CS259D: Data Mining for Cyber Security
Re-authentication: Practical requirements

- Accuracy
- Quick response
- Difficult to forge
Data

- Controllable set
  - 30 controllable users

- Field set
  - 1000 real field users

- Raw data: <ACTION-TYPE, t, x, y>
  - ACTION-TYPE: mouse-move, mouse click

- Data preprocessing
  - Identify every point-and-click action
    - Continuous mouse movement followed by click
• **Direction**
  ◦ For consecutive points A, B: angle between line AB and horizontal line

• **Angle of Curvature**
  ◦ For any three consecutive points A, B, C: angle between AB and BC

• **Curvature Distance**
  ◦ For any three consecutive points A, B, C: ratio between length of AC to length of perpendicular distance from B to AC
Metrics
Mouse movement characterization

• Dependence on different platforms
  ◦ OS, screen size & resolution, mouse pointer sensitivity, brand of mouse, desk space available near mousepad
  ◦ Affects measurements such as speed, acceleration

• Uniqueness of angle-based metrics across users
Mouse movement characterization

![Graph of CDF vs Direction Angle and Speed](image)
Distance between distributions

- Binned PDFs: \( \{p_1, p_2, \ldots, p_n\}, \{q_1, q_2, \ldots, q_n\} \)
- Distance:

\[
\sum_{i} |p_i - q_i|
\]
Distance between distributions

<table>
<thead>
<tr>
<th>Setting</th>
<th>Machine Type</th>
<th>Mouse Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-A</td>
<td>Dell Precision T3500</td>
<td>Dell MOC5UO Two-Button Scroll-Wheel</td>
</tr>
<tr>
<td>1-B</td>
<td>Apple Macbook MB990LL/A</td>
<td>Apple A1152 One-Button Trackball</td>
</tr>
<tr>
<td>1-C</td>
<td>Apple Macbook MB990LL/A</td>
<td>Dell MOC5UO Two-Button Scroll-Wheel</td>
</tr>
</tbody>
</table>
Classifier

- 2-class SVM
- RBF kernel

Decision:
- Threshold
- Majority vote
  - Multiple models using sampled data
Results: discrimination in same environment

- 500 training blocks, threshold 0.5, 3/5 majority

<table>
<thead>
<tr>
<th>Number of Clicks</th>
<th>FRR</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.57%</td>
<td>18.79%</td>
</tr>
<tr>
<td>3</td>
<td>2.59%</td>
<td>10.81%</td>
</tr>
<tr>
<td>5</td>
<td>2.02%</td>
<td>7.67%</td>
</tr>
<tr>
<td>10</td>
<td>1.27%</td>
<td>5.23%</td>
</tr>
<tr>
<td>15</td>
<td>1.03%</td>
<td>3.13%</td>
</tr>
<tr>
<td>20</td>
<td>0.70%</td>
<td>3.32%</td>
</tr>
<tr>
<td>25</td>
<td>0.86%</td>
<td>2.96%</td>
</tr>
</tbody>
</table>
Results: discrimination in different environments

- Train model on data from a work desktop
- Test on data from a home laptop
Results: Partial movements

- Continuous mouse movements without ending in a click
- Compare to point-and-clicks
  - More noisy
  - Much more frequent
    - 0.53 mouse clicks per minute
    - 6.58 partial movements per minute

<table>
<thead>
<tr>
<th></th>
<th>Equal Error Rate</th>
<th>Verification time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point-and-click</td>
<td>1.3%</td>
<td>38 minutes</td>
</tr>
<tr>
<td>Partial movement</td>
<td>1.9%</td>
<td>3 minutes</td>
</tr>
</tbody>
</table>
Research problem

- Angel-based metrics + frequent patterns
Administrativia

- No class on Thursday
  - Work on homework instead 😊
Entry-point authentication on mobile devices

- **Usability**
  - Inconvenient for quick activities

- **Security**
  - Short passwords
  - Increased screen lock time-outs
  - Disable unlock
  - Higher risk of theft
Trigger actions

- Sliding horizontally over the screen
  - Browse through images
  - Navigate to next page of icons
- Sliding vertically over the screen
  - Reading email, documents, webpages
  - Browsing menus
Data acquisition

- Android phones
- Tasks: read documents, compare images
- Raw features:
  - Event code (e.g., finger up, finger down, finger move, multi-touch)
  - Event time
  - Device orientation
  - x, y coordinates of finger
  - Finger pressure
  - Area on the screen covered by the finger
  - Finger orientation with respect to screen orientation
Stroke

- Sequence of touch data starting with touching the screen, ending with lifting the finger
- Sequence of vectors:
  - $s_n = (x_n, y_n, t_n, p_n, A_n, o_{n^f}, o_{n^ph}); (1 \leq n \leq N)$
# 30 Features

<table>
<thead>
<tr>
<th>Rel. mutual information</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.58%</td>
<td>mid-stroke area covered</td>
</tr>
<tr>
<td>19.63%</td>
<td>20%-perc. pairwise velocity</td>
</tr>
<tr>
<td>17.28%</td>
<td>mid-stroke pressure</td>
</tr>
<tr>
<td>11.06%</td>
<td>direction of end-to-end line</td>
</tr>
<tr>
<td>10.32%</td>
<td>stop $x$</td>
</tr>
<tr>
<td>10.15%</td>
<td>start $x$</td>
</tr>
<tr>
<td>9.45%</td>
<td>average direction</td>
</tr>
<tr>
<td>9.43%</td>
<td>start $y$</td>
</tr>
<tr>
<td>8.84%</td>
<td>average velocity</td>
</tr>
<tr>
<td>8.61%</td>
<td>stop $y$</td>
</tr>
<tr>
<td>8.5%</td>
<td>stroke duration</td>
</tr>
<tr>
<td>8.27%</td>
<td>direct end-to-end distance</td>
</tr>
<tr>
<td>8.16%</td>
<td>length of trajectory</td>
</tr>
<tr>
<td>7.85%</td>
<td>80%-perc. pairwise velocity</td>
</tr>
<tr>
<td>7.24%</td>
<td>median velocity at last 3 pts</td>
</tr>
<tr>
<td>7.22%</td>
<td>50%-perc. pairwise velocity</td>
</tr>
<tr>
<td>7.07%</td>
<td>20%-perc. pairwise acc</td>
</tr>
<tr>
<td>6.29%</td>
<td>ratio end-to-end dist and length of trajectory</td>
</tr>
<tr>
<td>6.08%</td>
<td>largest deviation from end-to-end line</td>
</tr>
<tr>
<td>5.96%</td>
<td>80%-perc. pairwise acc</td>
</tr>
<tr>
<td>5.82%</td>
<td>mean resultant length</td>
</tr>
<tr>
<td>5.42%</td>
<td>median acceleration at first 5 points</td>
</tr>
<tr>
<td>5.39%</td>
<td>50%-perc. dev. from end-to-end line</td>
</tr>
<tr>
<td>5.3%</td>
<td>inter-stroke time</td>
</tr>
<tr>
<td>5.14%</td>
<td>80%-perc. dev. from end-to-end line</td>
</tr>
<tr>
<td>5.04%</td>
<td>20%-perc. dev. from end-to-end line</td>
</tr>
<tr>
<td>5.04%</td>
<td>50%-perc. pairwise acc</td>
</tr>
<tr>
<td>3.44%</td>
<td>phone orientation</td>
</tr>
<tr>
<td>3.08%</td>
<td>mid-stroke finger orientation</td>
</tr>
<tr>
<td>0.97%</td>
<td>up/down/left/right flag</td>
</tr>
<tr>
<td>0%</td>
<td>change of finger orientation</td>
</tr>
</tbody>
</table>
Example feature

- Coordinates of the two endpoints
Example feature

- Median velocity of the last five points
  - “ballistic” scrolling
- Mean resultant length
  - 1 for straight stroke, 0 for random angles
- Length of the trajectory
- Direct distance between endpoints
- Largest perpendicular distance between end-to-end line & trajectory
- Stroke duration
- Inter-stroke time
Feature selection: informativeness

- $I_F = I(F; U)/H(U)$
- Most informative features
  - Area covered by fingertip
  - 20% percentile of stroke velocity
  - Fingertip pressure
  - Direction of the stroke
  - Locations of endpoints
- $x$ coordinate more informative than $y$ coordinate
Feature selection: correlation
Classification

- **k-NN**
  - Using a k-d tree
  - Euclidian distance
  - k between 1-7
    - Cross-validation

- **SVM**
  - RBF kernel
    - 5-fold Cross-validation

- Combine scores of multiple strokes
  - Threshold the combined score
Results

- EER = 13% based on single stroke
- EER = 1-2% for 11-12 strokes
- Reading text: Median one stroke per 3.9 sec
- Image comparison: one stroke per 1.0 sec
- Verification time with 11 strokes: 11-43 sec
Results

- inter-week authentication
  - EER = 0-4%
- Inter-session authentication
  - EER = 2-3%
- Short-term authentication
  - EER = 0%
References

• “An Efficient User Verification System via Mouse Movements”, 2011

• “Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication”, 2013