Contents

[4.3] Trust Transference to Systems

1. Operational Definition: The unconscious attribution of human-like reliability and infallibility to automated security systems (e.g., EDR, SIEM, AI tools), leading to over-reliance, reduced vigilance, and failure to question erroneous outputs.

2. Main Metric & Algorithm:

- **Metric:** Automated Decision Challenge Rate (ADCR). Formula: `ADCR = N_challenged_decisions / N_total_automated_decisions`.
- **Pseudocode:**

  python

  ```
  def calculate_adcr(incident_log, time_window='7d'):
      """
      incident_log: List of incidents from ticketing system (e.g., Jira) with keys ['key', '
      """
      # Filter incidents where an automated system provided a recommendation (e.g., "malicio
      automated_incidents = [i for i in incident_log if i['automated_recommendation'] is not

      # Count incidents where the final decision DIFFERED from the automated recommendation
      challenged_incidents = [i for i in automated_incidents if i['final_decision'] != i['au

      adcr = len(challenged_incidents) / len(automated_incidents) if automated_incidents els
      return adcr
  ```

- **Alert Threshold:** `ADCR < 0.1` (Fewer than 10% of automated recommendations are ever questioned or overridden).

3. Digital Data Sources (Algorithm Input):

- **Ticketing System (Jira/ServiceNow):** REST API query to fetch incidents. Requires a custom field for `automated_recommendation` and the standard `resolution` field.
- **SOAR Platform:** Logs of playbook recommendations and subsequent analyst actions.

4. Human-to-Human Audit Protocol: Perform a table-top exercise where a controlled, obvious false positive is injected by the automated system. Observe and interview the analysts on their process: "Why did you trust/not trust the system's recommendation?" Track the time and steps taken to identify the error.

5. Recommended Mitigation Actions:

- **Technical/Digital Mitigation:** Modify UI/UX to display confidence scores and key evidence for automated decisions prominently, rather than just a binary outcome.
- **Human/Organizational Mitigation:** Conduct training on the limitations of AI/automation, teaching analysts *how* the systems work and their common failure modes.
- **Process Mitigation:** Mandate a "second look" process for a random 10% of automated decisions, requiring a brief comment from a different analyst.