Contents

[8.1] Shadow projection onto attackers

1. Operational Definition: A psychological defense mechanism where an individual or group attributes their own unacceptable or negative impulses, intentions, and traits (their "Shadow") exclusively onto external attackers, creating a blind spot to internal threats and fostering an "us vs. them" mentality that weakens holistic defense.

2. Main Metric & Algorithm:

- **Metric:** Projection-to-Internal-Threat Ratio (PITR). Formula: `PITR = (Mentions of external attribution in incident reports) / (Mentions of internal factors in incident reports)`.
- **Pseudocode:**

  python

  ```
  def calculate_pitr(incident_reports, start_date, end_date):
      """
      incident_reports: Collection of post-incident report documents
      """
      # 1. NLP Keyword Lists
      external_keywords = ['hacker', 'nation-state', 'external', 'cybercriminal', 'them', 'c
      internal_keywords = ['misconfiguration', 'error', 'employee', 'insider', 'process', 't

      total_external_mentions = 0
      total_internal_mentions = 0

      for report in incident_reports:
          if start_date <= report.date <= end_date:
              text = report.title + " " + report.root_cause_analysis
              # 2. Count keyword occurrences (simple bag-of-words approach)
              external_count = count_keywords(text, external_keywords)
              internal_count = count_keywords(text, internal_keywords)

              total_external_mentions += external_count
              total_internal_mentions += internal_count

      # 3. Calculate PITR, avoid division by zero
      if total_internal_mentions > 0:
          PITR = total_external_mentions / total_internal_mentions
      else:
          PITR = total_external_mentions  # If no internal factors are cited, ratio is high
      return PITR
  ```

- **Alert Threshold:** `PITR > 3.0` (External threats are mentioned 3x more often than internal factors in root cause analysis)

3. Digital Data Sources (Algorithm Input):

- **Incident Management Platform (Jira, ServiceNow):** Fields: `post_mortem_report`, `root_cause_analysis`, `title`.
- **NLP Text Processing:** A simple keyword counter or a more sophisticated sentiment/theme analysis model trained to identify attribution language.

4. Human-to-Human Audit Protocol: A neutral third party (e.g., an auditor or external consultant) reviews a sample of incident reports and conducts interviews with the report authors. They use structured questioning: "What evidence led you to conclude the threat was primarily external? How thoroughly did you investigate potential internal contributing factors?" The goal is to identify assumptions and unchallenged narratives.

5. Recommended Mitigation Actions:

- **Technical/Digital Mitigation:** Configure the incident management system to require a mandatory field in the root cause analysis that must list at least one potential internal process or human factor that could have contributed to the incident's impact, even if external.
- **Human/Organizational Mitigation:** Conduct training on cognitive biases and defense mechanisms for the incident response team, specifically focusing on Jung's concept of the Shadow and how it manifests in cybersecurity.
- **Process Mitigation:** Integrate a mandatory "Red Team" or "Devil's Advocate" step into the incident review process, where a designated individual's role is to argue against the prevailing narrative and propose alternative, internal-cause hypotheses.