Security Operations Triage

Max Jones - 2305556

Goal

- Triage thousands of alerts a day
  - Investigate each alert separating real incidents from the noise
  - Prioritise by risk and impacts
  - Contain confirmed threats fast, under human control
  - Success metrics: mean time to triage

Agents and Roles

- Orchestrator: takes in the alerts, removes duplicates, sends each on to the right agent, tracks each case.
- Enrichment Agent: adds context such as IP, user details, asset criticality.
- Investigation Agent: follows the evidence and builds an incident timeline.
- Prioritisation Agent: gives each case a risk score and prioritises them
- Containment Agent: runs basic actions such as isolation, disabling an account or IOC
- Reporting Agent: writes a case summary, updates tickets and produces shift handover reports

Tools

- Orchestrator – SIEM(Sentinel) alert feed
- Enrichment – Threat intel (VirusTotal, AbuseIPDB)
- Investigation – SIEM log search
- Prioritisation – Risk scoring model
- Containment – EDR - isolate host, kill process (CrowdStrike)
- Reporting – ServiceNow, Jira, Slack (Ticketing)

Memory

- Short-Term: Case memory – the current alert, evidence gathered and reasoning steps.
- Long-tern: Knowledge base: past incidents and their verdicts, known false-positive patterns and the asset inventory
- Procedural memory: SOC playbooks and runbooks retrieved through RAG
- Feedback memory: analyst corrections stored so a misjudgement will not be made again
- Audit log: every decision, tool call and action recorded as an immutable trail for compliance and review

Human Checkpoint

- Auto-close the low-risk alerts above a certain confidence level with a sample to be reviewed for QA
- Human approval required for some actions
  - Containment on critical assets
  - Disabling privileged accounts
  - Any action that can disrupt the business
  - Kill Switch to pause any agent or roll back any action
- Escalate to an analysts when confidence is low or evidence conflicts.
- Kill Switch to pause any agent or roll back any action
- Feedback loop so an analysts decisions feed into memory for the agents to improve over time