v. 14, n. 1, p. 30-45, 2025
ISSN 2237-9223

REVISTA BRASILEIRA DE CRIMINALÍSTICA

DOI: http://dx.doi.org/10.15260/rbc.v14i1.893

Análise de logs de rede sem fio com suporte à tomada de decisão na infraestrutura Tecnologia da informação

R.F. Rodrigues [a,*], F.C.L. Ferreira [a], D.A. Gomes [a]

[a] *Programa de Pós-Graduação em Ciências Forenses, Universidade Federal do Sul e Sudeste do Pará, Marabá (PA), Brasil*

**Endereço de e-mail para correspondência: roberto.fr@unifesspa.edu.br. Tel.: +55-94-98400-2653.*

Recebido em 12/12/2024; Revisado em 19/02/2005; Aceito em 26/03/2025

Resumo

Nos últimos anos, o governo federal brasileiro tem dado mais atenção à segurança cibernética, culminando, em 2022, no desenvolvimento de uma política para a gestão de *logs* de sistemas computacionais. Esses *logs*, gerados em grande volume, podem ser analisados para identificar padrões e compreender eventos, falhas e violações de segurança. A análise inadequada desses registros dificulta investigações detalhadas em casos de anomalias e crimes cibernéticos, devido à quantidade de dados em formatos pouco amigáveis. Na Universidade Federal do Sul e Sudeste do Pará (Unifesspa), assim como em muitos órgãos públicos, há um grande volume de *logs* gerados em resposta a anomalias nos sistemas de TIC, que podem ser utilizados para prevenir riscos e bloquear ações perigosas em tempo real. Este trabalho visa desenvolver um sistema que automatize o gerenciamento de *logs* de forma eficiente, apoiando investigações e decisões no setor de infraestrutura TIC. Técnicas de análise de *logs* foram implementadas, permitindo que a equipe de TI tome decisões de maneira ágil e precisa. Os resultados indicaram que a aplicação desenvolvida alcançou uma impressionante eficiência de 99,18% na categorização dos registros durante sua validação. Apenas 0,82% do total, o que corresponde a aproximadamente 2.508 registros, não foram categorizados corretamente. A análise detalhada desses dados não apenas facilita a identificação de problemas, mas também é crucial para mitigar riscos e assegurar a segurança da rede. No futuro, planeja-se aprimorar a aplicação, incorporando um dashboard que permitirá uma visualização mais intuitiva e uma consulta otimizada dos dados gerados.

Palavras-chave: Análise de dados, forense digital, análise de logs.

Abstract

In recent years, the Brazilian federal government has given more attention to cybersecurity, culminating in 2022 in the development of a policy for the management of computer system *logs*. These *logs*, generated in large volumes, can be analyzed to identify patterns and understand events, failures, and security breaches. Inadequate analysis of these records hinders detailed investigations in cases of anomalies and cyber crimes due to the amount of data in less user-friendly formats. At the Federal University of the South and Southeast of Pará (Unifesspa), as well as in many public agencies, there is a large volume of *logs* generated in response to anomalies in ICT systems, which can be used to prevent risks and block dangerous actions in real-time. This work aims to develop a system that automates *log* management efficiently, supporting investigations and decision-making in the ICT infrastructure sector. *Log* analysis techniques have been implemented, allowing the IT team to make decisions quickly and accurately. The results indicated that the developed application achieved an impressive 99.18% efficiency in categorizing records during its validation. Only 0.82% of the total, which corresponds to approximately 2,508 records, were not categorized correctly. Detailed analysis of this data not only facilitates the identification of problems but is also crucial for mitigating risks and ensuring network security. In the future, there are plans to enhance the application by incorporating a dashboard that will allow for more intuitive visualization and optimized querying of the generated data.

Keywords: Data analysis, digital forensics, log analysis.

1. INTRODUÇÃO

No contexto da Tecnologia da Informação (TI), uma infraestrutura de rede que hospeda um sistema ou serviço geralmente tem a capacidade de gerar registros de atividades e padrões de uso de uma aplicação, servidor ou sistema de Tecnologia da Informação e Comunicação (TIC), chamados de *Logs*. Esses registros podem ser considerados como indicadores de eventos relacionados a condições inadequadas ou falhas no processo de execução ou autenticação de um sistema. Segundo [1, 11, 12], os *logs* são textos semiestruturados gerados por instruções de *log* no código-fonte do software. Esses *logs* podem ser gerados de múltiplas fontes, como *software* de segurança, antivírus, *firewall*, sistemas de prevenção e detecção de intrusão, sistemas operacionais, estações de trabalho, aplicações e equipamentos de rede [2]. Para [3], nos últimos anos, os *logs* de software se tornaram essenciais e verdadeiros aliados dos profissionais de TI para garantir a confiabilidade de muitos sistemas, pois muitas vezes representam os únicos dados disponíveis que registram informações sobre o tempo de execução do software.

A análise de *logs* consiste em atividade benéfica para os administradores pois fornecem uma visão geral do que está acontecendo no sistema e podem identificar incidentes, violações de políticas, atividades fraudulentas e problemas operacionais que estão ocorrendo na infraestrutura [2]. Os dados de *logs* gerados contêm informações que são fontes extremamente valiosas de informações para diagnosticar a causa raiz de problemas complexos em um sistema de TIC [4]. De acordo com a Norma Brasileira (NBR) 27001:2013 [5], os registros de *logs* são tão valiosos que devem ser produzidos, mantidos, protegidos de acesso não autorizado e analisados de forma crítica em intervalos regulares.

Contudo existe uma grande complexidade na análise desses dados e na obtenção de informações claras e precisas, dada a amplitude de informações produzidas e o formato desses textos semiestruturados. Para [6], a detecção de ameaças é comumente obtida por meio da análise de *logs*, que é uma abordagem amplamente utilizada. No entanto, com o aumento exponencial dos dados produzidos e a modernização dessas aplicações, há necessidade de métodos de análise melhorados que vão além da abordagem tradicional. De acordo com [4], desenvolver a habilidade de analisar *Logs* de forma rápida e precisa é essencial para reduzir o tempo de inatividade do sistema e detectar problemas operacionais antes ou durante a ocorrência.

Segundo [7], a análise de *log* é o processo de transformar dados brutos em informações úteis para resolver problemas. Na área de análise de *log* e diagnóstico de falhas, a correlação de dados é uma questão importante. Tem como objetivo reconstruir a falha do sistema agrupando *logs* que possuem a mesma fonte de falha. Isso se baseia no fato de que muitos detectores de software e hardware são acionados pela mesma falha [8]. Além disso, os registros também são úteis na realização de auditorias e análises forenses, no apoio a investigações internas, no estabelecimento de linhas de base e na identificação de tendências operacionais e problemas de longo prazo [2].

Nos últimos anos, o governo federal brasileiro tem dado atenção à importância desse tema. Em 2022 a secretaria de governo digital do Ministério da Gestão e Inovação em Serviços Públicos, com o intuito de contribuir com os órgãos do governo federal brasileiro, desenvolveu um modelo de política de gestão de registros de *logs* para fins de auditoria [2]. Esse documento fornece boas práticas para implementação de gestão de *logs*, de auditoria e orientações para tratamento de riscos ligados às temáticas de privacidade e segurança da informação relativos aos seus sistemas informacionais, contratos administrativos e processos de trabalho do órgão ou entidade [2].

Diante disso, a perícia forense computacional desempenha um papel fundamental nesse contexto, especializada na investigação de crimes digitais e focada na apuração dos fatos e na coleta de dados que servirão como evidências que combinam conhecimentos da área de informática com o campo jurídico [14]. Utilizando uma variedade de procedimentos técnicos periciais para identificar crimes e seus responsáveis, a perícia atua sempre em conformidade com as diretrizes da investigação, assegurando que as evidências sejam preservadas e não se percam [14]. Além disso, a perícia forense tem como objetivo determinar a dinâmica, a materialidade e a auditoria de ilícitos relacionados à área de informática, sendo seu principal foco a identificação e o processamento de evidências digitais como provas materiais de crimes, por meio de métodos técnico-científicos, conferindo-lhes validade probatória em juízo [15].

No contexto da Universidade Federal do Sul e Sudeste do Pará (Unifesspa), a rede sem fio enfrenta diversos desafios que impactam diretamente sua eficiência e segurança. Um dos principais problemas está relacionado à dificuldade em monitorar e analisar os dados de logs gerados pelas autenticações dos usuários. Esses logs, que são fundamentais para identificar comportamentos suspeitos, falhas operacionais e incidentes de segurança, não estão sendo devidamente processados e transformados em informações acessíveis e úteis para os profissionais de TI. A ausência de uma solução especializada para a análise de logs afeta diretamente a capacidade da equipe em detectar e responder de forma ágil a problemas operacionais e de segurança. Isso pode resultar em riscos potenciais para a

integridade e disponibilidade da infraestrutura de rede. Isso inclui questões como acessos não autorizados, falhas de autenticações, vazamentos de credenciais, entre outros. Além disso, a ausência de uma ferramenta de análise de *logs* pode levar a atrasos na identificação de padrões de uso que poderiam ajudar na otimização da rede e no planejamento estratégico. Esse cenário não só prejudica a eficiência operacional, mas também impacta negativamente a experiência do usuário final, que pode enfrentar problemas de conectividade e instabilidades no serviço.

Outro problema crítico é a ineficiência na apresentação e organização desses dados. Sem uma análise estruturada e clara, os administradores enfrentam dificuldades em realizar investigações detalhadas ou tomar decisões embasadas para resolver problemas cotidianos. Isso acarreta atrasos na detecção e solução de incidentes, comprometendo a segurança e experiência do usuário final. Dentro deste contexto, este trabalho visa desenvolver um sistema que automatize o gerenciamento de logs de forma eficiente e eficaz. O objetivo é fornecer um suporte robusto às investigações e à tomada de decisões no setor de infraestrutura de TIC. Com isso, pretende-se viabilizar a rápida identificação de problemas operacionais, a prevenção de falhas potenciais e a manutenção da confiabilidade e eficiência dos sistemas de rede. Além de que, a implementação desse sistema contribuirá para a mitigação de riscos, o aprimoramento da segurança, a agilidade na capacidade de resposta e a melhoria da qualidade dos serviços oferecidos tanto aos administradores quanto aos usuários da rede sem fio.

Além disso, pretende-se realizar a integração entre boas práticas de gestão de *logs* e a atuação da perícia forense, o que é crucial para fortalecer a segurança da informação e a transparência nas ações de segurança digital no contexto do governo federal.

No contexto da análise de evidências digitais, a implementação de um sistema de análise de *logs* que centralize os dados coletados pode oferecer uma série de vantagens significativas para os investigadores. Ao centralizar as informações, a ferramenta estabelece um repositório de backup robusto, que impede que um atacante apague todos os registros durante um acesso não autorizado ao sistema de computação.

Essa centralização não apenas preserva a integridade dos dados, mas também facilita a investigação ao permitir que os analistas acessem rapidamente informações cruciais sobre atividades suspeitas. Com um sistema de *logs* centralizado, é possível identificar padrões de comportamento, rastrear ações realizadas por usuários e detectar anomalias em tempo hábil de tratamento.

Além disso, a análise de *logs* centralizada pode contribuir para a criação de um histórico detalhado das atividades do sistema, o que é essencial para a elaboração de relatórios forenses e para a apresentação de evidências em processos judiciais. Dessa forma, a adoção de uma ferramenta eficaz de análise de *logs* não apenas fortalece a segurança da informação, mas também aprimora a capacidade dos investigadores de responder a incidentes de segurança, aumentando a proteção contra ameaças cibernéticas e garantindo a confiança no ambiente digital.

2. MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho, foram implementadas técnicas para análise dos *logs*, aliada ao desenvolvimento de uma aplicação que capacite a equipe de TI na tomada de decisões de forma mais ágil e precisa. Essa aplicação não apenas simplificará a interpretação dos dados de autenticações, mas também fornecerá feedbacks valiosos para aprimorar a eficiência e a segurança da rede sem fio.

Os *logs* consistem em dados semiestruturados ou não estruturados que podem ser gerados por uma ampla variedade de serviços e aplicações, abrangendo desde servidores, *firewalls* e roteadores, até sistemas de antivírus e *antimalware*. Para analisar esses dados, é necessário realizar algumas etapas essenciais. Neste contexto, as etapas foram divididas em coleta de dados, normalização de registros e análise dos dados.

A metodologia empregada nesta pesquisa foi desenvolvida de forma a permitir sua fácil replicação em ambientes que utilizam a tecnologia de rede sem fio da empresa *Ruckus Networking*. A *Ruckus Networking* é uma marca de equipamentos e software de rede com fio e sem fio. A empresa é reconhecida por oferecer uma infraestrutura de rede projetada para atender às exigências rigorosas de diversos ambientes empresariais, que fornece uma plataforma única para gerenciar várias redes de acesso com provisionamento *zero-touch* que elimina a intervenção manual e configura automaticamente os dispositivos de rede. Isso permite que as empresas escalem a implantação de dispositivos em vários locais e implementem um gerenciamento simples e intuitivo. Essa abordagem não apenas proporciona uma base sólida para a implementação, mas também garante que os resultados obtidos sejam consistentes e aplicáveis em ambientes que empregam tecnologia equivalente.

A partir do exposto, a Figura 1 retrata a metodologia adotada para chegarmos aos resultados deste trabalho, onde será seguido o seguinte percurso:

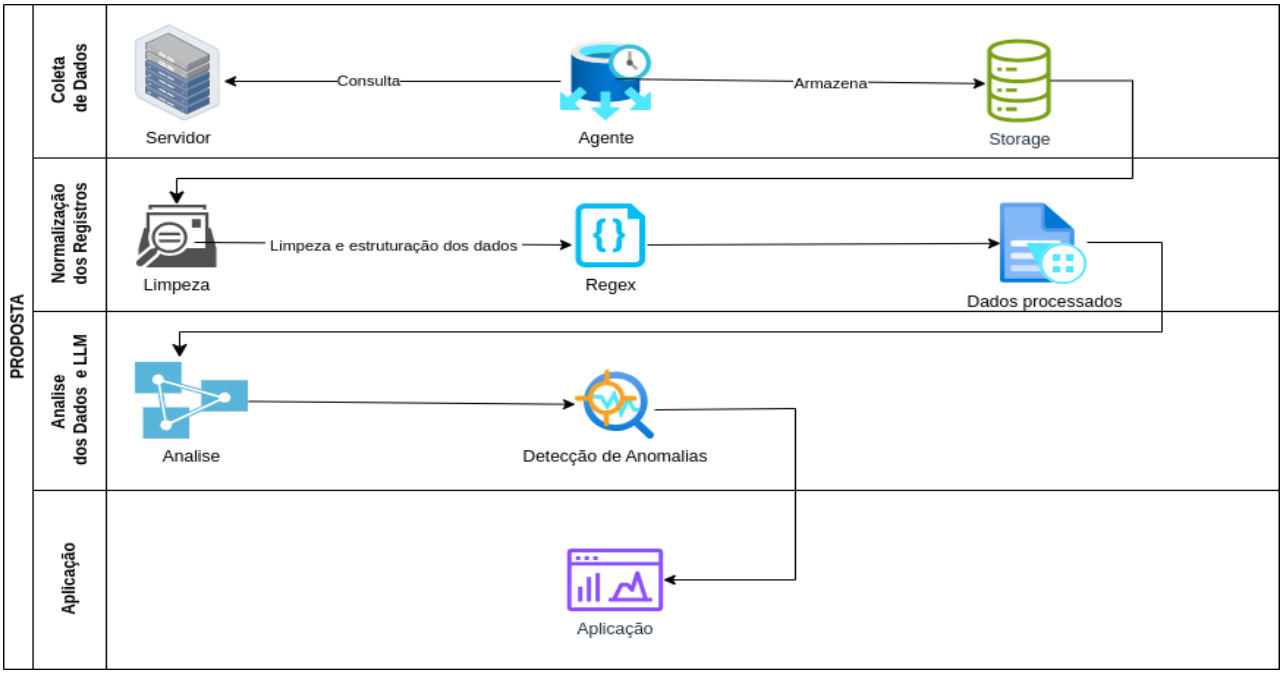

Figura 1. Representação da metodologia de trabalho da proposta

Coleta de Dados: Nesta etapa, é realizada a coleta dos registros provenientes das atividades e eventos na rede sem fio.

Normalização dos Registros: Em seguida, os dados coletados passaram por um processo de normalização, garantindo a consistência e eliminando redundâncias, preparando-os para uma análise precisa e confiável.

Análise dos Dados: Nesta etapa, os registros normalizados foram submetidos a uma análise de vestígios detalhada.

Desenvolvimento da Aplicação: Por fim, será desenvolvida a aplicação que proporcionará de forma intuitiva o acesso aos recursos necessários para a realização de análises periciais.

Para alcançar os resultados esperados neste trabalho, empregamos diversas ferramentas que foram essenciais para atingir os objetivos desejados. Essas ferramentas desempenharam um papel crucial no desenvolvimento do projeto, oferecendo suporte significativo e contribuindo para seu sucesso. As principais ferramentas utilizadas até o momento incluem:

Google Colab - O Google Colab é uma ferramenta baseada em nuvem que permite a incorporação de cadernos de códigos. Uma grande vantagem é que todo o código é executado na linguagem Python. Além disso, a possibilidade de acessar recursos de computação poderosos, como GPUs e *Tensor Processing Unit* (TPU)s, que aceleram o processamento de tarefas de computação intensivas, é outra vantagem do Google Colab.

Python - Python é uma linguagem de programação de alto nível que possui diversas aplicações, desde a automação de tarefas até a implementação de algoritmos de *machine learning* (ML). No âmbito da análise de *logs*, a utilização do *Python* no Google Colab oferece uma série de vantagens. Além disso, a grande quantidade de bibliotecas disponíveis para *Python*, como PANDAS, NUMPY e MATPLOTLIB, permite a manipulação e visualização eficiente dos dados de *logs*.

Biblioteca *Python* PANDAS - A biblioteca pandas é uma biblioteca do Python amplamente utilizada para análise e manipulação de dados. Ela fornece estruturas de dados e funções essenciais para simplificar o processo de manipulação e análise de dados. O pandas permite a manipulação de estruturas de dados tabulares como o DataFrame, que é uma estrutura de dados que organiza os dados em linhas e colunas em duas dimensões. Além disso, oferece funcionalidades avançadas para a manipulação de dados, permitindo operações como filtragem, agrupamento, ordenação e transformação de conjuntos de dados.

Biblioteca *Python* RE- A biblioteca re do Python suporta expressões regulares REGEX. Os padrões de texto conhecidos como expressões regulares permitem a busca, extração e modificação eficaz de strings. Ela oferece uma variedade de funções e abordagens para trabalhar com expressões regulares de forma flexível e poderosa. A utilização da biblioteca re é especialmente útil no contexto da análise de *logs*, onde é necessário extrair informações específicas dos registros de *logs*.

Biblioteca *Python* PLOTLY - A biblioteca PLOTLY é uma biblioteca de visualização interativa e de alta qualidade para *Python*. Ela permite a criação de diversos tipos de gráficos, como gráficos de linha, gráficos de

dispersão, gráficos de área, gráficos de barras, gráficos de caixa, histogramas, mapas de calor, entre outros.

Biblioteca *Python* STREAMLIT - É uma biblioteca *Python* de código aberto que combina recursos poderosos de visualização com a simplicidade da sintaxe do *Python*. O STREAMLIT é uma ferramenta incrivelmente poderosa e flexível, que permite aos desenvolvedores criar aplicativos da *Web* impressionantes e interativos, permitindo atualizações rápidas e em tempo real do aplicativo *web*.

2.1. Coleta de Dados

Nesta etapa de desenvolvimento da pesquisa, foi disponibilizado pela equipe de infraestrutura de Tecnologia da Informação um conjunto de registros de *logs* do serviço de rede sem fio. Esse conjunto de registros é composto por dados de *logs* de um período de sessenta dias, abrangendo os anos de 2021 a 2023. No total, foram disponibilizados cinquenta e dois arquivos de *logs*, que juntos contém um milhão, cento e cinquenta e um mil, duzentos e cinquenta registros. Esses dados são a base para nossa análise e para o desenvolvimento da ferramenta de visualização dos *logs* da rede sem fio.

Segundo [13] os eventos de *logs* incluem informações como carimbo de data e hora, gravidade, mensagem, origem da mensagem, mudança dinâmica ocorrendo em software e hardware, entre muitos outros. É crucial notar que os registradores de dados não são padronizados ou seguem um determinado formato estrutural unificado. Os dados coletados nos *logs* contém informações detalhadas, incluindo dia da semana, mês, data, horário, ID, status do *login*, nome do usuário, rede *Wi-Fi*, porta de conexão, protocolo de tunelamento e endereço *MAC* (*Media Access Control*) do dispositivo. Em Dados 1, apresentamos alguns registros de logs que são comumente encontrados em dispositivos de *Wi-Fi*. É importante ressaltar que a organização e a padronização desses dados podem variar conforme as diretrizes estabelecidas por diferentes fabricantes ou sistemas. Para garantir a segurança e a privacidade dos dados analisados, os usuários registrados nos *logs* foram anonimizados substituindo o *login* pela palavra user.

Dados 1 - Registros de *logs* coletados

1- Sun Nov 7 15:17:44 2021 : Auth: (842830) Login OK: [user@unifesspa.edu.br] (from client wifi port 1 cli A8-96-75-A7-91-AB)

2- Fri May 5 20:14:28 2023 : Auth: (5919820) Login incorrect: [user@unifesspa.edu.br] (from client wifi-c3 port 2 cli 9A-8F-BC-B1-94-79)

3- Fri Oct 22 09:29:18 2021 : Auth: (38373) Invalid user (Rejected: User-Name contains whitespace): [user@unifesspa.edu.br] (from client wifi port 4 cli 68-7D-6B-46-D5-E3)

Os dados provenientes dos cinquenta e dois arquivos de *logs* coletados foram mesclados em um único arquivo, utilizando a implementação apresentada no Código 1, que assegura uma integração eficaz das informações.

Código 1 - Mesclagem de dados

```
padrao_arquivos = os.path.join(caminho_pasta, '*.log*')
caminhos_arquivos = glob.glob(padrao_arquivos)
dados_totais = []
for caminho_arquivo in caminhos_arquivos:
  with open(caminho_arquivo, 'r') as arquivo:
    linhas = arquivo.readlines()
    dados_totais.extend(linhas)
dados_df = pd.DataFrame({'Logs': dados_totais})
print("Dados salvos em dados_totais.csv")
```

Este processo de mesclagem não apenas simplifica o manuseio dos dados, mas também otimiza a análise subsequente, garantindo que todas as informações relevantes estejam centralizadas e organizadas de maneira coerente.

2.2. Normalização dos Registros

Nesta etapa, é realizada a limpeza dos registros, removendo os dados duplicados. Essa etapa visa garantir a integridade e a precisão dos dados armazenados, evitando redundâncias e inconsistências que possam impactar negativamente as análises e consultas futuras.

Para otimizar a análise dos registros, foi empregada uma técnica de processamento de dados que consiste na eliminação de todas as linhas que contêm a sigla do protocolo *Transport Layer Security (TLS)*. Essa decisão foi tomada porque, para cada registro de autenticação, eram geradas uma linha com informações redundantes e que não contribuem para os objetivos da pesquisa. Essa ação é especialmente relevante no contexto dos *logs* sem fio que utilizam a tecnologia da *Ruckus Networks*, pois permite filtrar dados desnecessários e focar nas informações essenciais. Ao garantir que os dados analisados sejam pertinentes e alinhados com os objetivos da pesquisa, essa abordagem não apenas melhora a qualidade da análise, mas também torna o processo mais eficiente.

Nos registros apresentados em Dados 2, são exibidas informações referentes a uma autenticação ocorrida na rede. Na linha 1, estão os dados dos registros que foram

removidos, enquanto a linha 2 exibe as informações mantidas de cada par de *logs*. Por exemplo, no primeiro registro, observamos uma autenticação bem-sucedida, que inclui os dados de login do usuário [user@unifesspa.edu.br]. Esses dados também se repetem na linha 2, evidenciando a consistência das informações mantidas. Dessa forma, concluímos que a informação sobre o túnel TLS utilizado na conexão não é relevante para os objetivos desta pesquisa, uma vez que essa informação se repete de maneira idêntica para cada par de logs de autenticações. Portanto, a repetição dos dados relacionados ao TLS não contribui para a compreensão dos padrões investigados nesta pesquisa.

No segundo registro de Dados 2, as informações mantidas incluem detalhes técnicos adicionais, como a data, o horário, o ID da autenticação, o status da conexão, a rede *Wi-Fi* utilizada, o login do usuário, a porta do cliente *Wi-Fi* e o endereço *Media Access Control (MAC)* do dispositivo. Esses dados são valiosos para a análise de rede e segurança, pois permitem uma avaliação mais aprofundada do comportamento dos usuários e facilitam a identificação de padrões que podem indicar potenciais vulnerabilidades ou anomalias na rede.

Dados 2 - Registros de *logs* com presença do protocolo TLS

1- Tue May 16 00:03:12 2023 : Auth: (14053907) Login OK: [user@unifesspa.edu.br] (from client wifi port 0 via TLS tunnel)
2- Tue May 16 00:03:12 2023 : Auth: (14053908) Login OK: [user@unifesspa.edu.br] (from client wifi port 1 cli 56-39-63-1A-68-88)

Posteriormente, é aplicada uma expressão regular (REGEX) para estruturar os dados de acordo com as informações disponíveis e que foram classificadas como fundamentais juntamente com a equipe de infraestrutura para serem submetidos à análise. Técnicas baseadas no processamento de *logs* utilizando REGEX são amplamente utilizadas para buscar padrões em texto [9, 10]. No contexto da análise de *logs*, a aplicação de expressões regulares possibilita a identificação de padrões, a extração de dados relevantes e a estruturação dos registros de forma a facilitar a análise e o entendimento das informações registradas. Essa abordagem é fundamental para a organização e interpretação eficaz dos dados de *log*, contribuindo para a geração de resultados precisos e a detecção de tendências ou eventos importantes. A seguir é apresentada no Código 2 a expressão regular utilizada para estruturar os dados.

Código 2- Expressão regular

```
log_pattern = re.compile(r'(\w{3}) (\w{3}) +(\d{1,2}) (\d{2}:\d{2}:\d{2}) (\d{4}):(\b(?:Auth|ERROR|Info)\b): \((\w\d+)\) (Login \w+|OK|Invalid user|incorrect|radutmp|entry.*?)(?: \((.*?)\))?: \[([\w.@-]+)\]?.*\(from client ([\w-]+).*?cli ([\dA-Fa-f:-]+)\)')
dados_df[['Dia da Semana', 'Mês', 'Dia', 'Horário', 'Ano', 'Autenticação','ID', 'Status', 'Error', 'Email', 'Client', 'MAC Address']] = dados_df['Logs'].str.extract(log_pattern)
```

Com a aplicação da expressão regular apresentada, os dados foram tabulados em um único arquivo, seguindo a estrutura definida pelas expressões regulares. Isso permite a organização e padronização dos dados de acordo com os padrões identificados nos *logs*, facilitando a posterior análise e extração de informações relevantes. Nesta expressão, para cada linha de *log*, os dados são processados, extraídas as informações necessárias e organizadas em colunas com os seguintes dados: dia da semana, mês, dia, horário, ano, autenticação, id, *status*, erro, e-mail, cliente e endereço MAC.

Essa estruturação dos dados permite uma análise mais organizada e facilita a identificação de padrões e tendências nos registros de *logs*. Através dessas colunas, é possível filtrar, agrupar e visualizar os dados de acordo com as necessidades e objetivos da análise. Isso contribui para uma melhor compreensão do ambiente de rede e auxilia na tomada de decisões mais informadas em relação a melhorias, otimizações e solução de problemas.

3. RESULTADOS E DISCUSSÃO

3.1. Análise Estatística

Após a aplicação do REGEX, os dados tabulares são submetidos a análises, onde são aplicados os filtros determinados em conjunto com a equipe de infraestrutura. Esses filtros podem incluir critérios específicos para identificar e isolar dados relevantes, como padrões de comportamento, eventos específicos ou anomalias que requerem atenção especial. A análise dos dados com a aplicação de filtros e análise estatística personalizada permite a extração de dados significativos e a identificação de informações cruciais para a tomada de decisões e aprimoramento da infraestrutura de rede. A seguir apresentamos algumas possibilidades de análise que é viabilizada com a utilização do software proposto.

Primeiramente, nesta etapa, foi realizada uma análise estatística dos dados registrados processados, conforme demonstrado na Figura 2. Foram identificados três status de classificação dos *logs*, sendo eles:

***Login Ok*:** Situação em que a autenticação com a rede ocorreu corretamente.

***Login Incorrect*:** Situação em que a autenticação não foi concluída por algum erro.
***Invalid User*:** Situação em que a autenticação na rede não ocorreu devido à inserção errônea das informações de login.

No primeiro gráfico da Figura 2, podemos observar que 89,9% das requisições de *login* ocorrem de forma assertiva. Já os erros de autenticação correspondem, juntos, a 10,1% do total de requisições. Isso equivale a cerca de cinquenta e seis mil duzentos e setenta e nove registros.

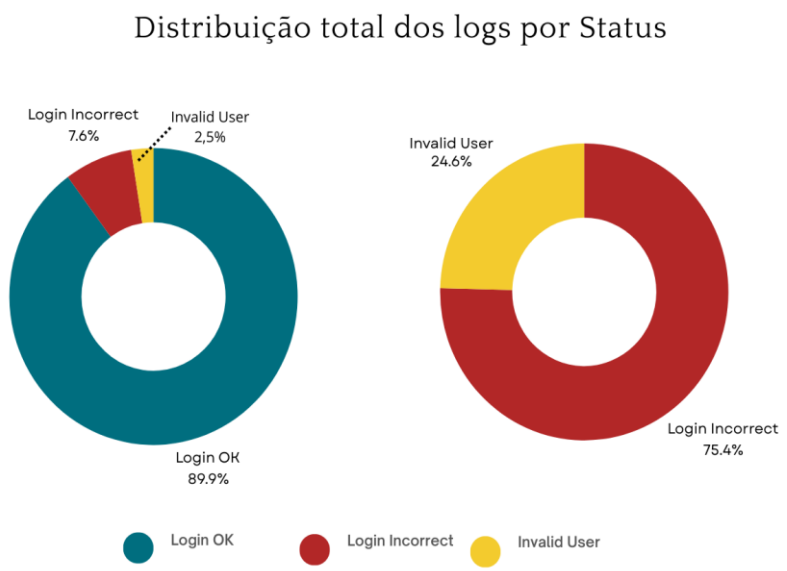

Figura 2. Distribuição total dos logs por Status

No segundo gráfico, representado na Figura 2, pode-se observar os percentuais de erros de usuário inválido e *login* incorreto. Podemos observar que 75,4% dos erros estão relacionados a situações em que a autenticação não foi concluída devido a algum erro de configuração dos protocolos de rede ou à inserção de caracteres como espaços ou arrobas duplicadas, enquanto 24,6% correspondem à situação em que a autenticação na rede não ocorreu devido à inserção incorreta das informações de *login*.

A Figura 3 ilustra a relação entre o uso das redes das Unidades I, II e III dos campi de Marabá e o número de erros de autenticação identificados em cada uma dessas redes. Ao analisarmos essa relação, observamos que a Unidade II apresenta um volume significativamente maior de autenticações durante o período de coleta de dados, além de registrar o maior número de falhas de autenticação.

Esse aumento no número de autenticações na Unidade II, em comparação com os outros campi, pode ser atribuído a vários fatores. Um deles é o maior fluxo de dispositivos conectados à rede, decorrente da realização de atividades acadêmicas que atraem um grande número de discentes e servidores durante o período em questão. Além disso, a análise da Figura 3 revela que a Unidade II também apresenta a maior concentração de uso da rede sem fio no período analisado, assim como o maior percentual de erros de *login*. Essas observações são cruciais para investigar se existem questões específicas de infraestrutura ou configuração que possam estar contribuindo para esse cenário. A compreensão desses fatores possibilitará a implementação de medidas corretivas direcionadas, visando melhorar a experiência de autenticação nos campi e, consequentemente, otimizar a conectividade e o desempenho da rede.

Após análise destes dados são realizadas consultas, essa etapa é crucial para extrair informações significativas que possam orientar a tomada de decisões e a implementação de melhorias na infraestrutura. A interpretação dos resultados obtidos permite compreender o impacto das ações, identificar tendências e padrões, e fornecer subsídios para aprimorar a eficiência e a segurança da infraestrutura de forma proativa.

A implementação apresentada no Código 3 apresenta a detecção do cenário onde são agrupadas as tentativas de *login* malsucedidas dos usuários e emitir um alerta para o administrador sempre que mais de dez tentativas consecutivas erradas são detectadas. Esse alerta contém informações detalhadas, incluindo o *login* do usuário, o tipo de erro identificado, o número total de requisições com erro, a data em que ocorreram, o intervalo de tempo entre as tentativas e a rede do campus onde os erros foram registrados. Essa abordagem visa proporcionar uma resposta rápida e eficaz a possíveis tentativas de acesso não autorizado. Além disso, esses alertas são essenciais para identificar problemas potenciais e garantir a segurança e a eficiência do sistema.

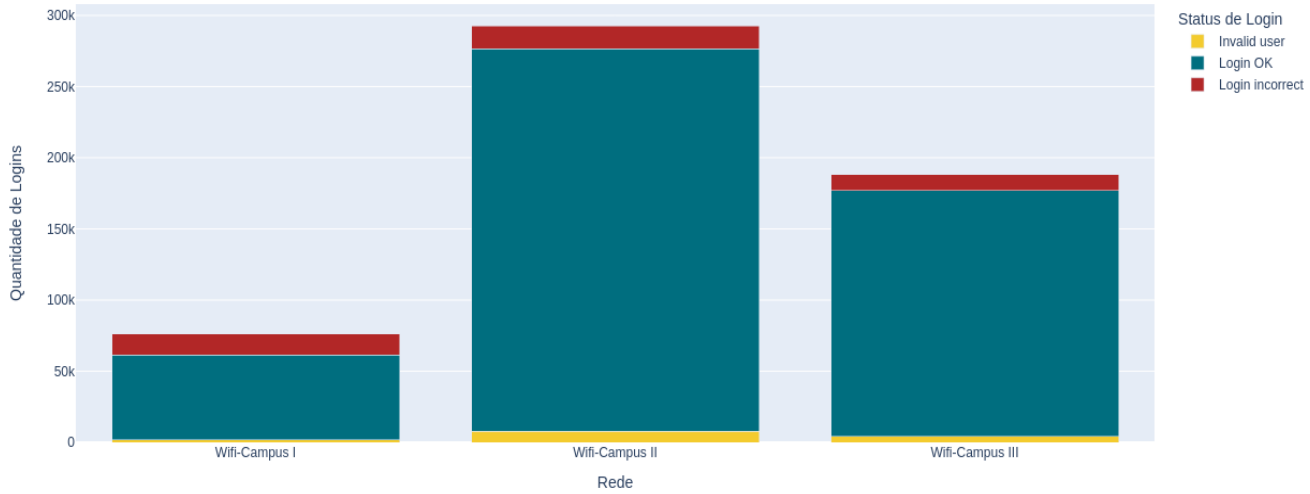

Figura 3. Quantidade de Logins por Rede X Status

Código 3 - Detecção de tentativas malsucedidas

```
tentativas_erro = dados_df[dados_df['Status'] == 'Login incorrect']
tentativas_erro_por_usuario_erro_data = tentativas_erro.groupby(['Email', 'Error', 'Data']).size()
usuarios_com_mais_de_10_erros=tentativas_erro_por_usuario_erro_data[tentativas_erro_por_usuario_erro_data > 10]
  if not usuarios_com_mais_de_10_erros.empty:
  for (usuario, error, data), tentativas in usuarios_com_mais_de_10_erros.items():
      print(f"Alerta para o administrador:")
      print(f"Usuário: {usuario}")
      print(f"Erro: {error}")
      print(f"Tentativas de login com erro: {tentativas}")
      print("Detalhes das tentativas:")
detalhes_tentativas = dados_df[(dados_df['Email'] == usuario) & (dados_df['Error'] == error) & (dados_df['Data'] == data)]
      print(f"Data: {data}")
      print(f"Intervalo de Horários:")
print(f"{detalhes_tentativas['Horário'].min()} - {detalhes_tentativas['Horário'].max()}")
      print(f"Rede: {detalhes_tentativas['Client'].unique()}")
      print("----------------------")
else:
  print("Não há usuários com mais de 10 tentativas de login com erro.")
```

Adicionalmente no Código 4, foi desenvolvido para verificar se um usuário possui mais de cinco dispositivos conectados simultaneamente. A implementação gera alertas para os administradores sempre que esses casos são identificados, incluindo as seguintes informações: *login* do usuário, data da ocorrência, quantidade de *logins* ativos durante aquele período, a rede em que a conexão foi detectada e os endereços *Media Access Control (MAC)* dos dispositivos. Essa abordagem visa garantir um controle mais eficaz sobre as conexões simultâneas, contribuindo para a segurança e a gestão da infraestrutura de rede.

Código 4 - Detecção de logins simultâneos

```
logins_por_usuario_data_mac = dados_df.groupby(['Email', 'Data'])['MAC Address'].nunique()
usuarios_com_mais_de_5_logins = logins_por_usuario_data_mac[logins_por_usuario_data_mac > 5]
if not usuarios_com_mais_de_5_logins.empty:
  qtd_usuarios = 0
  for usuario, data in usuarios_com_mais_de_5_logins.index:
 dados_usuario_data = dados_df[(dados_df['Email'] == usuario) & (dados_df['Data'] == data)]
mac_address_distintos = dados_usuario_data['MAC Address'].unique()
if len(mac_address_distintos) > 5:
client = dados_usuario_data['Client'].iloc[0]
  print(f"Alerta para o usuário {usuario}:")
  print(f"Data: {data}")
  print(f"Quantidade de logins ativos: {len(mac_address_distintos)}")
 print(f"MAC Address dos dispositivos conectados simultaneamente: {mac_address_distintos}")
```

```
    print(f"Client: {client}")
    print("----------------------")
 qtd_usuarios += 1
print(f"Total de usuários com mais de 5 logins em MAC Address distintos: {qtd_usuarios}")
else:
   print("Não há usuários com mais de 5 logins em MAC Address distintos no mesmo dia.")
```

Com essas medidas, busca-se fortalecer a proteção da rede e minimizar os riscos associados a acessos indevidos, garantindo uma melhor experiência para todos os usuários. A implementação desses alertas é um passo importante na nossa estratégia de segurança, permitindo uma resposta rápida e eficaz a possíveis ameaças.

3.2. Detecção Automática de Eventos

A aplicação dos códigos desenvolvidos nesta pesquisa ao conjunto de dados coletados demonstra um avanço significativo na segurança da infraestrutura monitorada. Esse progresso foi alcançado por meio da implementação e aplicação desses códigos, que tiveram como objetivo identificar eventos indicativos de tentativas repetidas de *login* em um curto espaço de tempo, o uso de múltiplos dispositivos conectados e a geração de alertas. Através da análise detalhada dos dados, foi possível detectar padrões de comportamento que indicam possíveis tentativas de acesso não autorizado, vazamento de credenciais e falhas de equipamentos.

Com a aplicação do Código 3 sobre os dados, foi possível identificar uma quantidade significativa de erros de *login* por usuário. Os dados analisados são convertidos em informações de alerta e apresentados à equipe de segurança. Esses alertas fornecem informações cruciais para a gestão e o monitoramento da infraestrutura, sendo essenciais para identificar problemas potenciais e garantir a segurança e a eficiência do sistema. A seguir no Alerta 1, é apresentado alguns dos trezentos e noventa e sete resultados retornados em formato de alerta dos dados coletados, proporcionando à equipe acesso simplificado às informações do usuário, ao erro gerado em suas tentativas frustradas de conexão, à quantidade de tentativas com erro, à data em que ocorreram as tentativas, ao intervalo de tempo e à rede onde o incidente aconteceu.

Alerta 1 - Alerta de detecção de tentativas malsucedidas

```
Alerta para o administrador:
Usuário: user@unfesspa.edu.br
Erro: eap_peap: The users session was previously rejected: returning reject (again.)
Tentativas de login com erro: 14
Detalhes das tentativas:
Data: 2023-05-16 00:00:00
Intervalo de Horários:
08:48:08 - 16:22:35
Rede: ['wifi']
—----------------------------------------------------------------------
Alerta para o administrador:
Usuário: user@unifesspa.edu.br
Erro: eap: No mutually acceptable types found
Tentativas de login com erro: 32
Detalhes das tentativas:
Data: 2023-05-02 00:00:00
Intervalo de Horários:
12:18:13 - 18:14:51
Rede: ['wifi-Campus3']
```

De posse dessas informações, o administrador de rede pode tomar algumas medidas para identificar problemas na rede e erros dos usuários durante a conexão, dentre elas podemos citar:

- **Monitoramento de erros de *login*** - Esses alertas fornecem dados cruciais, como a quantidade de tentativas com erro e os usuários afetados, permitindo que o administrador identifique padrões ou comportamentos suspeitos que possam indicar tentativas de acesso não autorizadas.
- **Análise de dados de conexão** - A partir das informações sobre as tentativas de conexão, o administrador pode investigar quais redes estão apresentando mais problemas e se há um número excessivo de dispositivos conectados simultaneamente e a quantidade de erros de conexão. Isso pode ajudar a identificar congestionamentos ou falhas na infraestrutura.
- **Relatórios e *feedback*** - Permitir que o administrador mantenha uma comunicação constante com a equipe de segurança e os usuários, coletando feedback sobre problemas enfrentados e ajustando as estratégias de monitoramento e resposta conforme necessário.

Com a aplicação do Código 04 sobre os dados, foi possível identificar um número significativo de usuários conectados simultaneamente a mais de cinco dispositivos. O alerta emitido ao administrador da rede fornece informações detalhadas sobre o usuário, a data da ocorrência, a quantidade de dispositivos autenticados, os endereços MAC dos dispositivos, a rede *Wi-Fi* utilizada e o intervalo de tempo em que as autenticações ocorreram. A seguir, é apresentado alguns dos duzentos e cinco alertas de usuários com mais de cinco *logins* simultâneos.

Alerta 2 - Alerta de detecção de logins simultâneos

Alerta para o usuário user@unifesspa.edu.br:
Data: 2023-05-12 00:00:00
Quantidade de logins ativos: 6
MAC Address dos dispositivos conectados simultaneamente: ['EA-E7-C0-2D-C0-A9' 'F6-4B-37-D3-FE-52' '4A-B1-78-4E-AE-50' '76-AE-B4-14-96-26' '1A-7B-B2-63-5F-49' '80-86-F2-F1-82-B8']
Client: wifi-c1
Intervalo de Horários:
08:48:08 - 16:22:35
Rede: ['wifi']
—----------
Alerta para o usuário user@unifesspa.edu.br:
Data: 2023-05-16 00:00:00
Quantidade de logins ativos: 7
MAC Address dos dispositivos conectados simultaneamente: ['80-86-F2-F1-82-B8' 'F6-4B-37-D3-FE-52' '62-7E-78-FB-D8-4E' '1A-7B-B2-63-5F-49' 'EA-E7-C0-2D-C0-A9' '4A-B1-78-4E-AE-50' '76-AE-B4-14-96-26']
Client: wifi-c3

Este alerta é fundamental para a área de segurança da informação, pois, com essas informações, o perito pode tirar conclusões essenciais para identificar potenciais violações de segurança na rede, especialmente em relação aos dados deste usuário. O fato de o usuário estar autenticado em vários dispositivos simultaneamente indica algumas situações de segurança preocupantes:

- **Vazamento de senha:** A autenticação do usuário em múltiplos dispositivos é um forte indicativo de que suas credenciais podem ter sido comprometidas.
- **Comprometimento de acesso a serviços institucionais:** Isso torna vulneráveis todos os acessos a outros sistemas institucionais, como o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), o Sistema Integrado de Patrimônio, Administração e Contratos (SIPAC), e-mails, Active Directory, entre outros.
- **Comprometimento de acesso a serviços externos:** A possibilidade de acesso não autorizado a serviços externos também deve ser considerada, uma vez que credenciais comprometidas podem ser utilizadas para acessar informações sensíveis fora da rede institucional.
- **Implementação de Medidas Preventivas:** Com base nas análises, o administrador pode implementar medidas preventivas, como a configuração de limites para o número de dispositivos que um usuário pode conectar simultaneamente, bloqueio deste usuário, forçando-o a realizar uma redefinição de senha, garantindo assim a segurança e a eficiência do sistema. Além de reforçar as políticas de segurança para evitar acessos indevidos.

Após a obtenção destes resultados gerados, revelando padrões relevantes nos dados, proporcionando uma visão mais clara do comportamento dos usuários na rede. As informações coletadas através destes dados no cenário apresentado não apenas sugerem um possível vazamento de senha ou contabilizam falhas de autenticação, mas também levantam preocupações sobre a integridade da segurança da informação em toda a rede da instituição estudada. A identificação de tais padrões é fundamental para que o administrador possa tomar medidas proativas. Com essas informações em mãos, é possível implementar restrições adequadas e reforçar as políticas de segurança, garantindo a proteção dos acessos a sistemas institucionais críticos. Portanto, a análise detalhada dos dados não apenas auxilia na identificação de problemas, mas também serve como uma ferramenta essencial para a mitigação de riscos, manutenção da segurança na rede, elaboração de relatórios forenses, preservação de evidências e geração de relatórios em processos judiciais.

3.3. Protótipo de Interface

Com base nas análises estatísticas e na implementação de códigos de detecção automática, conforme detalhado nas seções 3.1 e 3.2, foi desenvolvido um protótipo da interface do sistema. Este protótipo integra o uso de uma API (Interface de Programação de Aplicativos). Este protótipo foi desenvolvido utilizando a biblioteca STREAMLIT com o objetivo de transformar os dados processados em informações claras e objetivas, facilitando a análise e a tomada de decisão pelos administradores e especialistas.

O design da interface foi cuidadosamente planejado para proporcionar uma experiência intuitiva e eficiente aos usuários. Alguns destaques incluem:

***Dashboard*:** Os gráficos e paineis de controle apresentam os dados de maneira clara e concisa, facilitando uma visualização rápida e uma compreensão precisa das informações.

Alertas: O sistema é capaz de gerar alertas, notificando os peritos sobre eventos suspeitos, permitindo uma tomada de decisão rápida e eficaz para mitigar um incidente de segurança da informação.

Dados: A funcionalidade de dados permite visualizar os registros de *logs* e aplicar filtros, proporcionando uma análise mais detalhada e personalizada das informações.

Pesquisa avançada: A interface oferece recursos de busca e filtragem, permitindo que os peritos localizem e analisem registros específicos de forma eficiente e precisa. Isso torna a investigação de incidentes mais ágil e assertiva.

A organização e integração dos componentes desenvolvidos na API com a interface da *dashboard* ofereceram uma experiência que favorece a usabilidade e a compreensão dos dados pelos peritos. Em um estudo apresentado por [16], é discutida uma metodologia para o desenvolvimento de um sistema de análise de *logs* visual que se destaca pela sua eficiência e abrangência, permitindo a análise e o processamento de logs de forma rápida e precisa. A metodologia empregada utiliza a linguagem *Python* para criar uma interface unificada que integra diferentes algoritmos de análise de *logs*, facilitando a interação e a interpretação dos dados.

Além disso, a centralização dos dados e a sua estruturação em uma aplicação com interface não apenas preserva a integridade das informações, mas também facilita as investigações, permitindo que analistas e peritos acessem rapidamente dados cruciais sobre atividades suspeitas e identifiquem padrões de comportamento, rastreamento de ações realizadas por usuários e detectar anomalias em tempo hábil de tratá-la. Com base nisso e no que foi apontado na seção 2.2, concluímos que a informação sobre o túnel TLS utilizado na conexão não é relevante para os objetivos desta pesquisa, uma vez que essa informação se repete de maneira idêntica para cada par de logs de autenticações. Portanto, a repetição dos dados relacionados ao TLS não contribui para a compreensão dos padrões investigados nesta pesquisa. Essa constatação permite que o foco da análise seja direcionado a aspectos mais significativos e diferenciadores, que realmente impactam a segurança e a eficiência das autenticações na rede. As Figuras 4, 5, 6 e 7 apresentam o protótipo da interface projetada com o intuito de aprimorar a análise de *logs* da rede sem fio da Instituição.

Com os dados apresentados na *dashboard*, conforme apresentado na Figura 4, o perito consegue visualizar de forma rápida e precisa o que está acontecendo na rede. A interface oferece informações cruciais, como a quantidade de erros identificados na data atual, a distribuição desses erros entre os clientes *Wi-Fi*, e os erros mais frequentes que foram detectados. Além disso, a dashboard destaca o horário de maior ocorrência dos erros, permitindo uma análise temporal que pode ser fundamental para a identificação de padrões e a tomada de decisões informadas.

Essa abordagem não apenas facilita a compreensão das informações, mas também capacita o perito a agir de maneira proativa, otimizando a resposta a incidentes e melhorando a eficiência operacional da rede. O protótipo desenvolvido foi projetado para atender a essas necessidades, garantindo que os dados sejam apresentados de forma intuitiva e acessível.

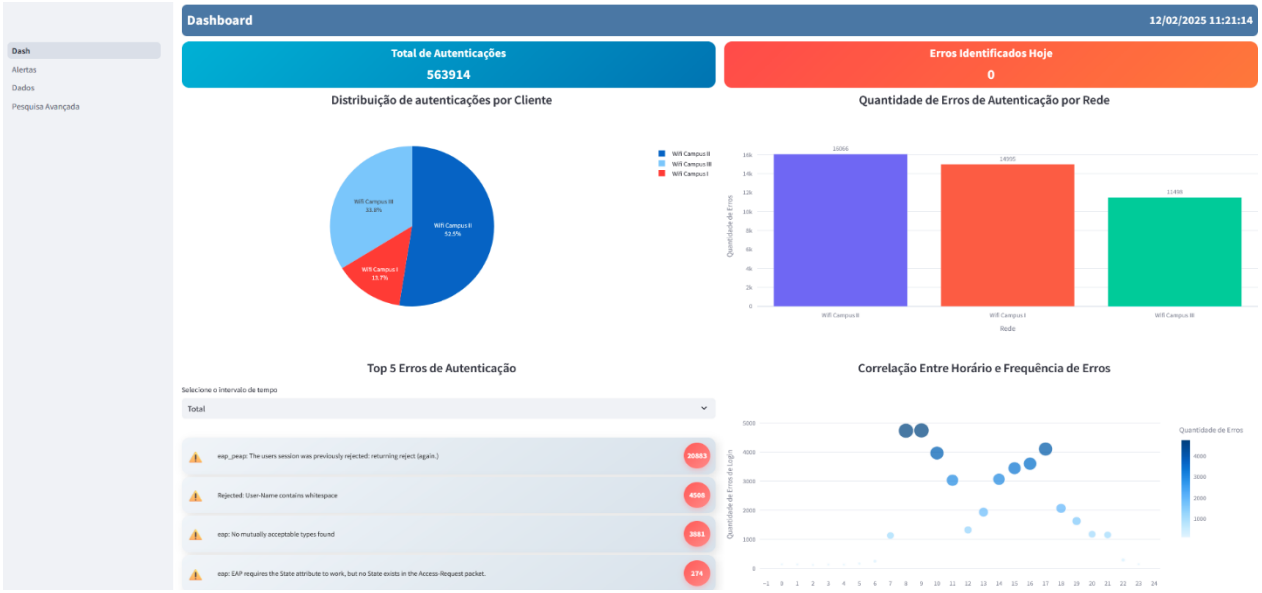

Figura 4. Interface do Sistema – Dashboard

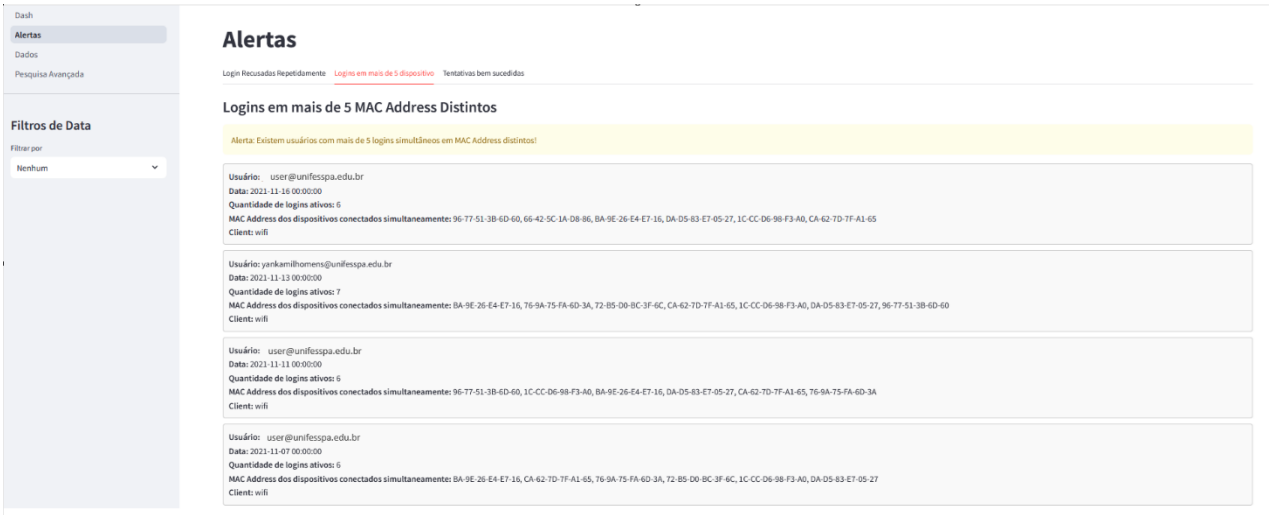

Figura 5. Interface do Sistema - Alertas

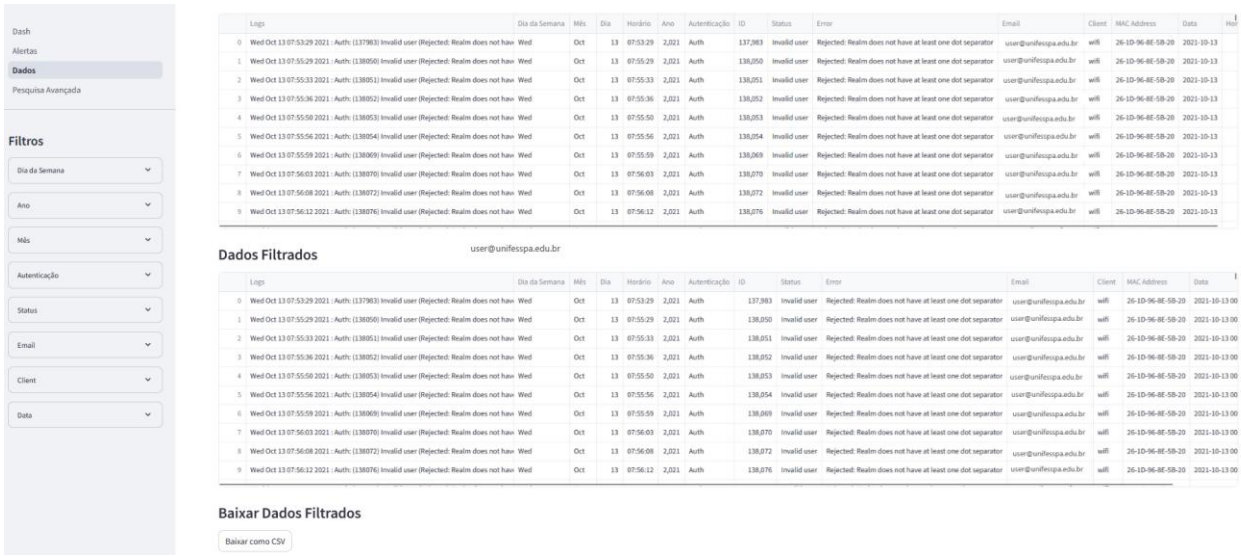

Figura 6. Interface do Sistema - Dados

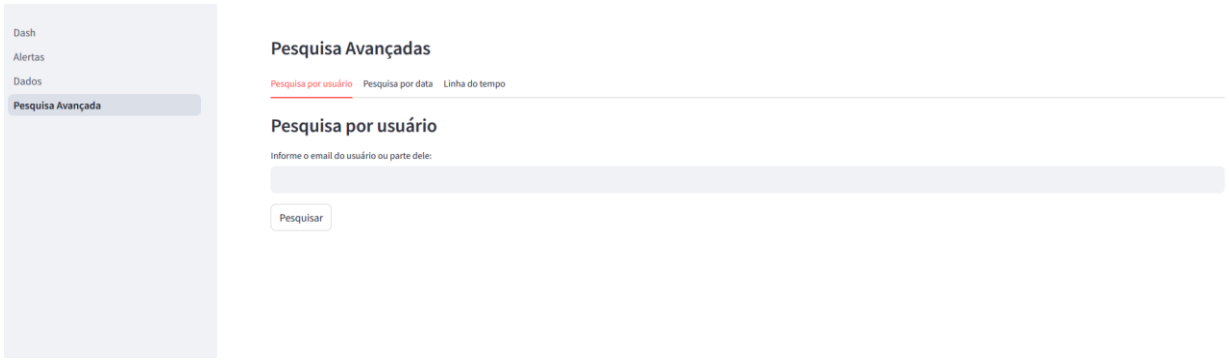

Figura 7. Interface do Sistema - Pesquisa Avançada

A Figura 05 apresenta a tela de alertas, onde são exibidos os alertas gerados a partir dos Códigos 03 e 04. Nesta interface, os dados são analisados e transformados em informações relevantes, organizadas em três abas distintas. Cada aba contém os registros identificados de diferentes tipos de dados, permitindo uma navegação intuitiva e uma visualização clara dos alertas. Essa estrutura facilita a compreensão dos eventos ocorridos, proporcionando ao administrador uma visão abrangente e detalhada das informações críticas relacionadas à segurança e ao desempenho da rede.

A Figura 06 apresenta os dados processados que estão sendo submetidos à análise. O objetivo desta tela é proporcionar ao analista uma visualização clara dos dados processados, permitindo a verificação de inconsistências. Além disso, a interface possibilita a aplicação de filtros adicionais que não estão programados na aplicação, oferecendo maior flexibilidade na análise e facilitando a identificação de padrões ou anomalias nos dados. Essa funcionalidade é essencial para garantir a precisão e a integridade das informações analisadas.

Por fim, a Figura 07 apresenta a tela de pesquisa avançada, que permite realizar buscas específicas de dados relacionados a um usuário. Nesta interface, o administrador pode visualizar os resultados de erros e alertas identificados, proporcionando uma análise detalhada das ocorrências. Além disso, a tela possibilita a realização de pesquisas em um lapso temporal, permitindo que o usuário gere uma linha do tempo com os dados de autenticação e os erros identificados. Essa funcionalidade é fundamental para rastrear eventos ao longo do tempo, facilitando a identificação de padrões e a análise de tendências, o que pode ser crucial para a resolução de problemas e a melhoria contínua da segurança e do desempenho da rede.

3.4 Validação dos Resultado

Para validar a aplicação, a ferramenta foi utilizada em um conjunto de logs coletados durante os meses de

outubro, novembro e dezembro de 2024. Esta metodologia não apenas possibilitou a verificação do desempenho e funcionalidade da aplicação, mas também facilitou a identificação de áreas potenciais para melhorias. A Figura 08 ilustra um conjunto de arquivos de logs que foram analisados.

A escolha desse conjunto de dados foi estratégica, visando garantir uma amostra representativa das autenticações realizadas na rede durante este período. A análise desses logs permitiu não apenas a verificação da eficácia da ferramenta proposta, mas também a identificação de padrões e anomalias que possam ser relevantes para a segurança da rede.

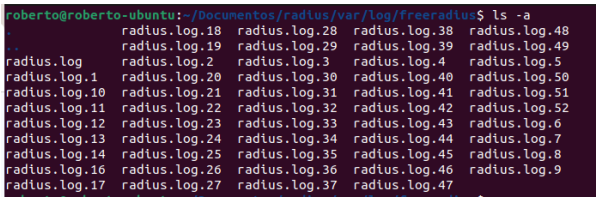

Figura 08 - Arquivos de logs

Após a coleta os dados foram submetidos a aplicação, onde avaliamos sua capacidade de processar informações de forma automática. Para validar a eficácia desse processamento, realizamos uma comparação detalhada por meio de análises estatísticas, antes e após a aplicação dos métodos de processamento. A Figura 09 apresenta a verificação da quantidade de registros contidos nos arquivos, totalizando seiscentos e cinquenta e seis mil setecentos e oitenta e um registros neste conjunto de dados.

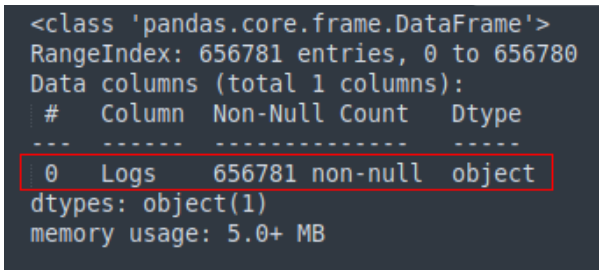

Figura 09 - Total de registros coletados

Em seguida, a aplicação realizou a limpeza dos registros que continham linhas com a descrição do protocolo TLS. Após esse processo de limpeza, conforme ilustrado na Figura 10, obtivemos um total de trezentos e seis mil seiscentos e vinte e oito registros.

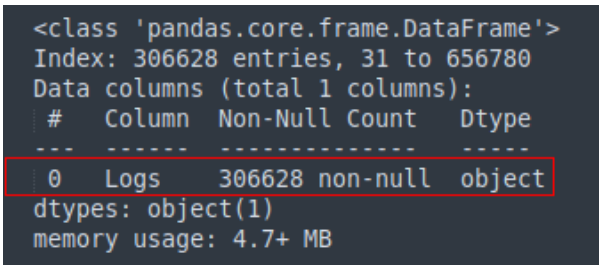

Figura 10 - Total de registros após eliminação dos dados TLS

Posteriormente, os dados foram submetidos a um processo de expressões regulares (REGEX) para a categorização e classificação das informações. Nesta etapa, conforme descrito na Seção 2.2, os dados foram estruturados em campos específicos, como Dia da Semana, Mês, Dia, Horário, Ano, Autenticação, ID Status, Error, Email, Client e MAC Address. Essa estruturação foi essencial para organizar os registros e facilitar a análise subsequente.

Na Figura 11, é possível observar o total de registros apresentados anteriormente, agora processados e contabilizados corretamente, com destaque em vermelho. No entanto, uma análise mais detalhada revelou que os logs das colunas destacadas em verde não correspondem exatamente ao total de registros processados. Foi identificada uma diferença de dois mil quinhentos e oito registros a menos, o que equivale a apenas 0,82% dos dados submetidos à aplicação. Esse resultado demonstra que a aplicação conseguiu estruturar com sucesso 99,18% dos registros de logs, evidenciando sua eficiência no processamento.

Outra observação relevante, destacada na cor amarela, refere-se ao total de quinze mil duzentos e oitenta e dois erros contabilizados. Esse número é distinto dos demais registros, pois considera exclusivamente os logs que contêm erros documentados em sua estrutura.

A análise detalhada desses erros é crucial para identificar inconsistências nas etapas subsequentes e aprimorar o processo de tratamento dos dados. Compreender a natureza e a frequência dos erros permitirá implementar melhorias significativas, garantindo a qualidade e a confiabilidade das informações processadas.

```
<class 'pandas.core.frame.DataFrame'>
Index: 306628 entries, 31 to 656780
Data columns (total 13 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Logs            306628 non-null  object
 1   Dia da Semana   304120 non-null  object
 2   Mês             304120 non-null  object
 3   Dia             304120 non-null  object
 4   Horário         304120 non-null  object
 5   Ano             304120 non-null  object
 6   Autenticação    304120 non-null  object
 7   ID              304120 non-null  object
 8   Status          304120 non-null  object
 9   Error           15282 non-null   object
 10  Email           304120 non-null  object
 11  Client          304120 non-null  object
 12  MAC Address     304120 non-null  object
dtypes: object(13)
memory usage: 32.8+ MB
```

Figura 11 - Dados estruturados com Regex

Após o processamento e a estruturação dos dados, foi possível visualizar, nos gráficos do *dashboard*, as informações que confirmam a correta leitura dos dados. Conforme apresentado na Figura 12, foram registrados um total de trezentos e quatro mil, cento e vinte registros. Essa quantidade significativa de registros, como demonstrado anteriormente na Figura 11, permite uma análise mais robusta e detalhada.

Além disso, na Figura 12 é possível observar a distribuição das autenticações entre os campi, onde 61% ocorrem no campus II, 30,9% no campus III e 8,11% no campus I. Essa análise revela uma concentração significativa das autenticações no campus II, indicando um padrão de uso que pode ser relevante para futuras estratégias de segurança e gestão de recursos.

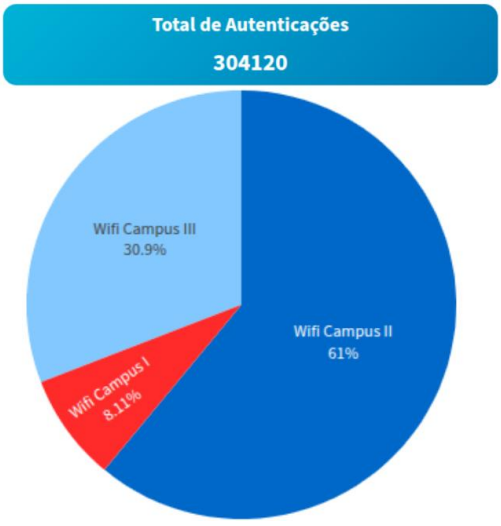

Figura 12 - Total de registros

Na Figura 13, é apresentada a distribuição dos erros de autenticação por campus, onde podemos observar que o total de dados é equivalente ao que foi apresentado na Figura 11. Além disso, nota-se que o campus II, que já possui a maior quantidade de autenticações, também apresenta a maior concentração de erros durante o período analisado.

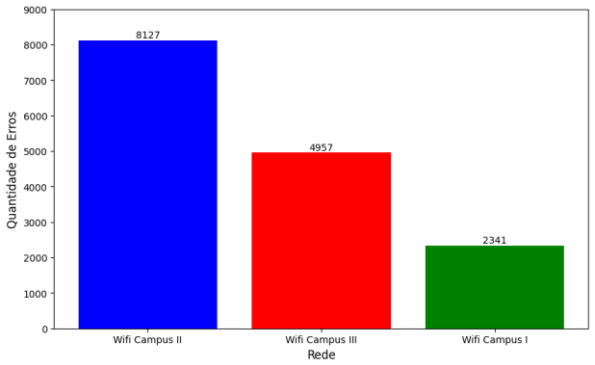

Figura 13 - Quantidade de erros de autenticação por rede

Na Figura 14 do *dashboard*, é apresentado o agrupamento e a quantidade dos cinco principais erros identificados pela aplicação, listados de acordo com sua frequência de ocorrência. Essa visualização permite uma análise clara e objetiva dos erros mais comuns, facilitando a identificação de áreas que necessitam de atenção e melhorias. A compreensão desses dados é fundamental para otimizar o sistema de autenticação e reduzir a incidência de falhas, contribuindo assim para uma experiência mais segura e eficiente para os usuários.

Figura 14 - Cinco principais erros

Na Figura 15, apresentamos a totalização dos dados de alertas gerados pela aplicação. As barras em vermelho representam o número de tentativas de login recusadas que ocorreram repetidamente em um curto intervalo de tempo. A aplicação está configurada para gerar alertas quando um erro se repete dez vezes consecutivas, registrando informações sobre o usuário, o tipo de erro e o intervalo em que ocorreu, a data e a rede. Para os dados analisados, foram registradas um total de 185 ocorrências, que estão destacadas em vermelho na Figura 15.

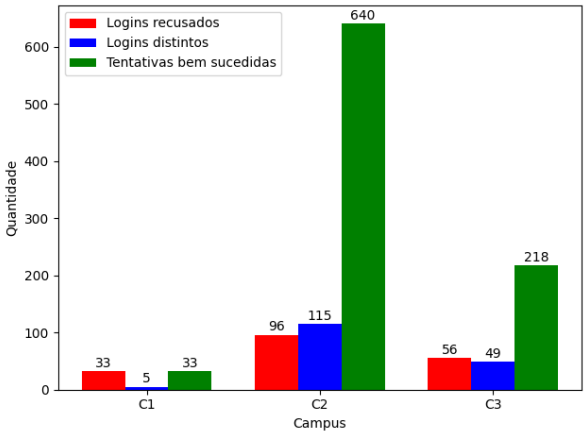

Figura 15 - Contabilização de alertas por campus

Na Figura 15, estão apresentados em azul os alertas identificados por campus, referentes ao total de logins realizados em mais de cinco dispositivos dentro de um intervalo de uma hora. Conforme ilustrado na figura, foram identificados 169 alertas.

Ainda na Figura 15, as barras em verde representam o quantitativo de logins que obtiveram sucesso consecutivo em mais de cinco dispositivos. Esse dado é relevante porque indica um comportamento potencialmente anômalo ou suspeito, já que o acesso simultâneo a múltiplos dispositivos pode estar associado a práticas como compartilhamento de credenciais ou uso indevido de contas. A análise desse tipo de ocorrência é essencial para identificar padrões de comportamento que possam comprometer a segurança da aplicação ou violar políticas de uso. Assim, a inclusão desse indicador no gráfico reforça a importância de monitorar e registrar atividades que fogem ao padrão esperado.

Além disso, podemos constatar que o Campus II apresenta, em todos os casos analisados, um número significativamente maior de alertas em comparação aos Campi I e III. Esse comportamento pode ser atribuído a diversos fatores, sendo o principal deles o maior fluxo de dispositivos conectados à rede. Esse fluxo elevado decorre da realização de atividades acadêmicas que atraem um grande número de discentes e servidores, especialmente durante períodos de maior movimentação, como semanas de provas, eventos acadêmicos e aulas práticas.

A alta densidade de usuários conectados simultaneamente aumenta a probabilidade de ocorrências relacionadas à segurança, como tentativas de login malsucedidas, acessos simultâneos em múltiplos dispositivos e outros comportamentos que geram alertas. Esse cenário reforça a necessidade de monitoramento contínuo e de políticas de segurança robustas para mitigar riscos e garantir a integridade da rede, especialmente em locais com maior concentração de usuários.

4. CONCLUSÃO

A presente pesquisa demonstra a importância dos dados gerados através de *logs* de aplicações e especificamente os dados de autenticação de redes sem fio, através deste registro e a implementação de códigos para realizar análise automática e gerar informações por meio de alerta a administradores de rede e segurança da informação. A aplicação se mostrou eficiente, e mostrou o quão esses dados podem ser valiosos e úteis se tratados da forma correta dentro de uma organização. Como vimos, as informações obtidas através destes registros podem proporcionar uma série de vantagens aos administradores e tornando mais fácil a implementação de medidas para sanar os incidentes e aperfeiçoar a segurança da rede e dos dados dos usuários.

Ademais, foi apresentado que a centralização da análise de *logs* pode ajudar na construção de um histórico minucioso das atividades do sistema, o que é fundamental para a elaboração de relatórios forenses e para a apresentação de evidências em processos judiciais. Assim, a implementação de uma ferramenta eficiente para análise de *logs* não apenas reforça a segurança da informação, mas também melhora a capacidade dos investigadores em lidar com incidentes de segurança, elevando a proteção contra ameaças cibernéticas e assegurando a integridade do ambiente digital, promovendo um espaço virtual mais seguro e confiável para todos os usuários.

Com isso, este trabalho traz sua contribuição significativa para o avanço do conhecimento na área de análise forense de dados de *logs* de autenticação em rede sem fio. As principais contribuições incluem: melhoria no gerenciamento de *logs*, facilitando a coleta, normalização, análise, monitoramento e configuração de alertas para identificar eventos anômalos; exploração forense da análise de logs, permitindo detectar e investigar incidentes de segurança cibernética; apoio à tomada de decisões estratégicas, fornecendo informações cruciais para segurança da rede e otimização de processos; e aumento da eficiência operacional, ao identificar gargalos e oportunidades de aprimoramento através da interpretação dos dados dos *logs*.

5. PROPOSTA FUTURA

Como proposta futuras, pretende-se aprimorar a aplicação, focando na estruturação de logs por meio da implementação de técnicas mais avançadas em

comparação ao uso de expressões regulares (REGEX). Embora as expressões regulares tenham se mostrado eficazes em várias situações, sua utilização pode apresentar algumas dificuldades, como a complexidade na escrita e na manutenção de padrões, além da possibilidade de erros que podem surgir em estruturas de logs mais complexas.

Além disso, pretende-se integra a esta aplicação uma *Large Language Model* (LLM) que será treinada dentro do contexto de inteligência artificial para operações de tecnologia da informação (AIOps) para permite que aos administradores e peritos identificar padrões, tendências e anomalias nos dados coletados não identificados nas consultas programadas.

AGRADECIMENTOS

Os autores agradecem à Unifesspa/CTIC, ao PPGCF/IGE, à CAPES, ao CNPq e à FAPESPA pelo apoio institucional e financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] S. He, P. He, Z. Chen, T. Yang, Y. Su, and M. R. Lyu, “A survey on automated log analysis for reliability engineering,” ACM Comput. Surv., vol. 54, no. 6, jul 2021.URL: https://doi.org/10.1145/3460345.

[2] S. de Governo Digital, “Modelo de política de gestão de registros (logs) de auditoria,” 2023. Acessado em 16 de fevereiro de 2024. URL: www.gov.br/governodigital/pt-br/privacidade_e_seguranca/ppsi/modelo_politica_logs_auditoria.pdf.

[3] S. He, P. He, Z. Chen, T. Yang, Y. Su, and M. R. Lyu, “A survey on automated log analysis for reliability engineering,” ACM Comput. Surv., vol. 54, no. 6, jul 2021. URL: https://doi.org/10.1145/3460345.

[4] B. Debnath, M. Solaimani, M. Gulzar, N. Arora, C. Lumezanu, J. Xu, B. Zong, H. Zhang, G. Jiang, and L. Khan, “Loglens: A real-time log analysis system,” 07 2018.

[5] A. B. de Normas Técnicas, “Abnt nbr iso/iec 27001:2013. Acessado em 16 de fevereiro de 2024. URL: https://bit.ly/ ABNTNBRISOIEC270012013.

[6] J. Svacina, J. Raffety, C. Woodahl, B. Stone, T. Cerny, M. Bures, D. Shin, K. Frajtak, and P. Tisnovsky, “On vulnerability and security log analysis: A systematic literature review on recent trends,” in Proceedings of the International Conference on Research in Adaptive and Convergent Systems, ser. RACS ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 175–180.URL: https://doi.org/10.1145/3400286.3418261.

[7] S. Alspaugh, B. Chen, J. Lin, A. Ganapathi, M. Hearst, and R. Katz, “Analyzing log analysis: An empirical study of user log mining,” in 28th Large Installation System Administration Conference (LISA14). Seattle, WA: USENIX Association, Nov. 2014, pp. 62–77. URL: https://www.usenix.org/conference/lisa14/conference-program/presentation/alspaugh.

[8] D.-Q. Zou, H. Qin, and H. Jin, “UiLog: Improving log-based fault diagnosis by log analysis,” J. Comput. Sci. Technol., vol. 31, no. 5, pp. 1038–1052, Sep. 2016.

[9] P. Dusane and G. Sujatha, “Logea: Log extraction and analysis tool to support forensic investigation of linux-based system,” in 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 909–916.

[10] J. Zhu, S. He, J. Liu, P. He, Q. Xie, Z. Zheng, and M. R. Lyu, “Tools and benchmarks for automated log parsing,” in 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), 2019, pp. 121–130.

[11] M. Fält, S. Forsström, Q. He, and T. Zhang, “Learning-based anomaly detection using log files with sequential relationships,” in 2022 6th International Conference on System Reliability and Safety (ICSRS), 2022, pp. 337–342.

[12] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, “Deep learning for anomaly detection in log data: A survey,” Machine Learning with Applications, vol. 12, p. 100470, 2023. URL: https://www.sciencedirect.com/science/article/pii/S2666827023000233.

[13] A. H. Shah, D. Pasha, E. H. Zadeh, and S. Konur, “Automated log analysis and anomaly detection using machine learning,” vol. 358, Virtual, Online, China, 2022, pp. 137 – 147, analysis detection;Anomaly detection;Clusterings;Labeled dataset;Labelings;Log;Log analysis;Logfile;Machine-learning.URL:http://dx.doi.org/10.3233/FAIA220378C

[14] RENAN CAVALHEIRO. Computação Forense: A Ciência da Solução de Crimes Digitais. Acessado em: 1 out. 2024. URL: https://academiadeforensedigital.com.br/computacao-forense-a-ciencia-da-solucao-de-crimes-digitais/

[15] ELEUTÉRIO, Pedro Monteiro da Silva; MACHADO, Marcio Pereira. Desvendando a Computação Forense. 1 ed. São Paulo: Novatec, v.7, 2011. 200 p. ISBN: 978-85-7522-260-7.

[16] Y. Wang, “Design of visual log analysis system,” in 2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE), 2023, pp. 1649–1652.