Resolução e classificação de alertas de centros de operações de segurança

Relatório de Aquisição de conhecimentos

2024 / 2025

1211131 Pedro Pereira

1200799 Rafael Branco

1240613 Ricardo Gonçalves

1222708 Ricardo Lemos

1240156 Svetlana Zamyatina

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Índice

Índice de Figuras

Índice de Tabelas

Não foi encontrada nenhuma entrada do índice de ilustrações.

1 Introdução / Objetivos do Desafio

Os sistemas periciais (SP) são programas criados para imitar o pensamento de especialistas humanos, baseando-se do seu conhecimento em setores específicos [1]. Por meio de uma combinação entre uma base de conhecimento e um mecanismo de inferência, estes sistemas resolvem problemas complexos e tomam decisões informadas. As áreas de aplicação destes sistemas vão desde medicina à engenharia e finanças. Um sistema pericial é composto pela aquisição do conhecimento de um especialista e sua representação, desenvolvimento de um motor de inferência e implementação de uma interface amigável para o utilizador.

O desafio proposto para as unidades curriculares de Paradigmas de Programação em Inteligência Artificial (PPROGIA) e Engenharia do Conhecimento (ENGCIA) é o desenvolvimento de um SP, no contexto da área de cibersegurança.

Desta forma, tendo em conta o background dos membros do grupo optou-se pelo desenvolvimento de um sistema pericial para **resolução e classificação de alertas de centros de operações de segurança** (SOC). Os SOCs gerem uma grande quantidade de dados e eventos diários, concebendo uma vasta quantidade de alertas provenientes de diferentes ferramentas e sensores [2]. Estes alertas precisam de ser analisados e tratados de acordo com sua criticidade e relevância, para garantir que ameaças reais sejam identificadas e resolvidas a tempo.

O crescente volume de ciberataques e a complexidade das ameaças atuais impõem uma carga significativa sobre os analistas de SOC, que muitas vezes precisam de lidar com grandes quantidades de alertas falsos positivos, além duma ampla gama de cenários de ataques que exigem conhecimento especializado. Isso torna os SP uma solução promissora, pois podem replicar o conhecimento de especialistas em cibersegurança e automatizar parte do processo de triagem e classificação dos alertas [3], [4].

O desenvolvimento deste sistema irá auxiliar os peritos e novatos no processo de diagnóstico melhorando o tempo de resposta a novos alertas. O sistema poderá apontar a causa do problema, sugerindo ações a tomar para a resolução do mesmo.

2 Fontes de Conhecimento

A construção de um sistema pericial de sucesso dependerá da quantidade e qualidade do conhecimento fornecido a ele. Nesta seção, vamos explorar as diferentes fontes de conhecimento usadas para alimentar o nosso SP. O grupo selecionou Paulo Valdeira, como perito para o projeto, analista de SOC na empresa *Redshift Global*. Para além disso o grupo teve acesso a certos documentos da empresa que descrevem os processos a tomar no caso da receção de um determinado tipo de alerta.

2.1 Paulo Valdeira

Paulo Valdeira é um atual SOC *Manager* da empresa *Redshift Global*. O perito exerceu funções na área de IT (*Information Tecnology*) há cerca de 21 anos, onde 10 desses anos são representados por atividades na área de SOC. O mesmo esteve cinco anos como analista SOC *tier 1*, três anos nos *tiers 2/3* e os dois últimos anos como SOC *Manager*.

Além da vasta experiência na área deste projeto, o perito apresenta as seguintes formações:

- *ETH Ethical Hacking*
- *CyberOPS*
- *Lead Implementer* ISMS – ISSO/IEC 27001:2022
- *Comptia Security+*

2.2 Documentos da empresa

Os documentos da empresa são recursos formais que descrevem procedimentos, políticas e orientações que devem ser seguidos pelos colaboradores no desempenho das suas funções. Esses documentos são fundamentais em várias áreas, pois garantem a consistência na execução de processos, promovem a aplicação de boas práticas e ajudam a mitigar riscos, especialmente em setores críticos como a segurança da informação.

Estes documentos apresentam instruções detalhadas sobre como realizar tarefas específicas. No caso da *Redshift Global*, esses manuais explicam as ações a serem tomadas quando um alerta de segurança é recebido, orientando os analistas de SOC sobre os passos corretos para lidar com incidentes. De modo a ser possível modelar o conhecimento, dividiu-se o mesmo por diferentes casos de uso, onde cada um representa um alerta diferente. Por

sua vez, cada caso de uso tem um diagrama e um conjunto de pontos associados que foram disponibilizados em formato de imagem. Deste modo, o grupo conseguiu guiar-se ao dividir as diferentes imagens pelos elementos, de forma ordenada, onde cada caso de uso desenvolvido foi discutido não só pelos membros do grupo, como com o perito Paulo Valdeira, que deu as suas indicações durante as reuniões semanais.

3 Sessões de aquisição de conhecimento

Foram realizadas duas sessões de aquisição de conhecimento com o perito, dia 1 de outubro e dia 10 de outubro de 2024. Segue uma breve descrição do conteúdo da reunião:

3.1 Sessão 1 – 01/10/2024

Na primeira sessão com o perito Paulo Valdeira, o grupo realizou uma breve apresentação das ideias iniciais para o projeto, com o objetivo de alinhar as expectativas e direções possíveis. O perito de seguida apresentou-se falando um pouco do seu histórico profissional e do trabalho que estava a desenvolver atualmente como analista de SOC, apresentando uma série de mais de 100 casos de uso que podem ser enfrentados no exercício da função de analista de SOC. Dada a amplitude desse leque de possibilidades, o grupo, em conjunto com o perito, decidiu concentrar-se em cinco casos de uso específicos. O perito forneceu uma breve explicação sobre cada um desses casos, acompanhada da documentação pertinente da empresa.

Posteriormente, cada caso de uso foi atribuído a um elemento do grupo e foi combinado que cada elemento do grupo ia elaborar durante a semana um diagrama *workflow*. O objetivo seria que, na próxima reunião, cada integrante apresentasse o seu caso ao perito para revisão e aperfeiçoamento, promovendo assim o refinamento do conhecimento exposto através da análise desses diagramas.

3.2 Sessão 2 – 10/10/2024

Esta sessão iniciou com uma breve apresentação do que iria ser tratado na mesma e uma pequena introdução do que fora feito. Cada um dos elementos havia desenhado um diagrama que representava o conhecimento associado aos diferentes alertas, portanto, os mesmos foram apresentados ao perito Paulo Valdeira, de modo que pudessem ser validados baseado no seu conhecimento.

Após a validação por parte do perito, foi possível realizar uma pequena demonstração do que seria a aplicação desenvolvida no modo pergunta-resposta, visto a integração entre os módulos de *frontend* e *backend-drools* já estar desenvolvida e com um pequeno exemplo apresentável.

Concluiu-se a sessão com os objetivos para a semana seguinte e com as melhorias não só nos diagramas, como no projeto de um modo geral.

4 Representação do conhecimento adquirido

4.1 Múltiplas falhas de Login para uma única conta

Baseado no diagrama apresentado na Figura 1 percebe-se que o conhecimento está modelado no formato de diagrama onde existem interações com o utilizador.

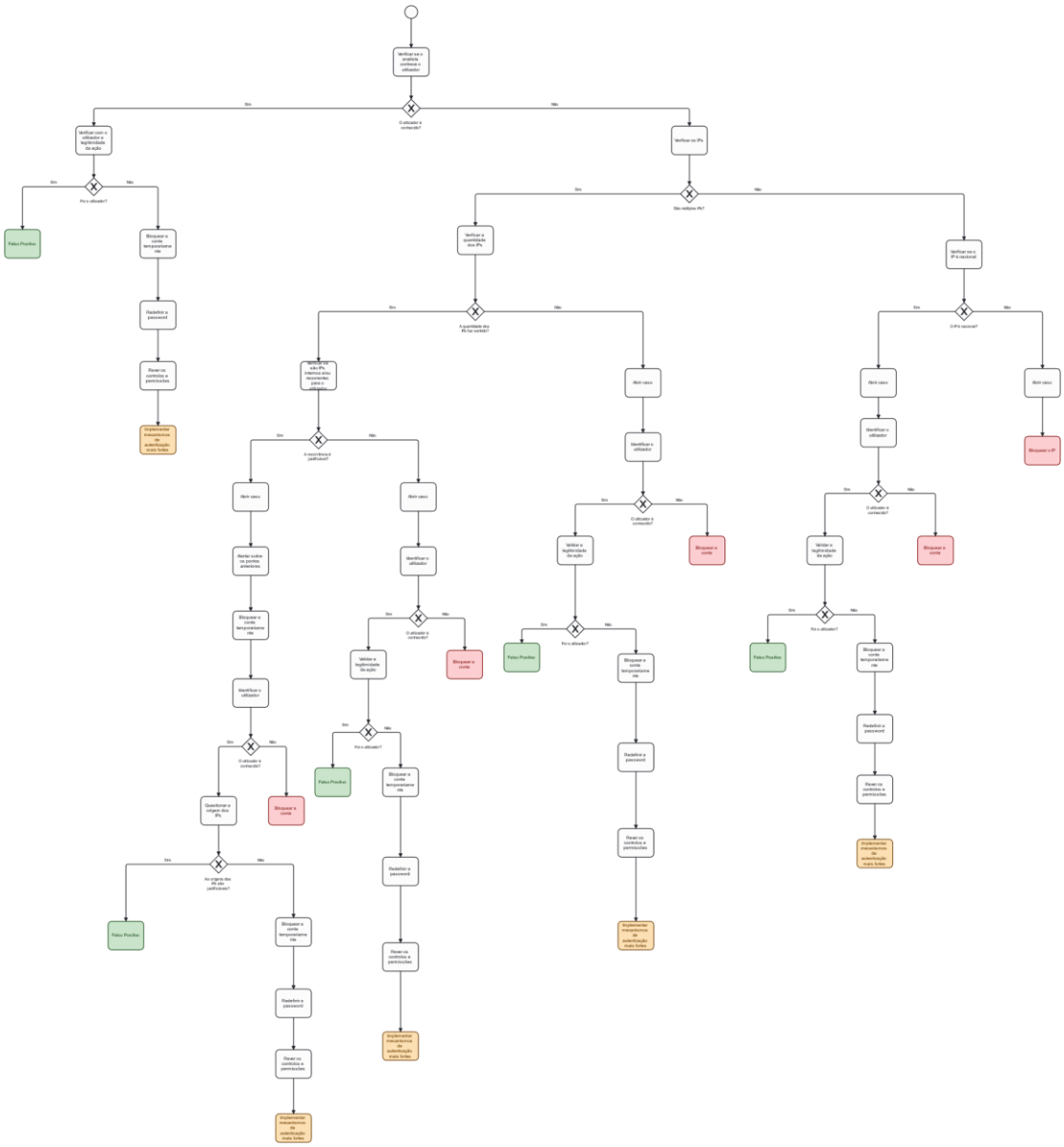

Figura 1 - Fluxograma múltiplos logins falhados para o mesmo utilizador

Este fluxograma está dividido num conjunto de questões diretas onde a interação é, maioritariamente, baseada em respostas “sim” ou “não”. O processo começa por validar se o utilizador é conhecido por parte do analista e caso aconteça, então, é necessário perguntar ao utilizador se foi realmente uma ação sua e, num caso positivo, assume-se um falso positivo,

caso contrário, procede-se a um conjunto de passos para controlar novamente o acesso à conta.

Caso o analista não conheça o utilizador, procede-se à triagem do alerta, a partir da deteção dos IPs e a sua quantidade, como também a nacionalidade e a recorrência. Tipicamente, estes casos são enviados para o *Helpddesk* onde, neste caso, está representado pelo texto "Abrir caso". Após a abertura do caso, o *Helpdesk* faz o mesmo processo de reconhecimento da pessoa e a resolução seria a mesma, caso contrário verifica as geopolíticas associadas, procede ao bloqueio do/dos IPs e realiza o processo de recuperação e controlo de conta. Este processo conta com o bloqueio temporário da conta, redefinição da *password*, revisão das permissões e a implementação de mecanismos de autenticação mais fortes.

4.2 Múltiplos inícios de sessão um utilizador em dispositivos diferentes

A Figura 2 é um fluxograma que representa um processo de avaliação de atividade de login simultâneo em várias localizações ou endereços IPs, para identificar se há comportamento suspeito ou malicioso em contas de utilizadores.

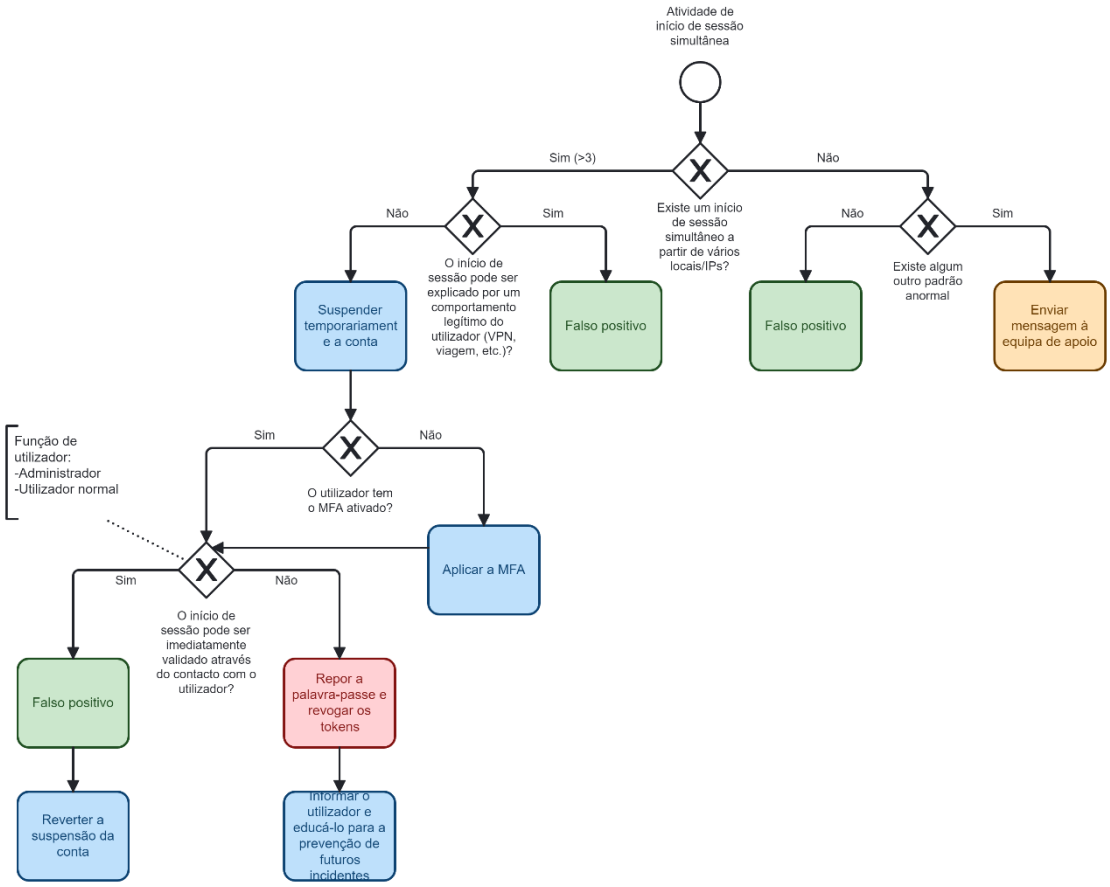

Figura 2 - Fluxograma múltiplos inícios de sessão um utilizador em dispositivos diferentes.

O fluxo começa com a identificação de atividades de login simultâneo por parte de um utilizador. Se houver mais de 3 logins simultâneos (em diferentes localizações ou IPs), o sistema considera que pode haver uma tentativa de acesso suspeita. O número >3 refere-se à quantidade de logins simultâneos que estão a acontecer a partir de diferentes locais ou endereços IPs. Caso não haja mais de 3 logins, segue-se para verificar outros padrões anormais de comportamento. Se houver um padrão anormal no comportamento do login, uma mensagem é enviada à equipa de suporte para investigar mais a fundo, caso contrário o caso é fechado como falso positivo.

Se houver mais de 3 logins simultâneos, de seguida verifica-se se estes podem ser explicados por comportamento legítimo (como VPN, viagens ou outro motivo válido), o resultado é falso positivo se o comportamento for verificado. Se não puder ser explicado, o próximo passo é suspender temporariamente a conta para evitar possíveis riscos e garantir que o utilizador tenha autenticação multifator ativada.

Por fim verifica-se se início de sessão pode ser validado através do contacto com o utilizador. Se o login puder ser validado ao entrar em contacto com o utilizador, é considerado

um falso positivo e a suspensão da conta é revertida. Se o login não puder ser validado, a palavra-passe do utilizador é reposta e os tokens de sessão são revogados. Além disso, o utilizador é informado e educado sobre como prevenir futuros incidentes de segurança.

4.3 Alterações efetuadas na firewall

No fluxograma seguinte (Figura 3) encontra-se representado o processo para verificar as alterações efetuadas numa firewall e se as mesmas podem ou não representar um perigo para o negócio. Com esta representação é possível saber quais os passos a seguir em determinadas situações e qual a melhor forma de intervir e de resolver o problema encontrado.

Assim que um novo alerta é lançado cabe ao analista SOC fazer a análise do mesmo, com o objetivo de encontrar algo que indique atividade suspeita. Se forem detetadas alterações na firewall é necessário perceber se estas foram feitas por uma fonte confiável e segura.

Para isso pode ser necessário proceder a análise de informação que possa tornar possível perceber se alteração foi efetuada de uma forma legitima ou com um objetivo malicioso.

Após a avaliação das alterações e do objetivo das mesmas é necessário proceder a ações de mitigação daquelas que podem ser as vulnerabilidades ou configurações incorretas que terão sido exploradas, de forma a restabelecer o normal funcionamento do sistema.

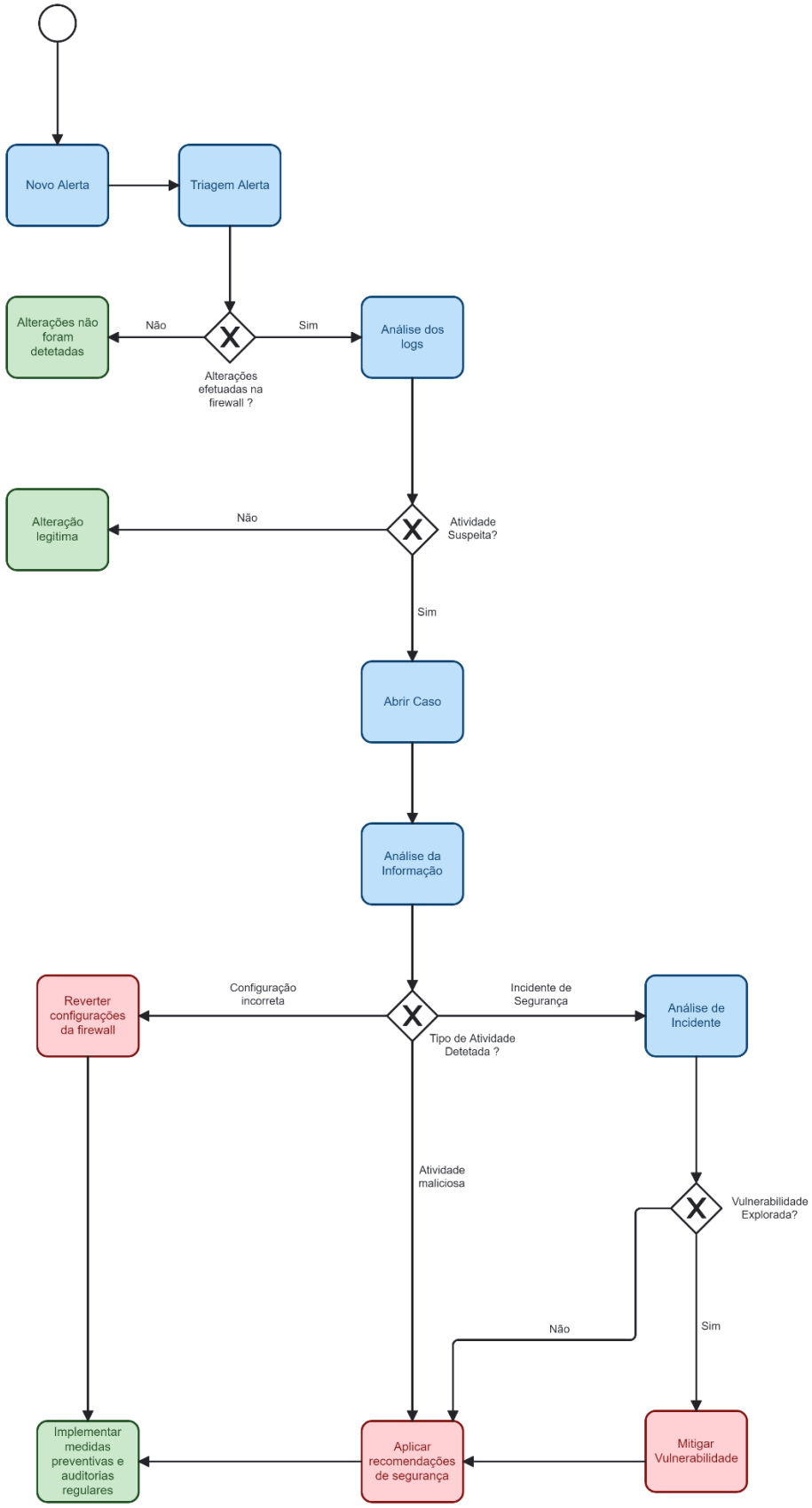

Figura 3- Fluxograma alteração de regras firewall

4.4 Nova conta de Utilizador

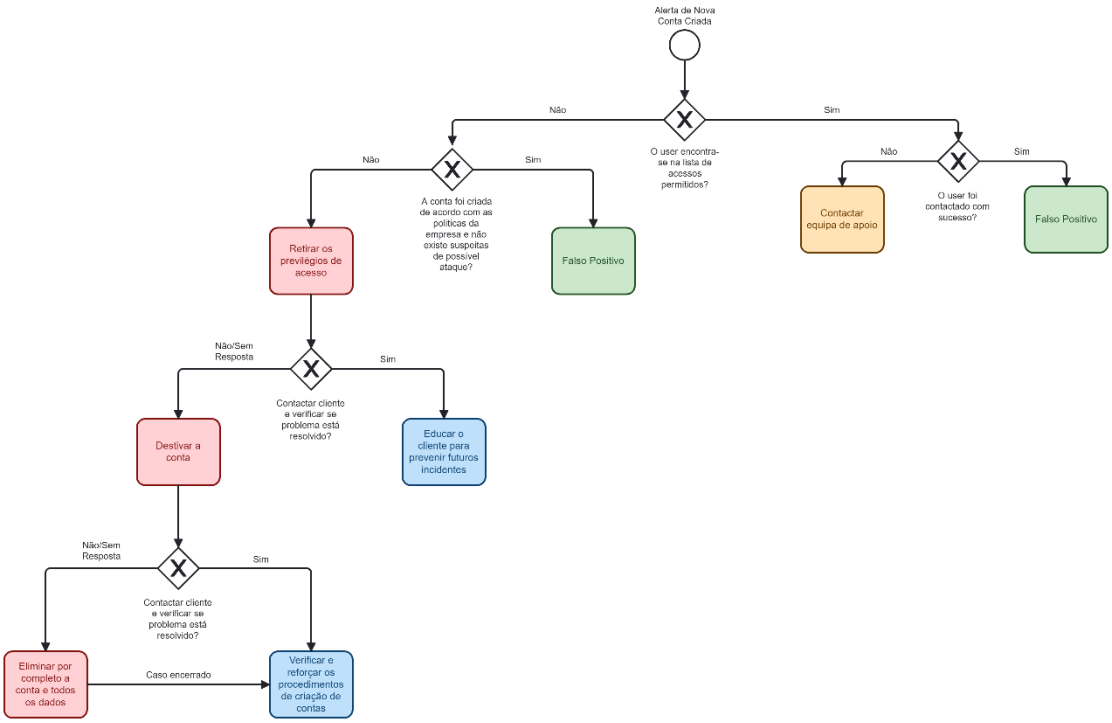

Figura 4 - Fluxograma nova conta de utilizador

Neste tipo de alerta (Figura 4) avaliamos o risco e as ações a tomar no caso da criação de uma nova conta de utilizador

Para fazer a deteção do alerta são monitorizados os eventos de criação de contas e utilizadores através de logs e ferramentas, gerando assim os alertas de possíveis ataques.

Após receber o alerta de nova conta criada, verificamos se o utilizador se encontra na lista de utilizadores conhecidos. Caso esteja na lista será notificado para averiguar a veracidade da conta criada e após a sua resposta positiva, o caso será considerado como falso positivo. Caso o utilizador não esteja presente na lista, não exista resposta ao ser notificado, ou indique que não foi ele a criar uma conta nova, será aberto um novo caso.

Após aberto, o caso começa a ser investigado. Inicialmente, se for confirmado o risco, são retirados os privilégios ou desativa-se a conta temporariamente, notificando o cliente sobre as ações. Se após notificar o cliente, não houver resposta ou ainda existir suspeitas de risco, a conta é removida completamente e o cliente é informado.

Em qualquer das etapas de resolução, assim que o caso for considerado resolvido, os processos de criação de contas são revistos, e caso necessário, implementam-se medidas de monitorização contínua para prevenir futuros incidentes.

4.5 Os dados do utilizador foram alterados

O fluxograma seguinte (Figura 5) representa um procedimento que os analistas do SOC seguirão ao executar tarefas de resposta a incidentes relacionados à alteração de conta de um utilizador. Funciona com base na verificação em várias etapas e é implementado como parte de um sistema de segurança.

Ao verificar um utilizador, o analista primeiro verifica se o utilizador é conhecido e se o mesmo fez a alteração à sua conta. Caso se verifique esta questão, considera-se um falso positivo, caso contrário passa-se para o próximo processo que é representado pelo levantamento dos IPs associados ao alerta.

Posteriormente, verifica-se o tipo de alteração e abre-se um caso, de modo que possa haver intervenção por outras partes. Neste tipo de casos é importante bloquear a conta temporariamente para que não haja alterações indesejáveis até que o caso esteja resolvido. O *Helpdesk* tentará reverter a mudança e contactar a vítima para que a mesma seja alertada da mudança e implementar medidas contracetivas com o objetivo de reduzir este tipo de ataques. Caso contrário, ou seja, não haja forma de contactar a vítima, passa-se ao bloqueio da conta, evitando problemas adicionais.

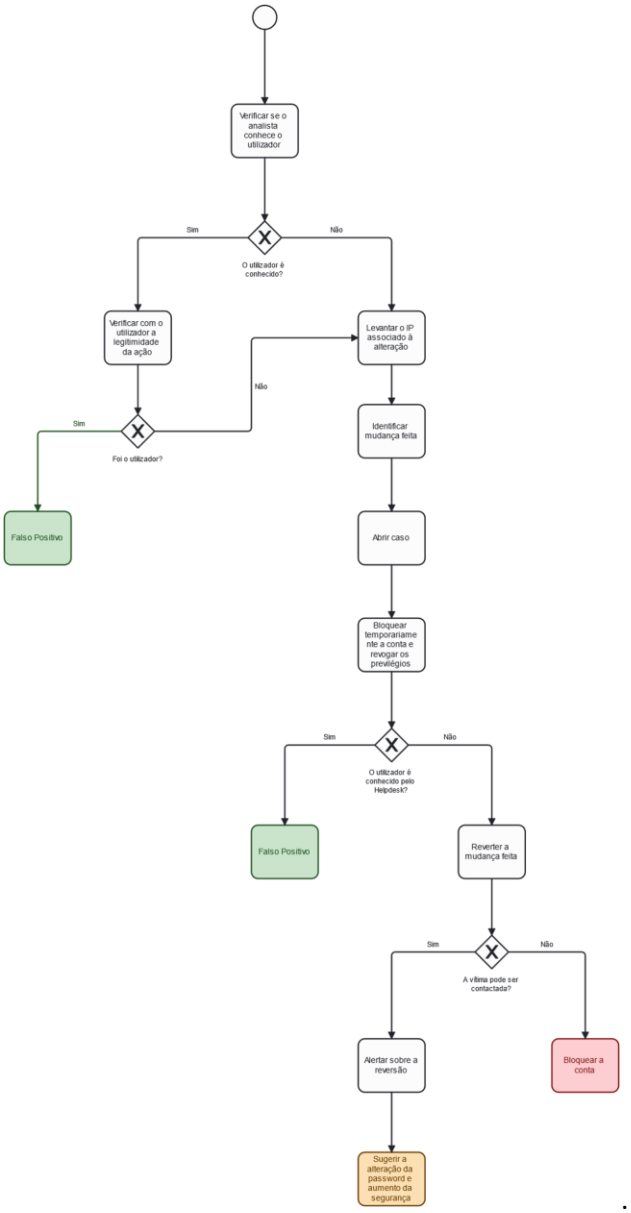

Figura 5 – Fluxograma de alteração de conta

5 Implementação

Nesta secção apresentaremos a arquitetura utilizada para a implementação do sistema pericial. Exploraremos a arquitetura do sistema, detalhando as tecnologias utilizadas, as decisões de design e os desafios enfrentados durante o desenvolvimento. Adicionalmente, serão apresentados os componentes individuais do sistema, como o *frontend* e o *backend*, e como eles interagem entre si.

5.1 Visão Geral da Arquitetura

O sistema pericial desenvolvido utiliza uma arquitetura cliente-servidor, onde o *frontend* em React comunica com um *backend* baseado em Java, que implementa as regras de negócio e a lógica de inferência utilizando *Drools*. A Figura 6 apresenta um diagrama de arquitetura simplificado, onde se destaca, os principais componentes e a interação entre eles:

- *Frontend* (*React*): Responsável pela interface com o utilizador, exibe os resultados processados e recolhe os dados para a consulta pericial.
- *Backend* (*Java/Drools*): Gere a lógica de negócio e as regras de inferência, processa as requisições recebidas e retorna as decisões baseadas nas regras definidas.

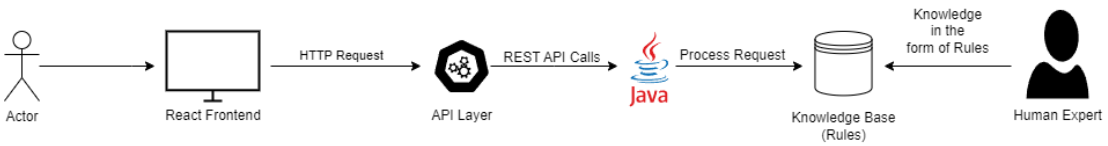

Figura 6. Diagrama da arquitetura do Sistema Pericial

A escolha de *React* para o *frontend* foi baseada na sua alta popularidade, robustez, capacidade de criar interfaces dinâmicas e reativas com facilidade e porque a equipa já tinha conhecimento prévio desta tecnologia. A integração com o *backend* através de APIs REST é facilitada pela simplicidade das bibliotecas de requisições HTTP como o Axios. O backend em Java foi um requisito da disciplina de ENGCIA, destaca-se especialmente na utilização de motores de regras como o Drools, que oferece uma poderosa infraestrutura para inferências baseadas em regras. Spring Boot é utilizado no ecossistema Java, o que facilita a construção de APIs robustas. Outras ferramentas como o *Git* e o *Maven* foram utilizadas para o controle de versões e para gestão de dependências e *build* do *backend* Java respetivamente.

5.2 Implementação do *Frontend*

A interface do utilizador foi projetada para ser intuitiva, permitindo que o utilizador faça consultas de forma eficiente e visualize os resultados das inferências do sistema. Os principais componentes estão apresentados na Figura 7 e incluem:

- Secção de Alertas (1): Permite ao utilizador escolher o alerta que pretende tratar.
- Secção de Histórico (2): Secção onde o utilizador pode aceder ao histórico de alertas tratados.
- Chat de Consultas (3): Onde o utilizador pode inserir dados que serão enviados ao backend para processamento, através de uma interação de pergunta e resposta.
- Botão “How?” (4): Exibe e justifica o caminho que o Sistema Pericial seguiu para obter uma dada conclusão.
- Botão “Drools/Prolog” (5): Permite a seleção do Sistema Pericial.

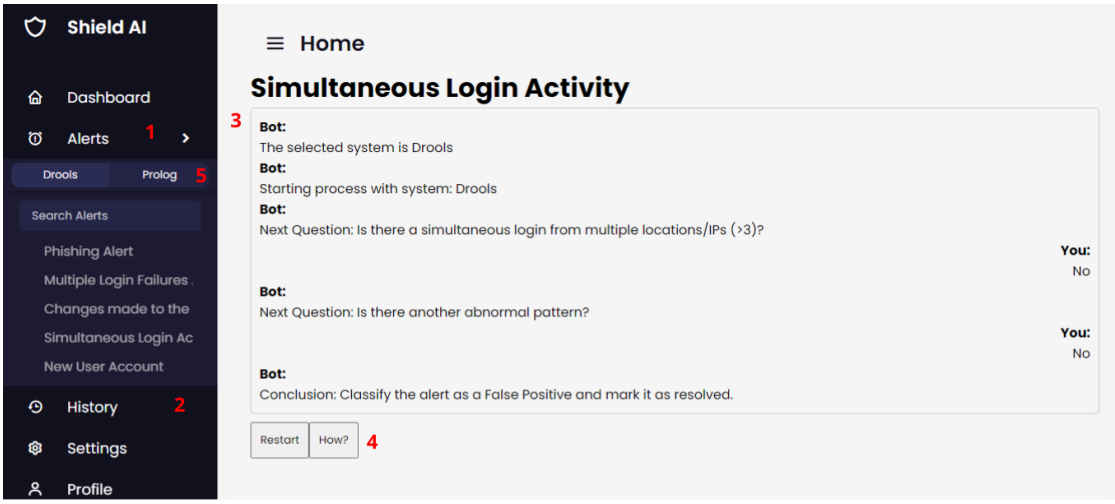

Figura 7. Componentes da interface gráfica do sistema

A lógica de negócio no frontend é responsável por organizar as requisições ao backend e exibir os resultados para o utilizador. Após a entrada do utilizador, os dados são formatados e enviados via API REST para o backend, que processa as regras de inferência. Uma vez que a resposta é recebida, os componentes são atualizados para refletir o resultado. A comunicação entre o frontend e o backend é feita através de APIs REST. Utilizamos a biblioteca Axios (ou Fetch API) para realizar requisições HTTP (POST, GET) de forma assíncrona.

5.3 Implementação do *Backend*

Os dados no backend são representados através de classes Java que modelam as entidades principais do sistema. Existem três principais tipos de classes:

- Evidencias: Reúne as informações obtidas nas interações com o utilizador.
- Resposta de alerta: Gere os resultados das inferências.
- Regras de Negócio: Contém as definições das regras de inferência que o Drools utilizará.

O backend é responsável por processar as requisições recebidas do frontend. A lógica de negócio principal envolve a execução das regras de inferência configuradas no Drools. Essas regras são definidas em arquivos .drl e são aplicadas sobre os dados recebidos nas consultas. O fluxo básico começa com o frontend a enviar uma consulta com dados relevantes, o backend recebe os dados e cria uma instância do objeto de evidencias, de seguida o motor de regras Drools processa as regras de inferência com base nesses dados e por fim os resultados são retornados ao frontend.

5.4 Exemplo de interação com o sistema

Nesta secção mostramos um exemplo de funcionamento do nosso sistema pericial no tratamento de um alerta de “Simultaneous Login Activity”, representado na Figura 8.

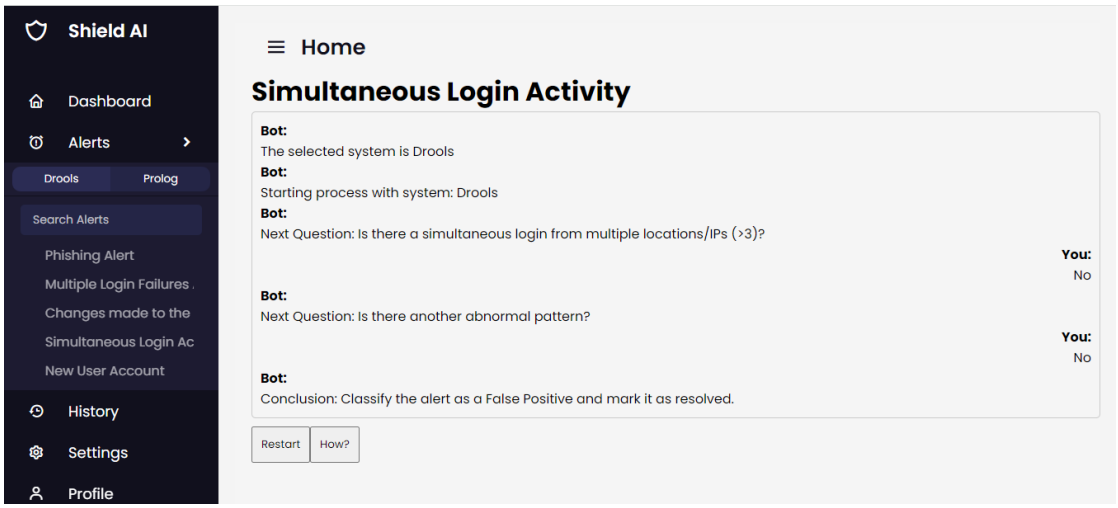

Figura 8. Exemplo de funcionamento do sistema

Inicialmente, o sistema começa com a seleção do motor de regras, neste caso, o Drools, conforme indicado pela mensagem inicial: "The selected system is Drools". Após essa seleção, o sistema inicia o processo de análise com a mensagem "Starting process with system: Drools", envio um pedido ao backend com um identificador desse alerta para que retorne a primeira

pergunta. Em seguida, o sistema começa a investigar possíveis logins simultâneos. A primeira pergunta feita ao utilizador é se há logins simultâneos a partir de múltiplas localizações ou IPs, onde o critério é mais de três logins: "Next Question: Is there a simultaneous login from multiple locations/IPs (>3)?". O utilizador responde negativamente a essa pergunta com um simples "No". Um novo pedido é enviado ao backend com a informação de que para a pergunta colocada a resposta foi negativa, para que assim o sistema possa retornar a próxima pergunta, ou uma conclusão. Na continuidade da análise, o sistema questiona se existe outro padrão anormal a ser considerado: "Next Question: Is there another abnormal pattern?". O utilizador, novamente, responde "No". Após a resposta um novo pedido é enviado ao backend e com base nas últimas evidencias o sistema conclui a análise e decide que o alerta pode ser classificado como um falso positivo. A decisão é então apresentada ao utilizador: "Conclusion: Classify the alert as a False Positive and mark it as resolved".

6 Conclusão

O objetivo principal do projeto foi desenvolver um sistema especialista pericial que possa auxiliar na resolução e classificação de alertas que chegam SOC. Através da implementação de um sistema baseado em regras/inferência e utilizando ferramentas como o Drools, foi possível criar um protótipo que incorpora a utilização de um sistema que automatiza parte do diagnóstico para reduzir o tempo de análise.

A quantidade crescente de alertas que os SOCs têm de gerir diariamente, juntamente com o problema da diferenciação entre ameaças reais e a quantidade de falsos positivos, mostra o potencial de soluções como a que foi proposta. Uma vez que os conhecimentos do sistema pericial se baseiam nos conhecimentos dos peritos em cibersegurança, a utilização do sistema pericial pode melhorar o processo de triagem. O sistema visa a clareza e facilidade de utilização por parte de qualquer utilizador independentemente do seu nível de experiência, separando os fluxos por tipo de alerta recebido, o que agiliza o diagnóstico e oferece soluções práticas para a correção dos problemas encontrados. A interação com o sistema é simples, baseada em perguntas objetivas e diretas, que se fundamentam na observação de comportamentos e na análise dos padrões de alerta. Mesmo na ausência de um especialista em cibersegurança, o operador consegue obter uma linha de ação clara e eficaz, permitindo que a resolução dos alertas seja realizada com eficiência e precisão.

Durante o desenvolvimento, enfrentaram-se desafios relacionados com a modelagem do conhecimento e a integração das regras no backend, mas as soluções encontradas mostraram-se eficazes. Este trabalho beneficiou da colaboração com o perito, Paulo Valdeira, bem como do acesso aos documentos da empresa necessários para a implementação de soluções em proximidade com o contexto atual do SOC.

Portanto, concluímos que este projeto tem o potencial de crescer, expandindo-se, adicionando novos casos de uso e integrando outros produtos como machine learning com o propósito de aumentar a controlabilidade da classificação dos alertas. E, dessa forma, a solução desenvolvida neste trabalho apresentam uma contribuição significativa na automação e na melhor operacionalização em contextos da cibersegurança, podendo ser uma ajuda tanto para um analista experiente como para um iniciante na área.

7 Bibliografia utilizada

[1] Shu-Hsien Liao, “Expert system methodologies and applications—a decade review from 1995 to 2004,” *Expert Syst Appl*, vol. 28, no. 1, pp. 93–103, 2005, doi: https://doi.org/10.1016/j.eswa.2004.08.003.

[2] M. Vielberth, F. Böhm, I. Fichtinger, and G. Pernul, “Security Operations Center: A Systematic Study and Open Challenges,” *IEEE Access*, vol. 8, pp. 227756–227779, 2020, doi: 10.1109/ACCESS.2020.3045514.

[3] B. U. Gelman, S. Taoufiq, T. Vörös, and K. Berlin, “That Escalated Quickly: An ML Framework for Alert Prioritization,” *ArXiv*, vol. abs/2302.06648, 2023, doi: 10.48550/arXiv.2302.06648.

[4] J. Huang, Z. An, S. Meckl, G. Tecuci, and D. Marcu, “Complementary Approaches to Instructable Agents for Advanced Persistent Threats Detection,” *Studies in Informatics and Control*, vol. 29, pp. 269–282, 2020, doi: 10.24846/V29I3Y202001.

8 Lista de terminologia específica

ENGCIA	*Engenharia do Conhecimento*
IP	*Internet Protocol*
IT/TI	*Information Tecnology/Tecnologias da Informação*
MFA	*Multifactor Authentication*
PPROGIA	*Paradigmas da Programação em Inteligência Artificial*
SOC	*Security Operations Center (centro de operações de segurança)*
SP	*Sistemas Periciais*
VPN	*Virtual Private Network*