

*Prepared by Matthew Shonman, Johns Hopkins University Information Security Institute*

```
simulation( )
    // Loops through all 40 emails. For every email, matches each cue to closest
    // instance in memory
    // Parameters:
    //     NUM_RUNS: number of simulation runs
    //     D: decay value for base-level term in ACT-R activation equation. Default value =
    //     0.5.
    //     THRESHOLD: when suspicion level rises above this value, the email is marked
    //     as "Phish"
    //     CUE_CUTOFF: when this number of cues in an email are scored, email analysis
    //     stops even if suspicion level is below the threshold. Default value = 13 (all cues).
    //     If the threshold has not been reached, score the email based on whether the
    //     majority of scored cues are threat or normal.
    //     NUM_LINKS: number of hyperlinks per email to check before stopping.
    //     S: ACT-R noise function parameter. Default value = 0.25.

    EMAIL_NUM = 1 // the timekeeping unit is one email.
    FOR email IN email_list:
        SUSPICION_LVL = 0
        CUE_NUM = 0
        NUM_THREATS = 0
        NUM_NON_THREATS = 0
        FOR cue IN email:
            CUE_NUM += 1
            activated_memory = score_memories(cue) // returns memory instance
                                                    // with highest activation score
                                                    // relative to current cue

            IF activated_memory[utility] == THREAT:
                SUSPICION_LVL += 1
                NUM_THREATS += 1
            ELSE:
                NUM_NON_THREATS += 1

        IF SUSPICION_LVL >= THRESHOLD:
            Set email[decision] as PHISH
            Break

        IF CUE_NUM > CUE_CUTOFF:
            IF NUM_THREATS > NUM_NON_THREATS:
                Set email[decision] as PHISH
            ELSE:
```

```

        Set email[decision] as NON-PHISH
        Break
    IF cue_type == "Hyperlink":
        // Loop through all hyperlinks for given email, until NUM_LINKS
        // hyperlinks have been processed. Ignore remaining hyperlinks.

```

```

        Set email[decision] as NON-PHISH // If all cues are scored and not enough are
        // threats
        EMAIL_NUM += 1 // for timekeeping

```

```

score_memories(CUE):
    // Re-scores all memory instances using ACT-R activation equation
    // Returns instance with highest score (if multiple instances have highest score,
    // one is chosen at random)
    // Parameters:
    //     CUE: current cue under evaluation

    FOR instance IN memory:
        B = ln[ (current_time - 1st_activation_time)^-D + (current_time -
        2nd_activation_time)^-d + ... + (current_time - most_recent_activation_time)^-d ]
        SIM:
            If current cue attribute score == current chunk attribute score:
                SIM == 1 * similarity weight
            Else:
                SIM == 0
        n = random[0, 1]
        E = S * ln((1-n)/n) // S is a parameter with default value 0.25
        Instance[score] = B + SIM + E

    Select instance with max(score)
    Add EMAIL_NUM to activation history for selected instance
    RETURN instance

```

---

### **Additional Descriptions**

#### Emails

- Array of 40 emails
- Each email is an array of cues
- Each cue contains:
  - Cue type (1-12)

- Ground truth (true score of attribute, 0-1) // User cannot access this
- Attribute (derived from content score, 0-1) // Analyst does access this
- Utility

#### Long-term memory

- Array of memory **chunks**
- Each chunk contains:
  - Cue type
  - Attribute score (0-1)
    - Hyperlink has 2 attributes; all others have 1
  - Utility (0-1)
  - Time (list of previous activation times; updated whenever chunk is activated)

#### Definitions:

- **Attribute**: is the chunk in question suspicious (1) or not suspicious (0)?
- **Utility**: in memory, was the original email associated with this chunk classified as phish (1) or non-phish (0)?