p. 190-195. Disponível em: https://doi.org/10.5753/errc.2019.9236. Acesso em: 16 dez., 2025.

PAVLIK, J. et al. Security information and event management in the cloud computing infrastructure. In: **IEEE 15TH INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL INTELLIGENCE AND INFORMATICS**, 2014, [S.l.]. Proceedings... [S.l.]: IEEE, 2014. Disponível em: https://doi.org/10.1109/CINTI.2014.7028677 . Acesso em: 16 dez., 2025.

PRATA, P. *et al*. Garantia de privacidade versus utilidade dos dados em anonimização: um estudo no ensino superior**. Risti**, Porto-Portugal, n. 40, p. 112-127, dez. 2020. Disponível em: https://scielo.pt/pdf/rist/n40/1646-9895-rist-40-112.pdf. Acesso em: 16 dez., 2025.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, M. D. P B. **Metodologia de pesquisa.** Porto Alegre: Penso, 2013.

SEVERINO, A. J. **Metodologia do trabalho científico**. 24 ed., São Paulo: Cortez, 2017.

SOUZA, J. F. Privacidade e dados pessoais: o debate ético sobre o uso de big data. Revista Ilustração, [S. l.], v. 5, n. 6, p. 27–51, 2024. Disponível em:https://doi.org/10.46550/ilustracao.v5i6.340. Acesso em: 16 dez., 2025.

TARIQ, A. *et al*. Open source SIEM solutions for an enterprise.**Information & Computer Security**, v. 31, n. 1, p. 88-107, 2023. Disponível em:https://doi.org/10.1108/ICS-09-2021-0146. Acesso em: 16 dez., 2025.

VAZÃO, A. *et al*. SIEM open source solutions: a comparative study. *In*: IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES (CISTI), 14., 2019, Coimbra. **Anais** [...]. Coimbra: IEEE, 2019. p. 1–5. Disponível em:https://doi.org/10.23919/CISTI.2019.8760980. Acesso em: 16 dez., 2025.