Contents

[3.1] Reciprocity exploitation

1. Operational Definition: The vulnerability of personnel to social engineering attacks that leverage the psychological principle of reciprocity, where a small favor or gesture from an attacker creates an unconscious obligation to comply with a subsequent malicious request.

2. Main Metric & Algorithm:

- **Metric:** Reciprocity Request Acceptance Rate (RRAR). Formula: `RRAR = (Number of accepted requests preceded by a "favor") / (Total number of requests preceded by a "favor")`.
- **Pseudocode:**

  python

  ```
  def calculate_rrar(email_logs, im_logs, access_logs, start_date, end_date):
      """
      email_logs, im_logs: Communication data
      access_logs: To check if a request was actioned
      """
      # 1. Identify communication threads containing a "favor" (e.g., "I helped you with X",
      threads_with_favors = find_threads_with_favor_keywords(email_logs, im_logs, start_date

      # 2. Extract the subsequent request from the same thread
      requests_after_favor = extract_subsequent_requests(threads_with_favors)

      # 3. Cross-reference with access/logs to see if the request was fulfilled
      accepted_requests = 0
      for req in requests_after_favor:
          if was_request_granted(req, access_logs):
              accepted_requests += 1

      # 4. Calculate RRAR
      total_requests = len(requests_after_favor)
      RRAR = accepted_requests / total_requests if total_requests > 0 else 0
      return RRAR
  ```

- **Alert Threshold:** `RRAR > 0.1` (Over 10% of requests following a perceived favor are granted)

3. Digital Data Sources (Algorithm Input):

- **Email/Instant Messaging APIs (MS Graph, Slack):** To scan for keywords related to favors and requests (`"I sent you"`, `"you can help me"`, `"as a thank you"`, `"could you please"`).
- **IAM/Access Logs & Ticketing Systems:** To determine if a request mentioned in communication was actually actioned.

4. Human-to-Human Audit Protocol: Integrate scenarios into security awareness training that test for reciprocity bias. For example, in a simulated phishing exercise, first provide a useful piece of genuine information or "help" to the target, then follow up with a malicious request. Measure the click-through/compliance rate compared to a control group that did not receive the initial favor.

5. Recommended Mitigation Actions:

- **Technical/Digital Mitigation:** Deploy email security gateways configured with advanced threat protection rules to flag external emails that contain both favor-related language and subsequent requests for action.
- **Human/Organizational Mitigation:** Train staff to recognize the reciprocity principle explicitly. Teach them to decouple the favor from the request: "Thank you for the information. I will process your request through the standard ticketing system for validation."
- **Process Mitigation:** Enforce a strict process that all access requests, especially those originating from informal channels, must be validated via the official ticketing system, regardless of the relationship or context.