On the Security and Usability of Challenge Questions

29 May 2009
Mike Just
University of Edinburgh
Outline

- What are Challenge Questions?
- Challenge Question Research
- Our Research
  - Collecting data
  - Analyzing data
- What Does it all Mean?
- Further Information
What are Challenge Questions? (1 of 3)

- What are 'Challenge Questions?'
  - Type of 'authentication credential'
  - Users register Question & Answer
  - To authenticate later, user is posed Question and asked to provide Answer

- 'Something You Have'
  - Access card
  - Smartcard
  - Mobile

- 'Something You Are'
  - Fingerprints
  - Iris/retinal scan
  - Facial scan

- 'Something You Know'
  - Passwords
  - PINs
  - Images

- 'Something You Memorize'
  - Challenge questions
  - Images

- 'Something You Already Know'
  - Challenge questions
  - Images
What are Challenge Questions? (2 of 3)

- Common Examples
  - 'What is my Mother's Maiden Name?'
  - 'What was the name of my first pet?'
  - 'What was the name of my primary school?'

- How do Challenge Questions support authentication?
  - The answers to the questions should be known only to the users that registered the questions, similar to how passwords should be uniquely known.
What are Challenge Questions? (3 of 3)

- How and why do we use Challenge Questions?
  - Almost exclusively as secondary/fallback authentication in case of lost primary credential
  - Sometimes used to complement primary credential
  - Often driven by desire to avoid costly help-desk calls
  - In some cases, 're-registration' is possible, but not always
    - Too expensive or takes too much time
    - Not all sites have a registration phase (that includes user identification with shared secrets)
  - So, some form of secondary authentication is desirable
    - Challenge Questions are today's ubiquitous choice
Challenge Question Research (1 of 3)

- What is studied w.r.t. Challenge Questions?
  1. Security (Attacker's Point-of-View)
     - How difficult is it to determine the answers to the questions?
     - Demonstration of security often involves *quantitative analysis*
  2. Usability (User's Point-of-View)
     - How easy is it to choose questions?
     - How easy is it to remember the answers?
     - Demonstration of usability often involves *qualitative research*
What has been studied w.r.t. Challenge Questions?

- Early '90s usability studies referred to 'word pairs,' and 'associative' or 'cognitive passwords'
- Focused on facts, opinions or interests. Studies [Haga et al.] suggested facts were easier to recall, but more easily guessable by friends or family
- Early '00 analysis focused on tolerating users forgetting or mis-typing answers with secret sharing [Ellison et al., Frykholm et al.]
- Recent work [Rabkin, Jakobsson et al.] has focused directly on the insecurity of administratively-chosen challenge questions, and on specific questions ('Mother's Maiden Name')
- Jakobsson et al. have published a novel solution based upon user preferences (binary), though more study is needed
Challenge Question Research (3 of 3)

- More recently ...

- Single user authentication

- Group authentication
Our Research (1 of 2)

- **Problem:** 'Systematic analysis of the security and usability of challenge questions is lacking'
- **Method:** Investigate security and usability of user-chosen challenge questions
- **Goals:** To answer the following:
  - Do users choose secure questions?
  - Do users choose memorable answers?
  - Can we lead realistic yet ethical authentication experiments?
Our Research (2 of 2)

- Lead three experiments with classes at the University of Edinburgh
  - Human Computer Interaction (HCI), Computer Security, and Biology class
- 170 participants submitted 500 questions
- Devised methods for measuring security and usability of the questions (and answers)
- Novel approach for collecting data
Collecting Data (1 of 3)

- Ethically challenging, but users readily submit
- Issues regarding participant behaviour
  - Equate credentials with other private information?
  - Contribute *real* information?
  - Degree of freedom with user-chosen questions
- Opportunities for improved Collector behaviour
  - Challenge to ourselves: Don't collect!
  - Avoid having to maintain information
  - Consistent message: Keep credentials to yourself!
Collecting Data (2 of 3)

Stage 1

Participant

Questions

Answers

Stage 2

Questions

Answers

MATCH?

Usability Analysis

Experiment

Security Analysis

Version 1 – Pen-and-Paper Only

Version 2 – Online & Pen-and-Paper
Participants use of 'real' Questions and Answers

- We asked if participants would use same Questions and Answers in real applications (e.g. Banking)
- Of the respondents (92%) indicating that they would likely reuse their questions, 61% indicated some influence from not submitting their answers

Participants and personal privacy

- We asked participants if they would be concerned if their friends or family members knew their Questions and Answers
- More than two-thirds of the questions raised 'no concern' at all for participants with < 10% meriting strong concern
Existing security analysis of Challenge Questions is limited, and ad hoc.

There are no clear guidelines for choosing 'good' questions and answers.

We're wanted a more systematic approach that would either
  - Provide some guidance for secure design, or
  - Recommend abandonment of the concept.
Security Analysis (2 of 7)

Increasing Information for Attacker

Answer alphabet and distribution, common answer sets
Questions, distributions of likely answers
User account, published data, social networks, friends, family, ...

Attack Methods
Blind Guess
Focused Guess
Observation

Answer Guess
Security Analysis – Blind Guess (3 of 7)

- Brute force attack
- Security Levels based on equivalence to passwords
  - 6-char alphabetic password ($2^{34}$)
  - 8-char alphanumeric password ($2^{48}$)
- Answer entropy: 2.3 bits (1st 8 chars), then 1.5 bits
- Results (by question)
  - Average answer length: 7.5 characters
  - 174 Low, 4 Medium, 2 High
- Results (by user)
  - Q1 – 59 Low, 1 Medium, 0 High
  - Q1, Q2 – 38 Low, 13 Medium, 9 High
  - Q1, Q2, Q3 – 5 Low, 19 Medium, 36 High
Blind Guess (cont'd)

- Unlike passwords, the alphabet for answers is just 26 lowercase letters (plus 10 digits in some cases).
- Use of a single question seems to provide insufficient protection against the simplest attack.
- But, multiple questions seem to help.
- Online attacks considered (targeted and random). Offline attacks would require more security ($2^{80}$).
Attacker knows the Challenge Questions

Security Levels same as for Blind Guess

Answer types and space

Results (by question)
- 167 Low, 0 Medium, 13 High

Results (by user)
- Q1 – 58 Low, 0 Medium, 2 High
- Q1, Q2 – 46 Low, 11 Medium, 3 High
- Q1, Q2, Q3 – 5 Low, 28 Medium, 27 High

Much room for refinement of 'Space'

<table>
<thead>
<tr>
<th>Q Type</th>
<th>%</th>
<th>Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proper Name</td>
<td>50%</td>
<td>$10^4 - 10^5$</td>
</tr>
<tr>
<td>Place</td>
<td>20%</td>
<td>$10^2 - 10^5$</td>
</tr>
<tr>
<td>Name</td>
<td>18%</td>
<td>$10^3 - 10^7$</td>
</tr>
<tr>
<td>Number</td>
<td>3%</td>
<td>$10^1 - 10^4$</td>
</tr>
<tr>
<td>Time/Date</td>
<td>3%</td>
<td>$10^2 - 10^5$</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>6%</td>
<td>$10^8 - 10^{15}$</td>
</tr>
</tbody>
</table>
Security Analysis – Observation (6 of 7)

- Attacker tries to obtain or observe the answer

- Security Levels defined qualitatively
  - Low – Answer publicly available
  - Medium – Answer not public, but known to F&F
  - High – Neither

- Levels assigned to questions by
  - Subjective analysis, and
  - Participant input (provided upper bound only)

- Results (by question)
  - 124 Low, 54 Medium, 2 High

- Results (by user)
  - 24 Low, 34 Medium, 2 High
  - Did not "sum" levels (used max)

- Much room for refinement of levels and analysis
Security Analysis – Overall (7 of 7)

- Overall rating is a 3-tuple (Blind, Focused, Observation)
- Results
  - All Low – 1 participant
  - All High – 0 participants
  - No Lows – 31 participants (50%)
  - (H,M,M) or (M,H,M) – 15 participants (25%)
  - (H,H,M) – 11 participants (20%)
- Perceived effort of Stranger to Discover Answers
  - Very difficult (47%), Somewhat difficult (42%), Not difficult at all (11%)
- Perceived effort of Friend/Family to Discover Answers
  - Very difficult (11%), Somewhat difficult (36%), Not difficult at all (53%)
Usability Analysis (1 of 3)

- Usability often refers to 'usable interface design'
- For usable authentication, similar principles apply
  - The user should be able to understand and execute their task
  - We're dealing specifically with information
- In this case, we're more concerned with mental capabilities, e.g., processing, memory
Usability Analysis (2 of 3)

- **Applicability**
  - Users have sufficient information to provide an answer to a question
  - E.g., 'What was my first pet's name?'
  - Relevant to administratively-chosen questions (not user-chosen)

- **Memorability**
  - Users can consistently recall the original answer to a question over time
  - Precise recall, 'blank'

- **Repeatability**
  - Users can consistently and accurately repeat the original answer to a question over time
  - E.g., 'Favourites' change over time, 'Street' versus 'Avenue'
Answer recall

- 44 errors (15%)
  - Reduces to 15 errors (5%) if we exclude 'capitalization' errors

Answer recall (from 99 users)

- 28 users (28%) made at least one error
  - Reduces to 14 users (14%) if we exclude 'capitalization' errors

Comments suggest that 'complicated answers' and allowance of free-form answers may be culprit

Florêncio & Herley (2007) found that 4.28% of Yahoo! users forget their passwords

Our results were after 23-28 days, with young students
Our results suggest significant concerns with the security and usability of challenge questions.

But, before we write-off challenge questions ...

- Multiple questions seem to help (security at least)
- Our assessment model is preliminary
- Our experiments were only with students
- Current implementations are terribly boring
What Does it All Mean? (2 of 2)

**Next Steps**
- Further refine security model and assessments (tighter entropy, question independence, observations)
- Dynamic assessments
- Broader usability studies
- New types of information for authentication (new questions)

**But, how to improve usability?**
- Fixed-form answers
- Tolerance for < 100% accuracy
Further Information

- Project web site
  - Includes some recent publications
- Email
  - mike.just@ed.ac.uk
## Usability Results

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td># Questions</td>
<td>51</td>
<td>66</td>
<td>180</td>
<td>297</td>
<td>100</td>
</tr>
<tr>
<td>Exact Answer</td>
<td>31</td>
<td>57</td>
<td>165</td>
<td>253</td>
<td>85.19</td>
</tr>
<tr>
<td>Any Error</td>
<td>20</td>
<td>9</td>
<td>15</td>
<td>44</td>
<td>14.81</td>
</tr>
<tr>
<td>Not Capitalization</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td>5.05</td>
</tr>
<tr>
<td>Completely diff</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Repeatability</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td># Users</td>
<td>17</td>
<td>22</td>
<td>60</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>Any Error</td>
<td>11</td>
<td>6</td>
<td>11</td>
<td>28</td>
<td>28.28</td>
</tr>
<tr>
<td>Not Capitalization</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>14</td>
<td>14.14</td>
</tr>
</tbody>
</table>

29 May 2009

Challenge Question Security and Usability