PROGRAMA DE PÓS-GRADUAÇÃO PROFISSIONAL EM ENGENHARIA ELÉTRICA - PPEE
UNIVERSIDADE DE BRASÍLIA

A EFETIVIDADE DO PURPLE TEAM NA CIBERSEGURANÇA: UMA ABORDAGEM DE INTEGRAÇÃO ENTRE RED E BLUE TEAM

Bruno Rodrigues dos Reis, Daniel Chaves Cafe
emails: brodriguesdosreis@gmail.com, dcafe@unb.br

RESUMO

A crescente sofisticação das ameaças cibernéticas evidencia limitações em abordagens isoladas de segurança ofensiva e defensiva, tradicionalmente representadas pelos modelos de Red Team e Blue Team. Embora o Red Team seja eficaz na identificação de vulnerabilidades por meio de simulações de ataque e emulação de adversários, e o Blue Team atue na detecção, resposta e monitoramento contínuo, a atuação dissociada entre essas equipes pode gerar lacunas na validação prática dos controles implementados. Nesse contexto, o Purple Team surge como modelo integrador, promovendo colaboração estruturada entre ofensiva e defesa com foco na melhoria contínua da maturidade em segurança cibernética.
Este estudo propõe uma abordagem metodológica híbrida baseada em três camadas: (1) emulação ofensiva por meio de técnicas fundamentadas no framework MITRE ATT&CK, com execução controlada de Táticas, Técnicas e Procedimentos (TTPs) por meio de ataques manuais e de um agente de inteligência artificial (IA), ambos operados a partir de uma máquina Kali Linux; (2) monitoramento e detecção em ambientes Windows e Linux, com coleta e análise centralizada de logs por meio da plataforma Wazuh, permitindo a implementação e o aprimoramento contínuo de regras de detecção; e (3) integração Purple com ajustes iterativos nas regras, mecanismos de auditoria e respostas defensivas, utilizando a ferramenta VECTR para organização, rastreabilidade e comparação estruturada das técnicas executadas antes e após as melhorias implementadas. A pesquisa busca avaliar empiricamente a efetividade do modelo Purple Team na redução do tempo de detecção (TTD) e resposta (TTR), no aumento da cobertura de técnicas adversárias e na evolução da maturidade operacional do *Security Operations Center* (SOC). Como contribuição, o estudo propõe um conjunto estruturado de métricas para avaliação prática de exercícios colaborativos, oferecendo diretrizes aplicáveis a organizações que desejam implementar o Purple Team de forma sistemática e mensurável.

Palavras-chave: Purple Team; Red Team; Blue Team; Inteligência Artificial; Wazuh; VECTR; Adversary Emulation;

ABSTRACT

The increasing sophistication of cyber threats highlights the limitations of isolated offensive and defensive security approaches, traditionally represented by the Red Team and Blue Team models. While the Red Team is effective in identifying vulnerabilities through attack simulations and adversary emulation, and the Blue Team focuses on detection, response, and continuous monitoring, the lack of integration between these teams may create gaps in the practical validation of implemented security controls. In this context, the Purple Team emerges as an integrative model that promotes structured collaboration between offense and defense, aiming at the continuous improvement of cybersecurity maturity.
This study proposes a hybrid methodological approach based on three layers: (1) offensive emulation grounded in the MITRE ATT&CK framework, with controlled execution of Tactics, Techniques, and Procedures (TTPs) performed through both manual attacks and a dedicated artificial intelligence (AI) agent operated from a Kali Linux machine; (2) monitoring and detection in Windows and Linux environments, with centralized log collection and analysis through the Wazuh platform, enabling the implementation and continuous refinement of detection rules; and (3) Purple integration through iterative adjustments of detection rules, auditing mechanisms, and defensive responses, using the VECTR tool to ensure organization, traceability, and structured comparison of executed techniques before and after improvements. The research aims to empirically evaluate the effectiveness

of the Purple Team model in reducing the Time to Detect (TTD) and Time to Respond (TTR), increasing adversarial technique coverage, and enhancing the operational maturity of the Security Operations Center (SOC). As a contribution, this study proposes a structured set of metrics for the practical evaluation of collaborative exercises, providing guidelines for organizations seeking to implement the Purple Team model in a systematic and measurable manner.
Keywords: Purple Team; Red Team; Blue Team; Artificial Intelligence; Wazuh; VECTR; Adversary Emulation;

1 INTRODUÇÃO

Nos últimos anos, incidentes cibernéticos de grande repercussão midiática têm evidenciado a crescente vulnerabilidade de organizações públicas e privadas a ataques sofisticados. No Brasil, episódios amplamente noticiados — como o vazamento de dados de aproximadamente 223 milhões de CPFs em janeiro de 2021 (PSafe, 2021), o ataque de ransomware que paralisou os sistemas do Superior Tribunal de Justiça (STJ) em novembro de 2020 (G1, 2020) e a derrubada do sistema ConecteSUS do Ministério da Saúde em dezembro de 2021 (Agência Brasil, 2021) — expuseram a insuficiência das estratégias tradicionais de proteção frente a adversários cada vez mais organizados e persistentes. Em escala global, o relatório IBM Cost of a Data Breach 2023 registrou custo médio de US$ 4,45 milhões por violação de dados, o maior valor histórico até então (IBM Security, 2023), enquanto o Verizon Data Breach Investigations Report (DBIR) 2023 apontou que 74% das violações envolveram o fator humano, incluindo engenharia social, erros e uso indevido de credenciais (Verizon, 2023). Esses dados reforçam a urgência de abordagens mais integradas e eficazes para a defesa cibernética organizacional.

A transformação digital e a crescente dependência de infraestruturas tecnológicas ampliaram significativamente a superfície de ataque das organizações, tornando os ambientes corporativos mais expostos a ameaças cibernéticas sofisticadas (GRACIS, 2022). Técnicas modernas de adversários incluem campanhas estruturadas de emulação de adversários, uso de Táticas, Técnicas e Procedimentos (TTPs) avançados e exploração coordenada de múltiplas camadas da infraestrutura. Nesse cenário, a eficácia dos modelos tradicionais de segurança baseados em atuação isolada torna-se cada vez mais questionável.

Historicamente, as organizações estruturam suas estratégias de segurança em duas vertentes principais: o Red Team, responsável por simular ataques realistas com o objetivo de identificar vulnerabilidades e validar a exposição do ambiente (VAISHNAVI; ANANYA; AKILESH, 2024), e o Blue Team, focado na detecção, monitoramento e resposta a incidentes por meio de operações de segurança (Security Operations) e engenharia de detecção (Detection Engineering) (CHINDRUS; CARUNTU, 2023). Embora ambos os modelos desempenhem papéis essenciais, a atuação dissociada entre ofensiva e defesa pode gerar lacunas operacionais, dificultando a retroalimentação estruturada dos mecanismos de proteção e a evolução contínua da maturidade de segurança (GAIFULINA, 2025).

Nos últimos anos, surge o conceito de Purple Team como abordagem integradora, propondo colaboração ativa entre Red e Blue Team com foco em aprendizado contínuo, validação iterativa de controles e melhoria mensurável de métricas operacionais (ROUTIN; THOORES; ROSSIER, 2022). Diferentemente de exercícios tradicionais de penetração ou testes pontuais de intrusão, o Purple Team enfatiza ciclos estruturados de emulação de ameaças, monitoramento, ajuste de regras e revalidação, promovendo evolução sistemática da capacidade defensiva através de ciclos contínuos de feedback (GAIFULINA, 2025).

Apesar da crescente adoção do modelo Purple Team no setor corporativo e sua disseminação em frameworks técnicos e treinamentos especializados, observa-se escassez de estudos acadêmicos que avaliem empiricamente sua efetividade com base em métricas mensuráveis (GAIFULINA, 2025). Grande parte da literatura disponível concentra-se em relatos práticos, relatórios técnicos ou materiais de mercado, o que evidencia uma lacuna científica relevante quanto à validação objetiva dos ganhos operacionais proporcionados pela integração entre ofensiva e defesa.

Diante desse contexto, este estudo propõe avaliar a efetividade do Purple Team por meio de uma abordagem metodológica estruturada em três camadas: (1) emulação ofensiva baseada no framework MITRE ATT&CK, com execução controlada de TTPs por meio de ataques manuais e de um agente de inteligência artificial (IA), ambos operados a partir de uma máquina Kali Linux; (2) monitoramento e detecção centralizados em ambiente SIEM, com coleta de métricas como Time to Detect (TTD), Time to Respond (TTR), taxa de detecção por técnica e análise de falsos positivos e negativos (GAIFULINA, 2025); e (3) integração iterativa Purple, com ajustes de regras, mecanismos de auditoria e protocolos de resposta, seguida de reexecução controlada dos mesmos cenários para comparação estatística antes e depois das melhorias implementadas.

A pesquisa busca responder à seguinte questão: em que medida a integração estruturada entre Red Team e Blue Team, sob o modelo Purple Team, contribui para a melhoria mensurável da capacidade de detecção, resposta e cobertura de técnicas adversárias em um Security Operations Center (SOC)?

Como contribuição, o estudo propõe um modelo estruturado de avaliação prática baseado em métricas operacionais e rastreabilidade técnica, oferecendo diretrizes aplicáveis para organizações que desejam implementar o Purple Team de forma sistemática, mensurável e orientada à maturidade em cibersegurança.

O presente artigo está organizado em cinco seções. Na Seção 2, é apresentada a revisão bibliográfica dos temas correlatos, abrangendo as diretrizes normativas para segurança cibernética — com destaque para o Programa de Privacidade e Segurança da Informação (PPSI 2.0) e a articulação deste trabalho com o Controle 18 (Testes de Intrusão) —, os principais frameworks de classificação de ataques cibernéticos (OWASP e MITRE ATT&CK) e o estado da arte sobre Red Team, Blue Team e Purple Team na literatura científica. Na Seção 3, descreve-se a metodologia experimental adotada, detalhando o ambiente laboratorial virtualizado, os critérios de seleção das TTPs, as três camadas da abordagem proposta — emulação ofensiva conduzida por ataques manuais e por um agente de inteligência artificial (IA) a partir do Kali Linux, monitoramento e detecção via Wazuh, e integração Purple operacionalizada pela metodologia PEIR e pela ferramenta VECTR — e as métricas operacionais utilizadas para avaliação dos resultados. A Seção 4 apresenta a análise consolidada dos resultados obtidos nos experimentos, com a comparação estatística entre os ciclos Baseline e Pós-ajuste. Por fim, a Conclusão sintetiza as contribuições do estudo, discute suas limitações e aponta perspectivas para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

2.1 Diretrizes Normativas para Segurança Cibernética

A crescente complexidade do ambiente cibernético tem impulsionado governos e organismos reguladores a estabelecerem políticas formais para a proteção de infraestruturas digitais. No contexto brasileiro, o Programa de Privacidade e Segurança da Informação (PPSI 2.0), instituído pela Portaria SGD/MGI nº 9.511/2025 e com vigência a partir de 1º de janeiro de 2026, representa o esforço centralizado da Administração Pública Federal para organizar e padronizar suas práticas de segurança e privacidade junto aos órgãos integrantes do Sistema de Administração dos Recursos de Tecnologia da Informação (SISP) (BRASIL, 2026). Desenvolvido pela Secretaria de Governo Digital (SGD) do Ministério da Gestão e da Inovação em Serviços Públicos (MGI), o PPSI 2.0 estrutura-se em três segmentos: governança (Controles 0), segurança da informação (Controles 1 a 18, fundamentados no CIS Controls e correlacionados ao marco regulatório brasileiro) e privacidade (Controles 19 a 25, alinhados à Lei Geral de Proteção de Dados Pessoais — LGPD).

O Controle 18 do PPSI 2.0 trata especificamente dos Testes de Intrusão (Pentest), definido como a atividade de "testar a eficácia e a resiliência dos ativos institucionais e das soluções de software, identificando e explorando vulnerabilidades e simulando os objetivos e ações de um invasor" (BRASIL, 2026). O referido controle é composto por cinco medidas: elaboração e manutenção de programa de testes de intrusão (18.1); realização periódica de testes externos (18.2); correção das descobertas identificadas (18.3); validação das medidas de segurança após cada ciclo de testes (18.4); e realização de testes de intrusão internos periódicos (18.5). Silva e Gondim (2025) observam que, embora o PPSI oriente os órgãos da Administração Pública Federal à aplicação do Controle 18, ele não prescreve uma metodologia padronizada para sua execução, evidenciando a necessidade de abordagens estruturadas — como a integração com as TTPs do MITRE ATT&CK — para elevar a eficácia e a rastreabilidade dos testes. Esse contexto normativo reforça a relevância do modelo Purple Team como mecanismo de operacionalização sistemática e mensurável do Controle 18, promovendo ciclos iterativos de emulação ofensiva, detecção e remediação alinhados aos requisitos do PPSI 2.0.

No âmbito internacional, estruturas normativas como a ISO/IEC 27001:2022 e o NIST Cybersecurity Framework (CSF) fornecem diretrizes amplamente adotadas para a gestão de riscos cibernéticos, contemplando controles técnicos e processuais que sustentam as atividades de segurança ofensiva e defensiva. A convergência entre esses frameworks e as abordagens práticas de Red Team, Blue Team e Purple Team evidencia a maturidade crescente do setor, que passa a integrar práticas operacionais antes restritas a ambientes militares e de inteligência em contextos corporativos e institucionais (GRACIS, 2022).

2.2 Tipos de Ataques Cibernéticos e Frameworks de Classificação

A compreensão sistemática dos ataques cibernéticos é condição fundamental para o desenvolvimento de defesas eficazes. Nesse sentido, dois frameworks de referência se destacam na literatura e na prática: o OWASP (Open Web Application Security Project) e o MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge).

O OWASP é uma fundação sem fins lucrativos que documenta e classifica as principais vulnerabilidades em aplicações web, sendo o OWASP Top 10 amplamente utilizado como referência para avaliações de segurança em sistemas orientados à web. Sua adoção como linha de base em testes de segurança permite que equipes ofensivas e defensivas compartilhem uma linguagem comum para a identificação e mitigação de vulnerabilidades de alto impacto. (OWASP FOUNDATION, 2025)

O framework MITRE ATT&CK representa um avanço significativo na categorização do comportamento adversarial, organizando as táticas e técnicas utilizadas por atacantes reais em uma base de conhecimento estruturada e continuamente atualizada. Com cobertura para ambientes Windows, Linux, macOS, nuvem e sistemas de controle industrial (ICS/OT), o MITRE ATT&CK fornece uma taxonomia padronizada para descrever Táticas, Técnicas e Procedimentos (TTPs) utilizados em campanhas de ameaças avançadas persistentes (APT). Essa padronização viabiliza a comunicação precisa entre equipes ofensivas e defensivas, constituindo o alicerce técnico sobre o qual o modelo Purple Team é construído (ROUTIN; THOORES; ROSSIER, 2022).

2.3 Red Team e Blue Team: Fundamentos e Literatura Científica

O Red Team consiste em um grupo de profissionais de segurança que simulam adversários reais com o objetivo de identificar vulnerabilidades em pessoas, processos e tecnologias antes que agentes maliciosos o façam. Diferentemente de testes de penetração convencionais (VAPT), o Red Team opera com escopo amplo, autonomia tática e utilização de TTPs baseadas em inteligência de ameaças, aproximando-se do comportamento real de atacantes avançados (VAISHNAVI; ANANYA; AKILESH, 2024). Essa abordagem permite não apenas a identificação de vulnerabilidades técnicas, mas também a avaliação da eficácia dos controles de detecção e resposta existentes.

O Blue Team é responsável pela defesa ativa da infraestrutura organizacional, atuando no monitoramento contínuo, detecção de anomalias, análise de indicadores de comprometimento (IOC) e resposta a incidentes. Suas operações são frequentemente estruturadas em um Security Operations Center (SOC), ambiente especializado que centraliza a coleta e correlação de eventos de segurança por meio de plataformas SIEM (Security Information and Event Management). O SOC moderno incorpora práticas de Detection Engineering, que envolvem a criação, validação e refinamento sistemático de regras e alertas com base em inteligência de ameaças e análise comportamental (GRACIS, 2022).

Chindrus e Caruntu (2023) demonstram, em estudo de caso baseado em competição de segurança cibernética, que a interação entre equipes ofensivas e defensivas em ambiente controlado gera aprendizado mútuo e aprimora a capacidade de detecção. Os autores identificam que a ausência de comunicação estruturada entre Red e Blue Team resulta em detecções tardias e ausência de retroalimentação efetiva dos controles implementados, corroborando a necessidade de modelos integrativos.

2.4 Purple Team: Estado da Arte e Modelo Integrador

O Purple Team emerge como resposta à fragmentação operacional entre segurança ofensiva e defensiva, propondo um modelo de colaboração estruturada em que Red e Blue Teams atuam de forma coordenada e iterativa. Conforme Routin, Thoores e Rossier (2022), o Purple Team não constitui uma equipe independente, mas sim uma função integradora que estabelece ciclos contínuos de emulação de ameaças, monitoramento, ajuste de controles e revalidação.

Gaifulina (2025) define a sinergia entre Red Team e Blue Team como fator determinante para o incremento da resiliência organizacional frente a ataques cibernéticos, destacando que a integração estruturada entre ofensiva e defesa permite a identificação de lacunas que permaneceram ocultas em operações isoladas. O autor aponta ainda a escassez de estudos empíricos que avaliem quantitativamente os ganhos proporcionados pelo modelo Purple Team, confirmando a lacuna científica que este estudo se propõe a endereçar.

Essa percepção é corroborada pelos achados bibliográficos realizados no âmbito desta pesquisa. As buscas nas principais bases científicas indexadas retornaram produção formal limitada — IEEE Xplore (4 conferências e 3 livros), ACM Digital Library (7 resultados) e Portal de Periódicos CAPES (2 publicações). Bases de cobertura ampla retornam volumes muito superiores; contudo, esses números brutos incluem literatura cinzenta (relatórios técnicos, white papers, teses e preprints) e correspondências espúrias dos termos genéricos “purple” e “team”, o que os torna inadequados como medida direta de produção científica. Refinando a busca no Semantic Scholar para a expressão exata associada ao contexto de cibersegurança, identificaram-se 23 trabalhos efetivamente pertinentes, distribuídos a partir de 2019, com concentração nos anos mais recentes e pico em 2025 (Figura 1). Tal cenário confirma que o Purple Team encontra-se em fase de consolidação científica, com predominância de relatos técnicos e publicações de mercado sobre estudos empíricos de metodologia rigorosa.

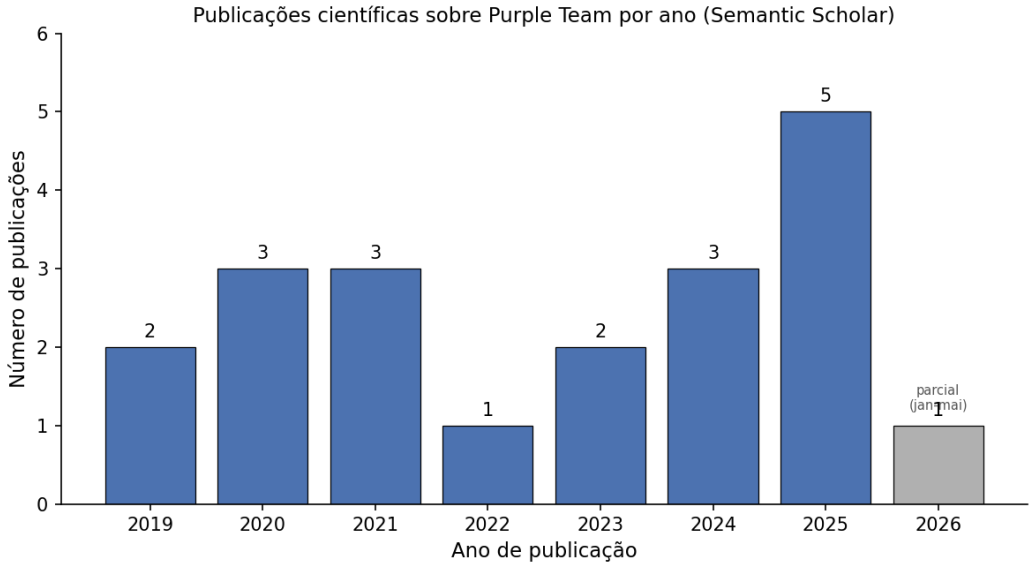
Tabela 1 — Levantamento bibliográfico sobre Purple Team por base de dados.

Base	Termo(s) de busca	Período	Cobertura	Resultados pertinentes
IEEE Xplore	"Purple Team", "Purple Teaming", "Red Team"	2021–2026	Indexada / revisada por pares	4 conferências + 3 livros
ACM Digital Library	"Purple Team"	2021–2026	Indexada / revisada por pares	7

Portal CAPES	"Purple Team"	2021–2026	Indexada / revisada por pares	2 (1 livro + 1 artigo)
Semantic Scholar	("purple team" OR "purple teaming") AND (security OR cybersecurity OR "red team" OR "blue team" OR SOC OR SIEM OR MITRE OR threat OR intrusion OR defense)	até 2026	Agregadora (busca refinada)	23

Fonte: o autor (2026).

Figura 1 — Distribuição anual de publicações científicas sobre Purple Team em contexto de cibersegurança (Semantic Scholar; busca refinada, n = 23). O ano de 2026 é parcial (jan–mai).



Fonte: o autor (2026).

A metodologia PEIR (Preparar, Executar, Identificar e Remediar) fornece estrutura operacional para a condução de exercícios Purple Team, organizando o ciclo de emulação em quatro fases distintas: planejamento dos cenários de ataque com base em inteligência de ameaças; execução controlada das TTPs mapeadas no MITRE ATT&CK; identificação das lacunas de detecção e resposta; e remediação iterativa dos controles. Essa estrutura metodológica orienta a organização dos experimentos e a coleta de métricas operacionais propostas neste estudo (ROUTIN; THOORES; ROSSIER, 2022).

3 METODOLOGIA

3.1 Caracterização da Pesquisa

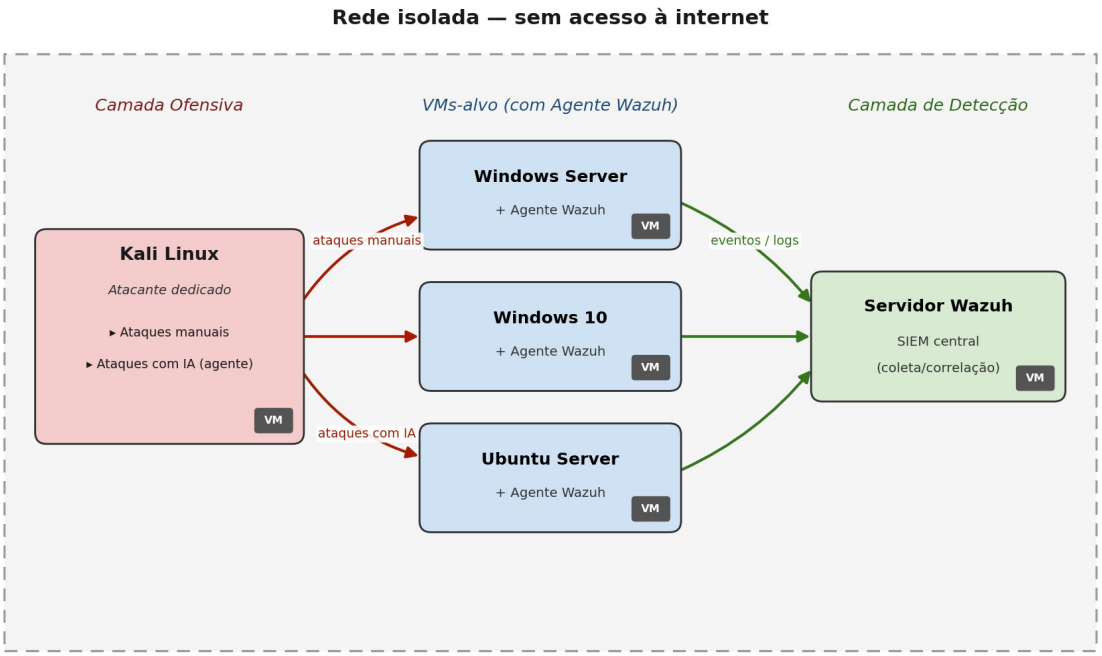
Esta pesquisa se caracteriza como experimental, de natureza quantitativa, com design de estudo de caso controlado em ambiente laboratorial. A abordagem adotada segue o paradigma de pesquisa aplicada, com foco na avaliação empírica de um modelo estruturado de integração entre segurança ofensiva e defensiva. O delineamento experimental baseia-se em um modelo de comparação pré e pós-intervenção, no qual os mesmos cenários de ataque são executados em dois momentos distintos: antes e após a implementação dos ajustes de detecção e remediação oriundos do ciclo Purple. Esse design permite a mensuração objetiva dos ganhos operacionais obtidos pela integração estruturada entre Red e Blue Team, com base em métricas operacionais coletadas em ambas as fases (GAIFULINA, 2025; SCARFONE et al., 2008).

3.2 Ambiente Experimental

O ambiente experimental foi construído em infraestrutura virtualizada, composta por máquinas virtuais (VMs) com sistemas operacionais Windows e Linux, de forma a reproduzir um cenário corporativo heterogêneo representativo dos ambientes reais encontrados em organizações. A topologia contempla: (1) uma máquina

atacante dedicada com Kali Linux, a partir da qual a camada ofensiva é operada em duas modalidades complementares — ataques manuais, conduzidos interativamente por um operador humano, e ataques automatizados, conduzidos por um agente de inteligência artificial (IA) dedicado —, simulando o comportamento de um atacante real com acesso interno à rede; (2) VMs-alvo com sistemas Windows (Windows Server e Windows 10) e Linux (Ubuntu Server), configuradas com serviços ativos e agentes de monitoramento instalados; e (3) um servidor Wazuh atuando como gerenciador central do SIEM, responsável pela coleta, normalização e correlação dos eventos de segurança gerados em todos os endpoints. A Figura 2 sintetiza a topologia descrita, identificando cada agente, seu papel no experimento e sua natureza virtualizada (VM).

Figura 2 — Diagrama de blocos da topologia do ambiente experimental: camada ofensiva operada a partir do Kali Linux (ataques manuais e ataques automatizados com agente de IA), VMs-alvo com agentes Wazuh e servidor Wazuh (SIEM central), todos virtualizados (VM) e interconectados em rede isolada.



Fonte: o autor (2026).

Todos os componentes estão interconectados em rede isolada, sem acesso à internet, garantindo o controle total do ambiente experimental e a segurança do laboratório. A isolação da rede assegura que os eventos registrados sejam exclusivamente originados pelos cenários de ataque planejados, eliminando ruídos provenientes de tráfego externo e aumentando a precisão das métricas coletadas.

3.3 Seleção das Técnicas Adversariais

A seleção das Táticas, Técnicas e Procedimentos (TTPs) que compõem os cenários de ataque foi realizada com base no framework MITRE ATT&CK para sistemas empresariais (Enterprise ATT&CK), adotado como linguagem padronizada para descrição e categorização do comportamento adversarial (ROUTIN; THOORES; ROSSIER, 2022). Os critérios de seleção consideraram: (1) relevância das técnicas para o contexto de infraestrutura corporativa, com foco em táticas de alto impacto operacional; (2) viabilidade de execução das técnicas tanto por meio de ataques manuais conduzidos no Kali Linux quanto por automação orquestrada pelo agente de IA; (3) cobertura equilibrada entre as principais fases da cadeia de ataque (kill chain), incluindo reconhecimento, execução, persistência, escalada de privilégios, movimentação lateral, exfiltração e evasão de defesas; e (4) disponibilidade de regras de detecção correspondentes configuráveis no Wazuh.

As técnicas selecionadas foram mapeadas previamente em uma matriz de cobertura ATT&CK, documentada no VECTR, permitindo rastreabilidade completa entre os cenários executados, os alertas gerados e os ajustes realizados nas regras de detecção ao longo do ciclo iterativo Purple.

3.4 Camada 1 — Emulação Ofensiva

A camada ofensiva é responsável pela execução controlada das TTPs selecionadas, simulando o comportamento de um adversário real em ambiente corporativo. Toda a operação ofensiva é conduzida a partir de uma máquina dedicada com Kali Linux, em duas modalidades complementares: (1) ataques manuais, executados interativamente por um operador humano; e (2) ataques automatizados, orquestrados por um agente de inteligência artificial (IA). Essa combinação permite contrastar o comportamento adversarial humano — adaptativo e não determinístico — com a execução autônoma e escalável proporcionada pela IA, ampliando a cobertura e o realismo dos cenários avaliados.

Os ataques manuais são executados a partir do Kali Linux por um operador humano que assume o papel de atacante dedicado, reproduzindo o comportamento de uma ameaça do tipo insider — usuário interno mal-intencionado ou comprometido, com acesso legítimo à rede e conhecimento prévio do ambiente. Essa modalidade explora vulnerabilidades de forma adaptativa e não determinística, contemplando técnicas como reconhecimento interno, movimentação lateral, escalada de privilégios interativa e evasão de defesas ajustada ao alvo. A execução manual aproxima os cenários de teste de situações reais de comprometimento, nas quais o adversário toma decisões contextuais dificilmente reproduzíveis por automações genéricas.

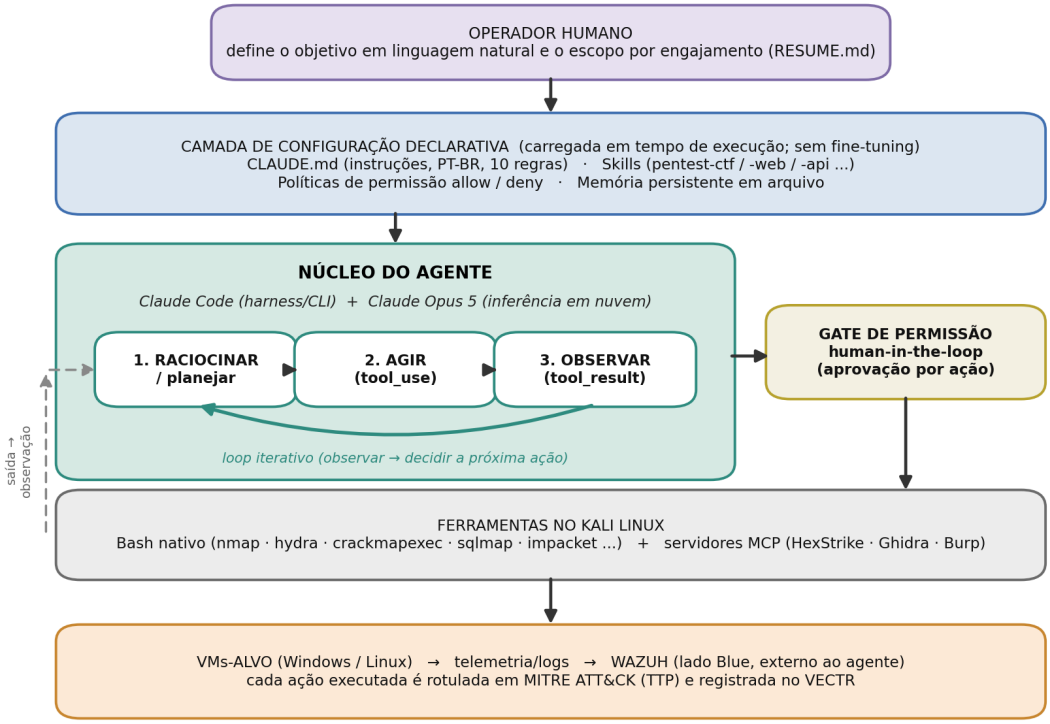
Os ataques automatizados são conduzidos por um agente de inteligência artificial dedicado, também hospedado no ambiente Kali Linux, responsável por planejar e executar de forma autônoma sequências de TTPs mapeadas no framework MITRE ATT&CK. A automação por IA assegura a reprodutibilidade exigida pelo delineamento experimental, pois permite que os mesmos cenários sejam reexecutados de forma consistente antes e após os ajustes defensivos, viabilizando a comparação estatística dos resultados (SILVA; GONDIM, 2025). Além disso, a capacidade do agente de encadear técnicas e adaptar a execução conforme as respostas do ambiente amplia a cobertura de comportamentos adversariais avaliados, complementando os ataques manuais com uma camada de automação inteligente e escalável.

A combinação das duas modalidades — execução manual interativa e automação conduzida por IA — busca equilibrar realismo e reprodutibilidade: enquanto os ataques manuais reproduzem a imprevisibilidade e a criatividade de um adversário humano, o agente de IA assegura a padronização e a repetibilidade necessárias à comparação estatística entre os ciclos Baseline e Pós-ajuste. Independentemente da modalidade, todas as técnicas executadas são mapeadas no MITRE ATT&CK e registradas no VECTR, enriquecendo a cobertura dos cenários avaliados e aumentando o realismo do exercício Purple Team.

3.4.1 Detalhamento do Agente de Inteligência Artificial

O agente de inteligência artificial responsável pela modalidade automatizada não constitui uma ferramenta ofensiva autônoma desenvolvida especificamente para esta pesquisa, tampouco um modelo submetido a treinamento ou a ajuste fino (fine-tuning) para a tarefa de teste de intrusão. Trata-se de um agente de propósito geral — operacionalizado pelo Claude Code, interface de linha de comando (CLI) da Anthropic, e movido pelo modelo de linguagem de grande porte Claude Opus 5 — especializado para a emulação ofensiva exclusivamente por configuração declarativa. O binário do agente é executado localmente na máquina Kali Linux, ao passo que a inferência do modelo ocorre de forma remota, em nuvem; não há, portanto, pesos, treinamento ou ambiente de inferência hospedados no laboratório. A especialização resulta de camadas de arquivos de instrução interpretados em tempo de execução: um arquivo de instruções de projeto — que define o idioma de saída, o papel assumido, as regras transversais de metodologia e o padrão de documentação —, módulos de habilidade (skills) que encapsulam procedimentos específicos por tipo de alvo, políticas de permissão e uma memória persistente em arquivo. A Figura 3 sintetiza essa arquitetura.

Figura 3 — Arquitetura do agente de inteligência artificial para a emulação ofensiva automatizada: camada de configuração declarativa, ciclo iterativo de raciocínio e ação (ReAct) executado sobre ferramentas do Kali Linux, supervisão humana obrigatória (human-in-the-loop) e rotulagem das ações executadas em MITRE ATT&CK.



Fonte: o autor (2026).

No que se refere à arquitetura de raciocínio, o agente opera segundo um ciclo do tipo ReAct (Reasoning and Acting), no qual raciocínio e ação são intercalados de forma iterativa (YAO et al., 2023). A partir do objetivo definido pelo operador, o modelo raciocina e decide uma ação, emite a chamada de uma ferramenta (tool_use), recebe o resultado da execução (tool_result) como observação e, com base nessa observação, decide o passo seguinte, repetindo o ciclo até a conclusão da tarefa. Não há planejador simbólico, árvore de busca explícita ou grafo de ataque formal: o encadeamento das técnicas emerge do raciocínio sobre a saída de cada ferramenta. O agente mantém estado em três níveis — o contexto volátil da sessão, uma memória persistente em arquivo e o registro do estado de cada engajamento —, o que lhe permite preservar decisões relevantes ao longo das etapas.

A camada de ação combina ferramentas nativas do ambiente e servidores complementares. A execução de comandos de shell torna invocável qualquer utilitário presente no Kali Linux — por exemplo, nmap, hydra, crackmapexec, impacket e sqlmap —, enquanto servidores baseados no protocolo MCP (Model Context Protocol) disponibilizam capacidades adicionais de análise. A saída de cada ferramenta retorna ao contexto do agente em forma textual e condiciona diretamente a decisão subsequente, materializando o elo entre observação e ação que caracteriza o ciclo iterativo.

Quanto à autonomia e à governança, o agente apresenta autonomia elevada na condução de uma tarefa, porém subordinada a supervisão humana obrigatória (human-in-the-loop). O objetivo e o escopo de cada engajamento são sempre definidos pelo operador em linguagem natural, e toda ação candidata é submetida a um mecanismo de permissão que autoriza ou bloqueia a execução de cada ferramenta. No delineamento experimental controlado desta pesquisa, tal característica é relevante: a contenção do escopo é procedimental — assegurada pela instrução e pela aprovação humana a cada passo —, em coerência com a operação em rede isolada descrita na Seção 3.2.

Por fim, o agente emprega o framework MITRE ATT&CK como ontologia de documentação e comunicação: cada ação executada é associada à respectiva tática e técnica (TTP) e registrada no VECTR, preservando a rastreabilidade exigida pelo ciclo Purple (Seção 3.6). Esse mapeamento é declarativo — apoiado no conhecimento do modelo e revisado pelo operador —, e não decorre de consulta programática à base do MITRE. A principal contribuição do agente para o delineamento experimental reside na padronização e na repetibilidade da campanha ofensiva: ao executar o mesmo conjunto predefinido de TTPs, contra os mesmos alvos, nos ciclos Baseline e Pós-ajuste, o agente reduz a variabilidade operacional entre as execuções e favorece a atribuição das diferenças observadas nas métricas aos ajustes defensivos, e não a variações na conduta do atacante — complementando o realismo adaptativo dos ataques manuais.

3.5 Camada 2 — Monitoramento e Detecção

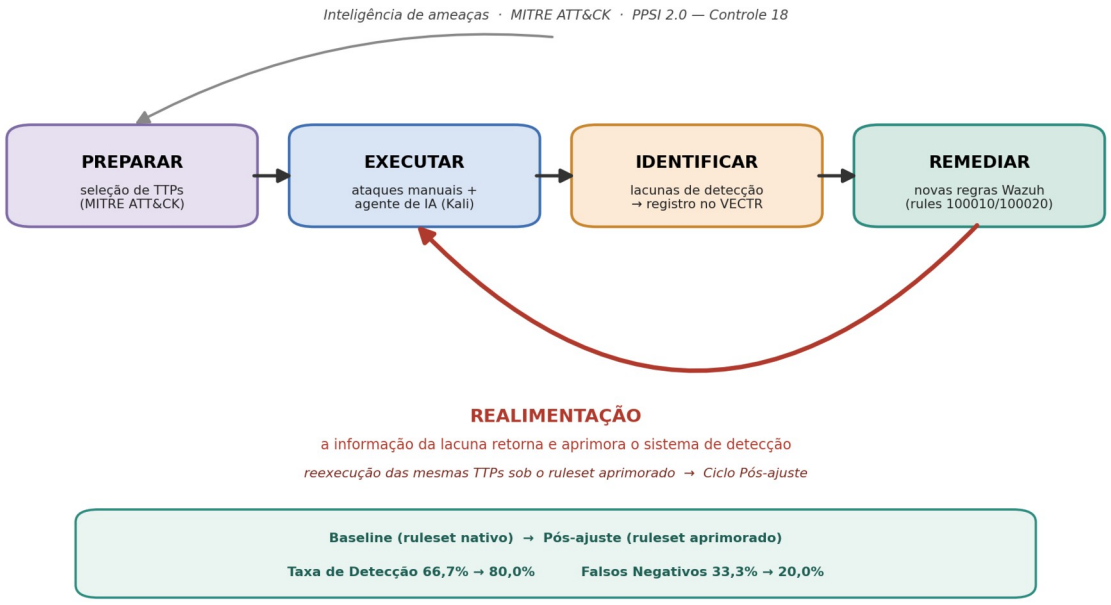
A camada de detecção é operacionalizada pela plataforma Wazuh, solução de código aberto que funciona simultaneamente como SIEM (Security Information and Event Management), sistema de detecção de intrusão em host (HIDS) e plataforma de resposta a incidentes. O Wazuh é responsável pela coleta centralizada de logs provenientes de todos os endpoints do ambiente experimental, incluindo logs de eventos do Windows (Windows Event Logs), registros do sistema Linux (Syslog, auditd), logs de aplicações e alertas de integridade de arquivos (FIM — File Integrity Monitoring).

Para cada TTP executada pela camada ofensiva, o Wazuh é configurado com regras de detecção correspondentes, desenvolvidas com base em indicadores de comprometimento (IOC) e padrões comportamentais associados às técnicas do MITRE ATT&CK. A ausência de regra de detecção ou a geração de alerta sem correlação adequada é registrada como lacuna de detecção, documentada no VECTR para orientar os ajustes da fase Purple. As métricas coletadas nesta camada incluem: o instante de execução de cada técnica (t_ataque), o instante de geração do primeiro alerta correspondente (t_alerta), e o instante de início da ação de resposta (t_resposta), que permitem o cálculo das métricas operacionais definidas na Seção 3.7.

3.6 Camada 3 — Integração Purple e Ciclo Iterativo

A camada Purple constitui o núcleo integrativo da metodologia proposta, operacionalizando o ciclo de feedback estruturado entre ofensiva e defesa por meio da metodologia PEIR — Preparar, Executar, Identificar e Remediar (ROUTIN; THOORES; ROSSIER, 2022). A ferramenta VECTR (Vulnerability and Exploit Consolidation Tool for Reporting) é utilizada para organizar, registrar e documentar todos os cenários de ataque, resultados de detecção e ajustes implementados ao longo do processo. A Figura 4 sintetiza o ciclo PEIR e o fluxo de realimentação entre as equipes ofensiva e defensiva.

Figura 4 — Ciclo PEIR (Preparar, Executar, Identificar e Remediar) e o fluxo de realimentação entre Red Team e Blue Team: as lacunas de detecção identificadas retornam como novas regras e telemetria, e a reexecução das mesmas TTPs sob o ruleset aprimorado caracteriza o Ciclo Pós-ajuste.



Fonte: o autor (2026).

O ciclo PEIR é conduzido em dois momentos experimentais distintos: (1) Ciclo Baseline: execução inicial das TTPs selecionadas sem ajustes prévios nas regras de detecção, representando o estado de maturidade defensiva pré-Purple. Os resultados são registrados no VECTR como linha de base, documentando quais técnicas foram detectadas, com qual latência e com quantos falsos positivos; (2) Ciclo Pós-Ajuste: após análise colaborativa dos resultados do Ciclo Baseline — etapa que caracteriza a integração Purple — são realizados ajustes nas regras de detecção do Wazuh, revisão de políticas de auditoria dos sistemas operacionais e atualização de playbooks de resposta. Em seguida, as mesmas TTPs são reexecutadas sob condições idênticas, e os novos resultados são registrados no VECTR para comparação direta com a linha de base.

Essa estrutura de dois ciclos permite a avaliação objetiva do impacto da integração Purple na maturidade operacional do SOC, quantificando a evolução das métricas de detecção e resposta entre os dois momentos experimentais.

3.7 Métricas de Avaliação

As métricas operacionais adotadas neste estudo foram definidas com base na literatura especializada em operações de SOC e avaliação de exercícios Purple Team (GAIFULINA, 2025; BALADARI, 2021), e são calculadas individualmente para cada TTP executada em ambos os ciclos experimentais:

Time to Detect (TTD): intervalo de tempo, em segundos, entre o instante de execução da técnica adversarial (t_ataque) e o instante de geração do primeiro alerta correspondente no Wazuh (t_alerta). Formalmente: TTD = t_alerta − t_ataque. Valores menores indicam maior capacidade de detecção oportuna.

Time to Respond (TTR): intervalo de tempo, em segundos, entre o instante de geração do alerta (t_alerta) e o instante de início da ação de resposta pelo analista ou por mecanismo automatizado (t_resposta). Formalmente: TTR = t_resposta − t_alerta. Reflete a agilidade do processo de resposta a incidentes.

Taxa de Detecção (TD): proporção de técnicas executadas que geraram ao menos um alerta válido no Wazuh, calculada como TD = (técnicas detectadas / total de técnicas executadas) × 100. Representa a cobertura efetiva das regras de detecção frente ao conjunto de TTPs avaliado.

Cobertura de Técnicas ATT&CK (CTA): proporção de técnicas do conjunto selecionado para as quais existem regras de detecção ativas e validadas no Wazuh, independentemente de terem gerado alertas no ciclo em questão.

Taxa de Falsos Positivos (TFP) e Taxa de Falsos Negativos (TFN): proporção de alertas gerados sem correspondência a uma técnica executada (falsos positivos) e de técnicas executadas que não geraram alerta (falsos negativos), respectivamente. Ambas as métricas são relevantes para a calibração da qualidade das regras de detecção.

Os resultados de ambos os ciclos serão tabulados e comparados por meio de análise descritiva e de variação percentual entre as métricas do Ciclo Baseline e do Ciclo Pós-Ajuste, permitindo quantificar os ganhos operacionais proporcionados pela integração Purple em cada dimensão avaliada.

4 RESULTADOS

4.1 Dados Obtidos nos Testes

Os experimentos descritos nesta seção foram conduzidos em junho de 2026, no laboratório virtual caracterizado na Seção 3.2. O ambiente foi composto por um servidor Wazuh (SIEM/HIDS) na versão 4.14.4, atuando como camada de detecção, e por três endpoints monitorados por agentes ativos e reportando ao servidor: duas estações Microsoft Windows 10 Pro, identificadas como Client-1 (192.168.66.130) e Client-2 (192.168.66.131), e uma estação Linux Ubuntu, identificada como joao-VirtualBox (192.168.66.128). A camada ofensiva, operada a partir do Kali Linux nas modalidades manual e por agente de IA, executou um conjunto de quinze execuções de técnicas adversariais (TTPs), correspondentes a oito técnicas distintas do framework MITRE ATT&CK, distribuídas ao longo da cadeia de ataque e abrangendo as táticas de Descoberta, Acesso a Credenciais, Persistência, Evasão de Defesa e Movimentação Lateral. Para assegurar a integridade do cálculo do Time to Detect (TTD), todos os relógios do ambiente foram sincronizados via NTP e referenciados em UTC, e o instante de disparo de cada técnica (t_ataque) foi registrado automaticamente por um script de instrumentação.

Os dados apresentados nesta seção referem-se ao Ciclo Baseline, no qual o conjunto de técnicas foi executado contra o ruleset nativo do Wazuh, sem qualquer customização de regras ou ampliação de telemetria. O propósito deste ciclo é estabelecer a linha de base de detecção da plataforma, ou seja, identificar quais técnicas são reconhecidas em sua configuração padrão e quais permanecem como lacunas de cobertura. A Tabela 2 consolida, para cada técnica executada, a tática associada, o alvo, a ocorrência ou não de detecção, a regra do Wazuh responsável pelo alerta (identificada por seu rule.id) e o respectivo TTD.

Tabela 2 – Resultados de detecção por técnica no Ciclo Baseline.

Técnica (MITRE ATT&CK)	Tática	Alvo	Detecção	Regra (rule.id)	TTD (s)
T1046 – Varredura de portas	Descoberta	Ubuntu	Não	—	—
T1110.001 – Força bruta SSH	Acesso a Cred.	Ubuntu	Sim	5503 / 5763	3
T1078 – Conta válida (SSH)	Evasão de Defesa	Ubuntu	Sim	5715	1
T1053.003 – Cron(persistência)	Persistência	Ubuntu	Não	—	—
T1070.002 – Limpeza de logs	Evasão de Defesa	Ubuntu	Não	—	—
T1046 – Varredura de portas	Descoberta	Client-1	Não	—	—
T1110.001 – Força bruta RDP	Acesso a Cred.	Client-1	Sim	60122 / 60204	3
T1078 – Conta válida (RDP)	Evasão de Defesa	Client-1	Sim	92657	8
T1110 – Força bruta (bloqueio)	Acesso a Cred.	Client-1	Sim	60204	*
T1021.001 – RDP (mov. lateral)	Mov. Lateral	Client-1	Sim	92653	*
T1046 – Varredura de portas	Descoberta	Client-2	Não	—	—
T1110 – Força bruta SMB	Acesso a Cred.	Client-2	Sim	60122 / 60204	19
T1110 – Força bruta WinRM	Acesso a Cred.	Client-2	Sim	60122	*

T1078 – Conta válida (SMB)	Evasão de Defesa	Client-2	Sim	92652	*
T1135 – Enum. de compartilh.	Descoberta	Client-2	Sim	92657	3

Legenda: "—" indica técnica não detectada (ausência de alerta); "*" indica técnica detectada cujo TTD não pôde ser isolado, em razão da sobreposição temporal com alertas de eventos correlatos (por exemplo, fluxos de força bruta encadeados ou logons originados na etapa de validação de credenciais). As regras são identificadas pelo respectivo rule.id no Wazuh.

Fonte: o autor (2026).

A partir dos resultados individuais, calcularam-se as métricas agregadas do Ciclo Baseline, consolidadas na Tabela 3, ao lado das do Ciclo Pós-ajuste, cuja comparação é retomada ao final desta seção. Das quinze execuções, dez geraram ao menos um alerta válido, resultando em Taxa de Detecção (TD) de 66,7%. Entre as técnicas detectadas, o tempo médio de detecção foi de aproximadamente 6,2 segundos, considerando-se as seis técnicas cujo TTD pôde ser medido de forma isolada, o que evidencia boa responsividade da plataforma para as técnicas efetivamente cobertas. Não se observaram alertas sem correspondência a um ataque real, de modo que a Taxa de Falsos Positivos (TFP) foi nula no período; em contrapartida, a Taxa de Falsos Negativos (TFN) alcançou 33,3%, correspondente às cinco execuções não detectadas. Cabe registrar que, no Ciclo Baseline, não foram configuradas ações de resposta automática (Active Response); por essa razão, a métrica Time to Respond (TTR) não é reportada neste ciclo, constituindo objeto de avaliação na fase de integração Purple.

Tabela 3 – Métricas agregadas do Ciclo Baseline e comparação com o Ciclo Pós-ajuste.

Métrica	Ciclo Baseline	Ciclo Pós ajuste
Taxa de Detecção (TD)	66,7%	80,0%
TTD médio (s)	6,2	≈ 6
Cobertura de Técnicas ATT&CK (CTA)	66,7%	80,0%
Taxa de Falsos Positivos (TFP)	0,0%	0,0%
Taxa de Falsos Negativos (TFN)	33,3%	20,0%

Fonte: o autor (2026).

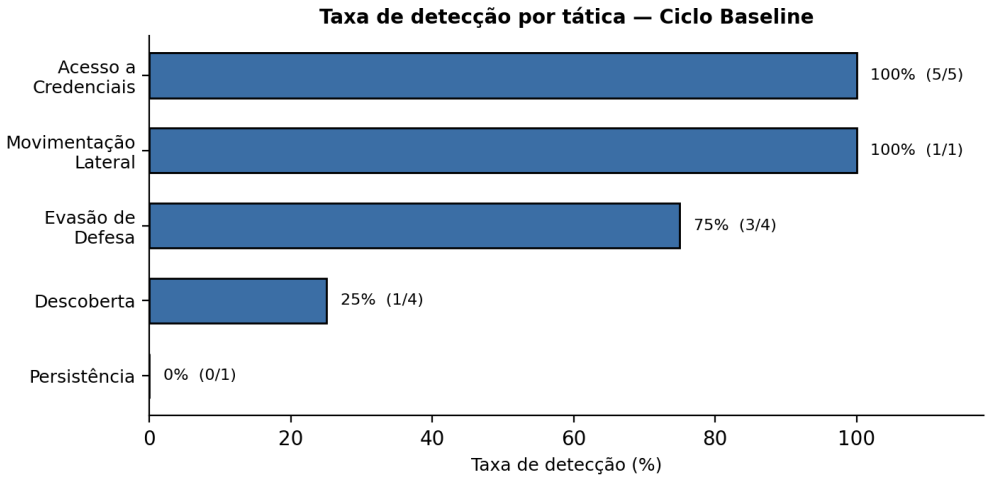
Para uma leitura consolidada do desempenho por categoria de ataque, a Tabela 4 e as Figuras 5 e 6 sintetizam a detecção segundo a tática do MITRE ATT&CK e o alvo, evidenciando em quais frentes o ruleset nativo se mostrou eficaz e em quais residem as lacunas. Observa-se que a totalidade das técnicas de Acesso a Credenciais e de Movimentação Lateral foi detectada, ao passo que as táticas de Descoberta e Persistência concentram a maior parte dos falsos negativos.

Tabela 4 – Síntese da detecção por tática no Ciclo Baseline.

Tática	Execuções	Detectadas	Não detectadas	Detecção (%)
Acesso a Credenciais	5	5	0	100,0
Movimentação Lateral	1	1	0	100,0
Evasão de Defesa	4	3	1	75,0
Descoberta	4	1	3	25,0
Persistência	1	0	1	0,0
Total	**15**	**10**	**5**	**66,7**

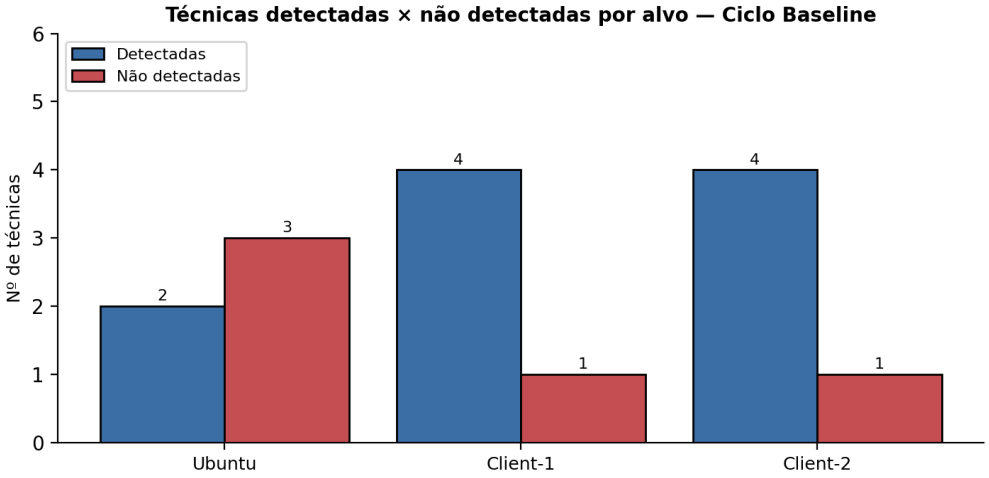
Fonte: o autor (2026).

Figura 5 – Taxa de detecção por tática no Ciclo Baseline.



Fonte: o autor (2026).

Figura 6 – Técnicas detectadas e não detectadas por alvo no Ciclo Baseline.



Fonte: o autor (2026).

A análise qualitativa dos resultados revela um padrão consistente quanto à cobertura do ruleset nativo. A plataforma demonstrou elevada eficácia na detecção de técnicas relacionadas a credenciais e autenticação: os ataques de força bruta sobre os serviços SSH, RDP, SMB e WinRM (T1110 e T1110.001), os logons bem-sucedidos com contas válidas (T1078), o movimento lateral via RDP (T1021.001) e a enumeração de compartilhamentos por sessão anônima (T1135) foram prontamente sinalizados, com destaque para a geração de regras de correlação de nível elevado, como a regra 60204 (Multiple Windows Logon Failures, nível 10) e a regra 5763 (SSHD brute force, nível 10). Em contrapartida, a plataforma mostrou-se cega a duas classes de atividade. A primeira é o reconhecimento de rede (T1046): a varredura de portas executada com a ferramenta Nmap não produziu qualquer alerta em nenhum dos três alvos. A segunda compreende as técnicas de furtividade no host Linux: a persistência por meio de tarefa agendada (T1053.003) e a evasão de defesa por limpeza de logs (T1070.002). No primeiro caso, a inserção de uma entrada maliciosa no arquivo /etc/crontab não foi sinalizada, uma vez que o módulo de monitoramento de integridade de arquivos (FIM) não contemplava esse caminho em sua configuração padrão, tendo sido registrada apenas a elevação de privilégio (sudo) associada à ação. No segundo, o truncamento do arquivo /var/log/auth.log não gerou alerta, por inexistir regra dedicada à detecção dessa operação anti-forense.

As cinco lacunas identificadas (a varredura de rede nos três endpoints, a persistência via cron e a limpeza de logs) não decorrem de uma incapacidade intrínseca da plataforma de detecção, mas da ausência de regras de correlação e de configurações de telemetria específicas no Ciclo Baseline. Constituem, precisamente, o insumo que caracteriza a operação Purple Team: a partir do conhecimento detalhado das técnicas ofensivas que escaparam à detecção, o Blue Team dispõe de informação suficiente para desenvolver e ativar novas regras e estender o

escopo do monitoramento. Esse processo de realimentação, núcleo do modelo Purple, é descrito na sequência, e seus efeitos são quantificados no Ciclo Pós-ajuste, no qual o mesmo conjunto de técnicas é reexecutado, de forma idêntica, sob o ruleset aprimorado. A comparação entre os dois ciclos constitui a evidência empírica central deste trabalho.

Concluído o diagnóstico do Ciclo Baseline, a fase de Integração Purple traduziu cada lacuna conhecida em uma ação concreta de aprimoramento da detecção, combinando a ampliação da telemetria coletada e a criação de regras de correlação customizadas no Wazuh. Para a persistência via cron (T1053.003), ativou-se o monitoramento de integridade de arquivos (FIM) em tempo real sobre o arquivo /etc/crontab e os diretórios de cron, associado a uma regra customizada de nível 10 (rule.id 100010). Para a evasão de defesa por limpeza de logs (T1070.002), habilitou-se o subsistema de auditoria do Linux (auditd) com uma regra de monitoramento (watch) sobre o arquivo /var/log/auth.log, vinculada a uma regra customizada de nível 12 (rule.id 100020). Para o reconhecimento de rede (T1046), configurou-se o registro das novas conexões TCP pelo firewall do host (iptables), encaminhado ao Wazuh, com uma regra de correlação destinada a sinalizar rajadas de conexões originadas de um mesmo endereço de origem. Essas três modificações estão sintetizadas na Tabela 5. Os artefatos correspondentes — regras customizadas, decoder e configurações de telemetria do Wazuh — são disponibilizados no Apêndice A, de modo a assegurar a reprodutibilidade do experimento.

Tabela 5 — Síntese das modificações defensivas implementadas na fase de Integração Purple.

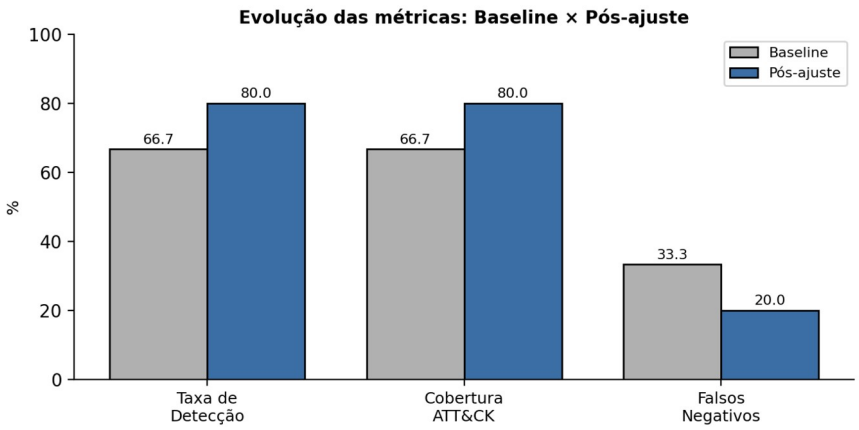
Lacuna no Baseline (Técnica MITRE ATT&CK)	**Modificação implementada na Integração Purple**	**Regra Wazuh (rule.id / nível)**	**Resultado no Pós-ajuste**
T1053.003 — Persistência via cron	Inclusão dos caminhos do cron no monitoramento de integridade de arquivos (FIM), em tempo real: /etc/crontab e diretórios de cron	100010 (nível 10)	Detectada (latência de segundos)
T1070.002 — Limpeza/truncamento de logs	Regra de auditoria (auditd) com watch sobre o log de autenticação /var/log/auth.log	100020 (nível 12)	Detectada (latência de segundos)
T1046 — Varredura de portas (Nmap)	Registro das novas conexões TCP pelo firewall do host (iptables) encaminhado ao Wazuh, com regra de correlação de rajadas	Correlação de rajadas	Não detectada (limitação de rede)

Fonte: o autor (2026).

No Ciclo Pós-ajuste, o mesmo conjunto de quinze técnicas foi reexecutado, de forma idêntica, sob o ruleset aprimorado. Das cinco lacunas identificadas no Baseline, duas foram efetivamente fechadas: a persistência via cron e a limpeza de logs passaram a ser detectadas pelas regras 100010 e 100020, respectivamente, ambas com latência da ordem de segundos. O reconhecimento de rede (T1046) permaneceu sem detecção efetiva, limitação discutida ao final desta seção. Em consequência, a Taxa de Detecção evoluiu de 66,7% para 80,0% e a Taxa de Falsos Negativos recuou de 33,3% para 20,0%, mantendo-se nula a Taxa de Falsos Positivos e preservando-se o tempo de detecção na ordem de segundos.

A Tabela 3 e a Figura 7 consolidam a comparação entre os dois ciclos. A integração Purple proporcionou um ganho de 13,3 pontos percentuais tanto na Taxa de Detecção quanto na Cobertura de Técnicas ATT&CK, com a correspondente redução dos Falsos Negativos. Esse resultado evidencia, de forma empírica, o efeito da realimentação Red→Blue característica do modelo Purple Team: a cada lacuna conhecida e corrigida, a capacidade de detecção do SOC eleva-se de maneira mensurável.

Figura 7 – Evolução das métricas entre os ciclos Baseline e Pós-ajuste.



Fonte: o autor (2026).

Por fim, registra-se uma limitação relevante. A varredura de portas (T1046) não foi detectada em nenhum dos dois ciclos. Diferentemente das demais técnicas, centradas em autenticação, persistência e manipulação de arquivos, que deixam rastros nos logs de host, o reconhecimento de rede manifesta-se predominantemente na camada de rede, de modo que sua detecção confiável tende a exigir sensores dedicados (por exemplo, sistemas de detecção de intrusão de rede como Suricata ou Zeek) integrados ao SIEM, o que extrapola o escopo baseado em host adotado neste experimento. Tal limitação não compromete o resultado central (a melhora mensurável proporcionada pela integração Purple), mas delimita o alcance do monitoramento exclusivamente baseado em host e indica uma direção natural para trabalhos futuros.

4.2 Discussão

A análise dos dados apresentados na Seção 4.1 permite interpretar, à luz dos objetivos definidos na introdução e da literatura revisada, os fatores que explicam a evolução observada nas métricas entre os ciclos Baseline e Pós-ajuste. O resultado central do experimento confirma a hipótese de que a integração estruturada entre Red Team e Blue Team, operacionalizada pelo modelo Purple Team, produz ganho mensurável na capacidade de detecção do SOC: a Taxa de Detecção evoluiu de 66,7% para 80,0% e a Taxa de Falsos Negativos recuou de 33,3% para 20,0%, sem qualquer aumento de falsos positivos. Esse ganho de 13,3 pontos percentuais não resultou da adoção de uma nova plataforma ou de maior poder computacional, mas exclusivamente do ciclo de realimentação Red→Blue: o conhecimento detalhado das técnicas que escaparam à detecção no Baseline orientou a criação de regras e a ampliação da telemetria que fecharam parte das lacunas no Pós-ajuste.

O comportamento da detecção por tática evidencia onde residem as forças e as fragilidades do ruleset nativo do Wazuh. As táticas de Acesso a Credenciais e de Movimentação Lateral foram integralmente detectadas já no Baseline, o que reflete a maturidade das regras de correlação voltadas a eventos de autenticação, como as regras 60204 (Multiple Windows Logon Failures) e 5763 (SSHD brute force), de nível elevado. Esse achado é coerente com a natureza dessas técnicas, que produzem registros abundantes e bem estruturados nos logs de host e, por isso, são naturalmente cobertas por rulesets generalistas. Em contrapartida, as táticas de Descoberta e Persistência concentraram os falsos negativos, revelando que atividades de baixo volume de log ou que dependem de telemetria não habilitada por padrão, como o monitoramento de integridade de arquivos e a auditoria do sistema, permanecem invisíveis até que sejam deliberadamente instrumentadas.

A fase de Integração Purple atuou precisamente sobre essas lacunas. A ativação do monitoramento de integridade de arquivos (FIM) sobre /etc/crontab e a habilitação do subsistema de auditoria (auditd) sobre /var/log/auth.log, associadas a regras customizadas (rule.id 100010 e 100020), converteram duas técnicas antes invisíveis, a persistência via cron (T1053.003) e a limpeza de logs (T1070.002), em eventos detectados com latência da ordem de segundos. Esse resultado oferece evidência empírica para a tese, defendida por Gaifulina (2025) e por Routin, Thoores e Rossier (2022), de que o valor do Purple Team não está na execução do ataque em si, mas no ciclo estruturado de identificação e remediação de lacunas. Confirma-se também a observação de Chindrus e Caruntu (2023) de que a ausência de comunicação estruturada entre ofensiva e defesa produz detecções tardias ou inexistentes, ao passo que a retroalimentação organizada eleva a cobertura de forma rastreável.

A estabilidade do Time to Detect, mantido na ordem de segundos em ambos os ciclos, indica que os ajustes ampliaram a cobertura sem degradar a responsividade da plataforma. Esse ponto é relevante porque o aprimoramento de regras de correlação pode, em determinados cenários, introduzir latência adicional; no experimento, contudo, as novas regras operaram sobre eventos pontuais, como a criação de arquivo e a chamada de sistema auditada, de detecção praticamente imediata. A reprodutibilidade exigida pela comparação pré e pós-

intervenção foi assegurada pela execução combinada das modalidades manual e automatizada por agente de IA a partir do Kali Linux: enquanto a operação manual aproximou os cenários do comportamento de um adversário interno, a automação garantiu que o mesmo conjunto de técnicas fosse reexecutado de forma idêntica nos dois ciclos, condição necessária para atribuir a variação das métricas aos ajustes defensivos, e não a diferenças na execução do ataque.

Os resultados devem ser interpretados à luz das limitações do estudo. A primeira é a abrangência do monitoramento, restrito à camada de host: a varredura de portas (T1046) permaneceu indetectada nos dois ciclos, pois sua assinatura manifesta-se predominantemente no tráfego de rede, cuja detecção confiável demanda sensores dedicados, como Suricata ou Zeek, integrados ao SIEM. A segunda é a não avaliação do Time to Respond, uma vez que o ciclo experimental não contemplou ações de resposta automática (Active Response). A terceira diz respeito à escala: o conjunto de quinze execuções, correspondentes a oito técnicas, e o caráter de laboratório controlado limitam a generalização estatística dos achados, tratados por análise descritiva e de variação percentual. Tais limitações não invalidam o resultado central, mas delimitam seu alcance e indicam direções para trabalhos futuros, em especial a incorporação de detecção em rede, a mensuração do TTR sob políticas de resposta automatizada e a ampliação do conjunto de TTPs avaliado.

4.3 Síntese dos Resultados

Confrontados com a questão de pesquisa formulada na introdução, os resultados dos dois ciclos experimentais respondem afirmativamente quanto à detecção e à cobertura de técnicas adversárias, delimitando de forma transparente o que o desenho adotado não permitiu mensurar. A integração Purple elevou a Taxa de Detecção e a Cobertura de Técnicas ATT&CK de 66,7% para 80,0% e reduziu a Taxa de Falsos Negativos de 33,3% para 20,0%, sem qualquer aumento de falsos positivos e com o tempo de detecção estável na ordem de segundos — o que indica que a ampliação da cobertura não foi obtida à custa de ruído operacional, fator crítico para a sustentabilidade de um SOC. A leitura por tática qualifica esse resultado: a atuação foi seletiva sobre as lacunas de maior criticidade tratáveis no escopo do experimento — a Persistência, antes sem qualquer cobertura, e a Evasão de Defesa passaram a ser detectadas, enquanto a Descoberta permaneceu como a tática menos coberta, em razão da limitação inerente ao monitoramento baseado em host para identificar varredura de rede.

A dimensão de resposta não compõe os resultados reportados, uma vez que o ciclo avaliado não contemplou ações de resposta automática (Active Response); o Time to Respond fica indicado como métrica a ser incorporada em ciclos futuros. Ainda assim, a evolução conjunta das métricas de detecção e cobertura, obtida sem incremento de falsos positivos e com latência estável, evidencia ganho de maturidade operacional do SOC entre os dois momentos experimentais (GRACIS, 2022; GAIFULINA, 2025). Por fim, os resultados validam, na prática, o conjunto de métricas operacionais proposto como contribuição deste trabalho: articuladas e rastreadas por meio da ferramenta VECTR e estruturadas pela metodologia PEIR, as métricas mostraram-se suficientes para quantificar, de modo reproduzível, o impacto da integração Purple, oferecendo às organizações um modelo aplicável para operacionalizar de forma mensurável exercícios colaborativos de segurança, alinhado, no contexto brasileiro, à execução sistemática do Controle 18 (Testes de Intrusão) do PPSI 2.0 (BRASIL, 2026; SILVA; GONDIM, 2025).

5 CONCLUSÃO

Este estudo investigou em que medida a integração estruturada entre Red Team e Blue Team, sob o modelo Purple Team, contribui para a melhoria mensurável da capacidade de detecção, resposta e cobertura de técnicas adversárias em um Security Operations Center. Para isso, propôs uma abordagem metodológica em três camadas (emulação ofensiva fundamentada no MITRE ATT&CK e operada a partir do Kali Linux, monitoramento e detecção centralizados na plataforma Wazuh, e integração Purple estruturada pela metodologia PEIR e documentada na ferramenta VECTR) e a submeteu a um experimento controlado de comparação pré e pós-intervenção. Os resultados confirmaram a hipótese central: o ciclo de realimentação entre ofensiva e defesa elevou a Taxa de Detecção de 66,7% para 80,0% e reduziu a Taxa de Falsos Negativos de 33,3% para 20,0%, sem qualquer aumento de falsos positivos e preservando o tempo de detecção na ordem de segundos.

A principal contribuição do trabalho é o conjunto estruturado de métricas operacionais (Taxa de Detecção, Cobertura de Técnicas ATT&CK, Taxa de Falsos Positivos e Negativos e Time to Detect), articulado a um protocolo reproduzível de dois ciclos que permite quantificar, de forma objetiva e rastreável, o impacto da integração Purple sobre a maturidade do SOC. Diferentemente de relatos predominantemente qualitativos ou de mercado, que caracterizam boa parte da literatura sobre o tema, o estudo oferece evidência empírica obtida em ambiente controlado e instrumentado. No contexto brasileiro, esse instrumental dialoga diretamente com o Controle 18 (Testes de Intrusão) do PPSI 2.0, ao prover uma forma sistemática e mensurável de operacionalizar e validar os ciclos de teste e remediação previstos na norma.

O estudo apresenta limitações que delimitam o alcance de suas conclusões. O monitoramento restringiu-se à camada de host, o que impediu a detecção da varredura de rede (T1046) em ambos os ciclos e indica que

cenários de reconhecimento exigem sensores de rede dedicados integrados ao SIEM. A dimensão de resposta não foi avaliada, uma vez que o ciclo experimental não contemplou ações automáticas, deixando o Time to Respond fora dos resultados reportados. Por fim, o conjunto de quinze execuções correspondentes a oito técnicas, conduzido em laboratório controlado, restringe a generalização estatística dos achados, tratados por análise descritiva e de variação percentual.

Como perspectivas para trabalhos futuros, recomenda-se a incorporação de sensores de detecção em rede, como Suricata ou Zeek, para cobrir as técnicas que se manifestam fora dos logs de host; a configuração de mecanismos de resposta automática (Active Response) que viabilizem a mensuração do Time to Respond; e a ampliação do conjunto de TTPs e do número de execuções, de modo a permitir inferência estatística mais robusta. Sugere-se ainda investigar de forma comparativa o desempenho das modalidades de execução manual e automatizada por agente de inteligência artificial, bem como replicar o protocolo em ambientes de maior escala e heterogeneidade, aproximando-o das condições reais de operação de um SOC corporativo. Tais extensões podem consolidar o modelo proposto como referência metodológica para a avaliação empírica de exercícios Purple Team.

REFERÊNCIAS

BALADARI, V. Red Team vs. Blue Team: Assessing Cybersecurity Resilience Through Simulated Attacks. European Journal of Advances in Engineering and Technology, v. 8, n. 4, p. 82-87, 2021.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. Guia do Framework de Privacidade e Segurança da Informação (PPSI 2.0). Versão 1.2. Brasília: SGD/MGI, 2026. Disponível em: https://www.gov.br/governodigital/pt-br/privacidade-e-seguranca/ppsi-2.0. Acesso em: abr. 2026.

CHINDRUS, C.; CARUNTU, C.-F. Securing the Network: A Red and Blue Cybersecurity Competition Case Study. Information, v. 14, n. 11, p. 587, 2023. Disponível em: https://doi.org/10.3390/info14110587.

GAIFULINA, A. The Concept of Red Team and Blue Team Synergy as a Factor in Enhancing an Organization's Resilience to Cyberattacks. The American Journal of Applied Sciences, v. 7, n. 11, p. 55-60, 2025. Disponível em: https://doi.org/10.37547/tajas/Volume07Issue11-06.

GRACIS, R. Next Generation SOC: Automated Operations and Machine Learning in Cybersecurity. 2022. Dissertação (Mestrado em Engenharia da Computação) – Politecnico di Torino, Turim, 2022.

MITRE CORPORATION. MITRE ATT&CK: Adversarial Tactics, Techniques, and Common Knowledge. [S. l.]: MITRE, 2025. Disponível em: https://attack.mitre.org. Acesso em: abr. 2026.

OWASP FOUNDATION. OWASP Top Ten 2025: The Ten Most Critical Web Application Security Risks. [S. l.]: OWASP Foundation, 2025. Disponível em: https://owasp.org/Top10/2025/. Acesso em: abr. 2026.

ROUTIN, D.; THOORES, S.; ROSSIER, S. Purple Team Strategies: Enhancing global security posture through uniting red and blue teams with adversary emulation. Birmingham: Packt Publishing, 2022.

SCARFONE, K.; SOUPPAYA, M.; CODY, A.; OREBAUGH, A. Technical Guide to Information Security Testing and Assessment. Gaithersburg: National Institute of Standards and Technology, 2008. (NIST Special Publication 800-115). Disponível em: https://doi.org/10.6028/NIST.SP.800-115.

SILVA, A. C. F. P.; GONDIM, J. J. C. Validação da Metodologia DOTA Integrada às TTPs: Uma Proposta para a Execução do Controle 18 do PPSI (Pentest). Programa de Pós-Graduação Profissional em Engenharia Elétrica – PPEE, Universidade de Brasília, Brasília, 2025.

VAISHNAVI, S.; ANANYA, S.; AKILESH, M. S. The Importance of Red Team Exercises in VAPT for Proactive Cybersecurity. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), v. 4, n. 6, p. 639-644, abr. 2024. Disponível em: https://doi.org/10.48175/IJARSCT-17687.

YAO, S. et al. ReAct: Synergizing Reasoning and Acting in Language Models. In: International Conference on Learning Representations (ICLR), 2023. Disponível em: https://doi.org/10.48550/arXiv.2210.03629. Acesso em: jul. 2026.

APÊNDICE A — ARTEFATOS DE DETECÇÃO DA FASE DE INTEGRAÇÃO PURPLE

Para assegurar a reprodutibilidade do experimento, apresentam-se a seguir os artefatos de detecção desenvolvidos na fase de Integração Purple e aplicados no Ciclo Pós-ajuste. Os identificadores de regra (rule.id) correspondem aos referenciados na Seção 4 e na Tabela 5. As regras e o decoder foram instalados no gerenciador Wazuh; as configurações de telemetria (FIM e auditoria), no agente do alvo Linux (Ubuntu). A totalidade dos artefatos — regras, decoders e configurações completas do Wazuh — está também disponível publicamente no repositório: https://github.com/sgtbrunorodrigues/purple-team-effectiveness.

A.1 Regras customizadas (local_rules.xml)

```
<group name="purple,local,">
  <!-- T1053.003 - Persistencia via Cron -->
  <rule id="100010" level="10">
    <if_group>syscheck</if_group>
    <field name="file">/etc/cron|/var/spool/cron</field>
    <description>Alteracao em arquivo de cron - possivel persistencia
[T1053.003]</description>
    <mitre><id>T1053.003</id></mitre>
    <group>persistence,attack,</group>
  </rule>
  <!-- T1070.002 - Limpeza de logs (auditd, key=logtamper) -->
  <rule id="100020" level="12">
    <if_group>audit</if_group>
    <field name="audit.key">logtamper</field>
    <description>Acesso de escrita ao /var/log/auth.log - possivel limpeza de logs
[T1070.002]</description>
    <mitre><id>T1070.002</id></mitre>
    <group>defense_evasion,attack,</group>
  </rule>
  <!-- T1046 - Varredura de portas (correlacao: 15+ conexoes em 10s) -->
  <rule id="100032" level="5">
    <if_sid>4100</if_sid>
    <description>Conexao de firewall registrada (netfilter via 4100)</description>
    <group>firewall,</group>
  </rule>
  <rule id="100030" level="10" frequency="15" timeframe="10">
    <if_matched_sid>100032</if_matched_sid>
    <description>Possivel varredura de portas (port scan) [T1046]</description>
    <mitre><id>T1046</id></mitre>
    <group>recon,attack,</group>
  </rule>
</group>
```

A.2 Decoder auxiliar para reconhecimento de rede (local_decoder.xml)

```
<decoder name="netfilter-new">
  <parent>kernel</parent>
  <prematch>NETFILTER-NEW:</prematch>
  <regex>SRC=(\d+\.\d+\.\d+\.\d+)</regex>
  <order>srcip</order>
</decoder>
```

A.3 Telemetria nos alvos (FIM/syscheck e auditoria)

```
<!-- FIM em tempo real sobre os caminhos de cron -->
<syscheck>
  <directories realtime="yes" check_all="yes"
report_changes="yes">/etc/crontab</directories>
  <directories realtime="yes" check_all="yes"
report_changes="yes">/etc/cron.d</directories>
  <directories realtime="yes"
check_all="yes">/etc/cron.daily,/etc/cron.hourly,/etc/cron.weekly,/etc/
cron.monthly</directories>
  <directories realtime="yes" check_all="yes">/var/spool/cron</directories>
</syscheck>
<!-- Coleta do log de auditoria e do kernel (netfilter/iptables) -->
<localfile><log_format>audit</log_format><location>/var/log/audit/audit.log</
location></localfile>
<localfile><log_format>syslog</log_format><location>/var/log/kern.log</location></
localfile>
```

A regra de auditoria (auditd) que alimenta a detecção da limpeza de logs (rule.id 100020) é: -w /var/log/auth.log -p wa -k logtamper.