Session Summaries Panel

IFIP Working Group 10.4

June 28th, 2026

Session 3 Behavior Biometrics and Human Factors

- Chair: Fuqun Huang; Rapporteur: Hui Lin

- Presentation 1: Behavior-aligned Mouse Dynamics Segmentation
  - Daqing Hou, Rochester Institute of Technology

- Presentation 2: Dear AI: Handle this. Biometrics in the New Age
  - Roy Maxion, Carnegie Mellon University

Behavior-aligned Mouse Dynamics Segmentation

- Overarching problem
  - Mouse dynamics including movement and clicks are widely used for authentication
  - Segmenting mouse dynamic events into samples provides contexts of sequences, producing more accurate authentication
- Types of mouse actions
  - Point-click
  - Pure movement: series of movement without click but with a specific intent
  - Compound actions: a combination of pure movements
- Existing approaches, not effective to consider pure movement
  - Action-driven: use specific actions as segmentation boundary
  - Time difference threshold: use long enough pausing to segment mouse activities
  - Static window: a fixed number of actions into a segmentation

Behavior-aligned Mouse Dynamics Segmentation

- Behavior-aligned segmentation: apply Fitts' law to segment mouse dynamic actions
  - Fitts' law: foundational HCI principle that describes the structure of hand movement in target acquisition tasks; originally used in military for cognitive evaluation
- Method outline:
  - Extract mouse action features
  - Calculate mouse movement speeds in two dimensional space
  - Two thresholds as design parameters
    - minima threshold: signaling low speed actions
    - peak threshold: signaling ballistic sub-movement
  - Use the combination of ballistic peaks and low speed signal to segment mouse dynamic movement
    - E.g., lookahead window of 500 ms to find ballistic peak or minima split point
    - Enhancement: use adaptive threshold and lookahead window

Behavior-aligned Mouse Dynamics Segmentation

- Dataset
  - Balabit: from remote server
  - ChaoChen: unconstrained computing tasks
  - DFL: personal devices
- Evaluation
  - Equal error rate (EER) of single-sample result
  - Score level fusion, providing the average model output scores of N-consecutive samples
  - Impact of neural network infrastructure
- Findings:
  - Slightly larger EER due to less events in samples in single-sample evaluation
  - Significant improvement in score-level fusion
  - Behavior-aligned segmentation is especially beneficial over early detection regime; when detection misuse of a system as soon as possible is desirable.

Dear AI: Handle this. Biometrics in the New Age

- Overarching problem
  - Should we trust or be skeptical of AI in science?
- Study 1: how does AI help in keystroke research?
- Study 2: how can AI aide trial defense?

Study 1: how does AI help in keystroke research?

- Keystroke biometrics
  - Use keystroke dynamics, e.g., holding time (measured between the key press and key release), to identify users' typing rhythms
  - An important behavior biometrics that can identify specific user
  - Used in two-factor authentication, insider threat, deception, etc.
- Keystroke experiments: a same user perform keystroke experiments with the same dataset under two different scenarios
  - In a controlled Lab environment, using USB keyboard.
  - In a uncontrolled Field environment, using PS/2 keyboard.
  - Very two different scatterplots
- AI experiments: generate the same prompts for three LLM models, i.e., Gemini Pro, ChatGPT, and Gemini Research, to search for the potential reasons why two different scatterplots are obtained
  - Use domain expertise to verify the responses from the LLM models.

Study 1: how does AI help in keystroke research?

- Gemini Pro suggestions: keyboard polling rates, software interrupts/latency
  - Can provide high-level suggestions.
  - Responses are generally superficial knowledge; should not take the advises.
  - Response from OS experts: should refer to USB specification.
- ChatGPT suggestions: keyboard scan rate, browser timing API, deboucing logic in keyboard, and keyboard hardware/firmware differences.
  - Responses are generally misguiding.
  - Require some knowledge of computer architecture, OS, firmware to better understand the scope of the problem.
- Gemini Deep Research: provides research paragraphs or results without providing citations
  - Responses are generally correct but not related.
  - Aims to make the technical contents easier to interpret
- LLM model can provide some hints, but not direct insightful suggestions or solutions.

Study 1: how does AI help in keystroke research?

- Further Investigation
  - Descriptive statistics of hold time (N-number summaries)
  - Histogram of hold time
  - Frequency table
  - Time-series data
- Our observation: likely caused by differences in apparatus and change of working environments (e.g., working in the lab vs at home)
  - USB keyboard: polling interval at the power of 2. Most frequent polling can be configured at 8ms
  - PS/2 keyboard: faster response time than USB, at 1ms frequency.

Study 2: how can AI aide trial defense?

- Court case:
  - Someone was accused of attempted production of child porn
  - FBI uses a forensic software to map some pictures to certain Kodak camera; FBI claim the software has 1 in million error rate
- Use ChatGPT to identify methodological flaws, according information obtained through Goljan09 seminal paper and PACER (Public Access to Court Electronic Records)
  - Answers are vaguely correct. Emphasize one technical issue as the most significant one, but it is just one of the important issues.
  - Cross-check with Kodak to further validate the results

Study 2: how can AI aide trial defense?

- ChatGPT suggests comparing metadata in unallocated space, which is not considered in trial
  - A good suggestion to search more evidences
- Photography tests: set piece images, adjust lighting conditions to see whether it can affect judgements
  - Use # of reference images vs PCE score (for camera fingerprinting)
  - Trial testimony: 1 in millions
  - Test experiment: 1 in 8, leading to reasonable doubts