2025年度未踏アドバンスト事業

AIエージェントによる自動ペネトレーションテストシステムの開発

ー　網羅的なペンテストのフルオートメーション化　ー

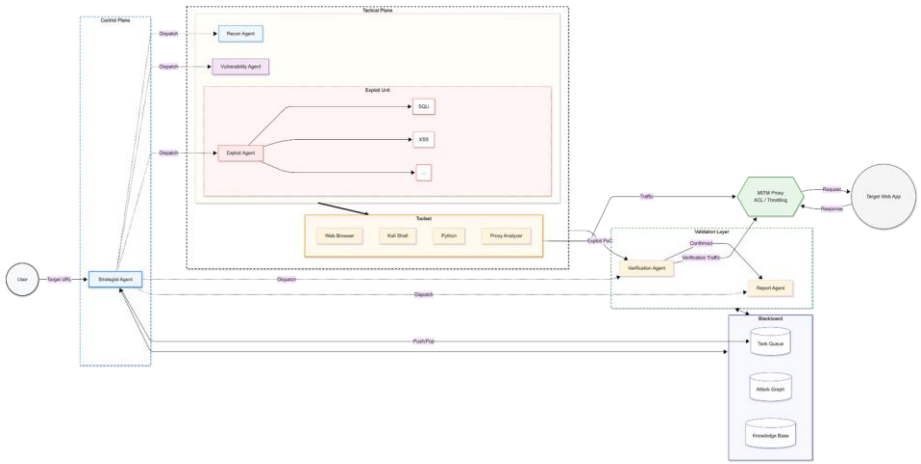
岡本 拓将　阿部 竜也

背景・目的

従来のペンテストは専門家の手作業に依存し、コスト、網羅性の面でスケールしづらい。この人間による律速を AI Agent で打破することにより、セキュリティチェックのシフトレフトや頻度向上を目指す。

そこで、
・ブラックボックス/グレーボックステストのフルオートメーション化
・スケーラブル&網羅的な検証
・ビジネスインパクトの自動評価
をコアコンセプトとしてペンテスト AI Agent を開発した。

アーキテクチャ



実証実績

フルオートメーションで実行した結果、下記結果を得られた。
実稼働システムに対して脆弱性の悪用リスクが立証されたため、忠実度の高いペンテストを実現できたと言える。

・PortSwigger Academy Lab：84%
　・制限時間：1h
・HackerOne の VDP 部門（90 Days）で世界86位
　・VDP：米国防総省の運用資産上で3件発見し、トリアージされた。
　・BBP：3件発見（既報告のためポイントなし）

技術的ブレイクスルー

★ 文脈理解
従来のスキャナでは、Logic Flaws や Race Condition などの自動検知は極めて困難だった。アプリケーション特性に応じた論理的プランニングにより、多角的な攻撃が可能になる。

★ スケーラブル&網羅的な検証
ペンテストの全工程を独立したエージェントが並列に処理することで、人間が数週間かけて行うペンテストを数時間単位に短縮可能になる。

★ ビジネスインパクトの自動評価
PoC の自動生成 → 実行により、検出された脆弱性の悪用リスクを客観的に評価可能になる。