On the Security and Usability of Challenge Questions

29 May 2009

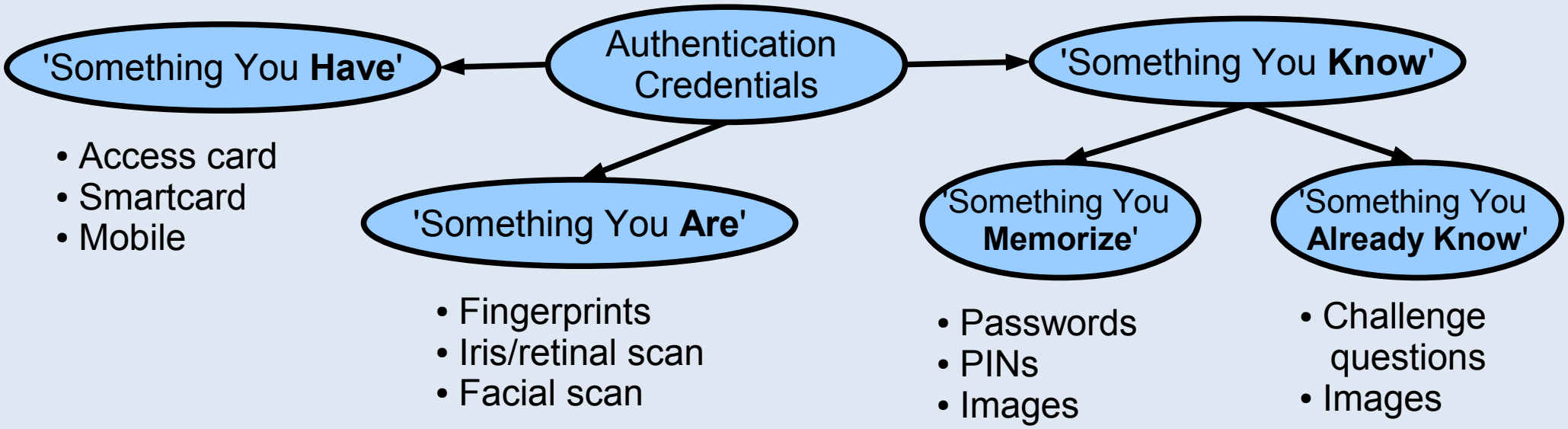
Mike Just
University of Edinburgh

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



What are Challenge Questions? (2 of 3)

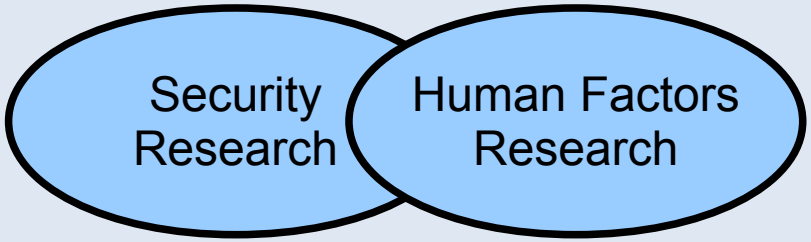
- Common Examples
  - 'What is my Mother's Maiden Name?'
  - 'What was the name of my first pet?'
  - 'What was the name of my primary school?'
- How do Challenge Questions support authentication?
  - The answers to the questions should be known only to the users that registered the questions, similar 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 desireable
    - Challenge Questions are today's ubiqutous 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*



Challenge Question Research (2 of 3)

- 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
  - Just, Aspinall, ”Challenging Challenge Questions,” *Trust 2009*, April 2009
  - Schechter, Bernheim Brush, Egelman, ”It's no secret: Measuring the security and reliability of authentication via 'secret' questions,” *IEEE Security and Privacy 2009*, May 2009
  - Just, Aspinall, ”Personal Choice and Challenge Questions: A Security and Usability Assessment,” *SOUPS 2009*, July 2009
- Group authentication
  - Toomim, Zhang, Fogarty, Landay, ”Access Control by Testing for Shared Knowledge,” *CHI 2008*, April 2008
  - Bonneau, ”Alice and Bob in Love: Cryptographic Communication Using Natural Entropy,” *Security Protocols 2009*, April 2009

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?

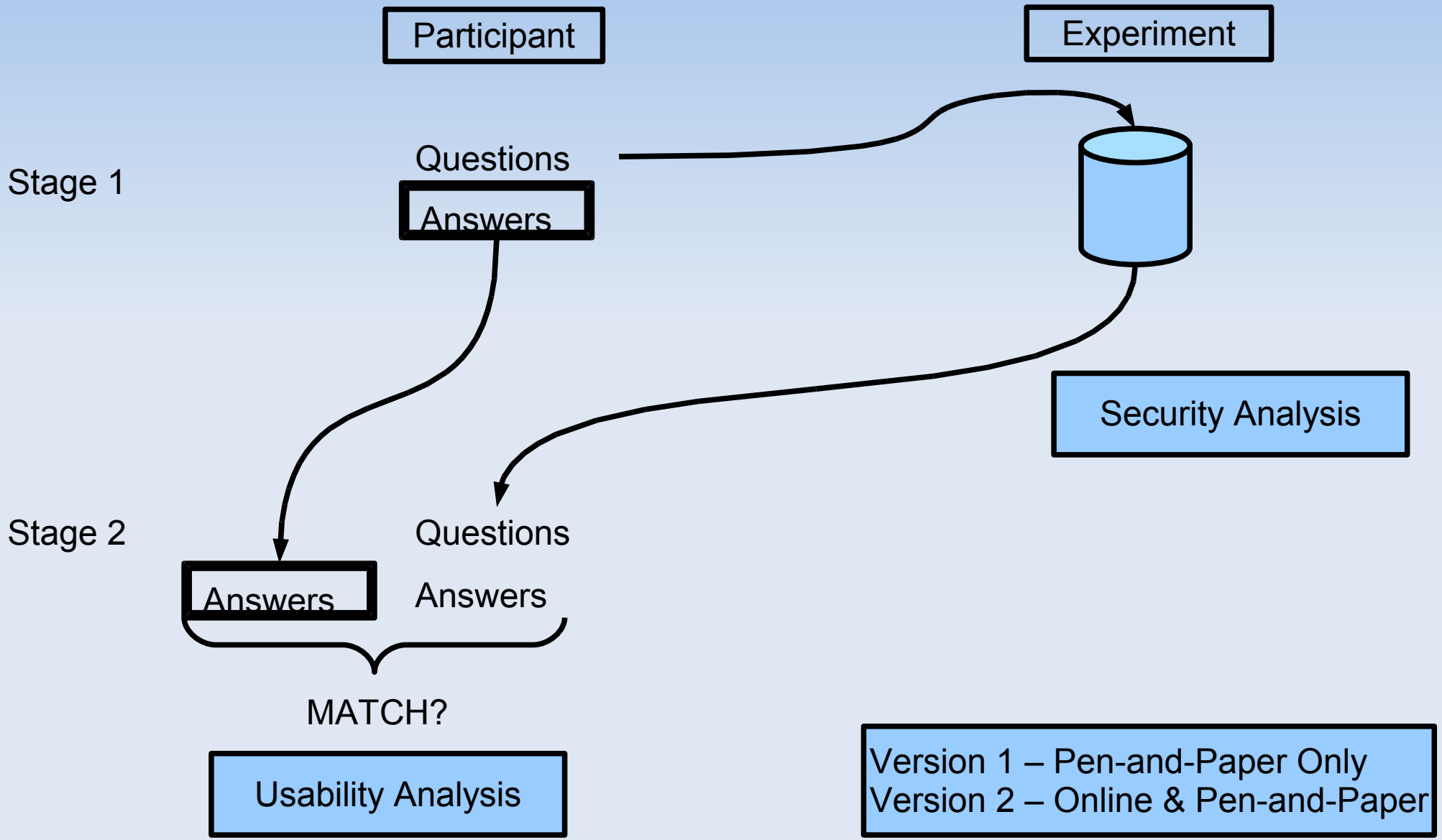
Our Research (2 of 2)

- 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)



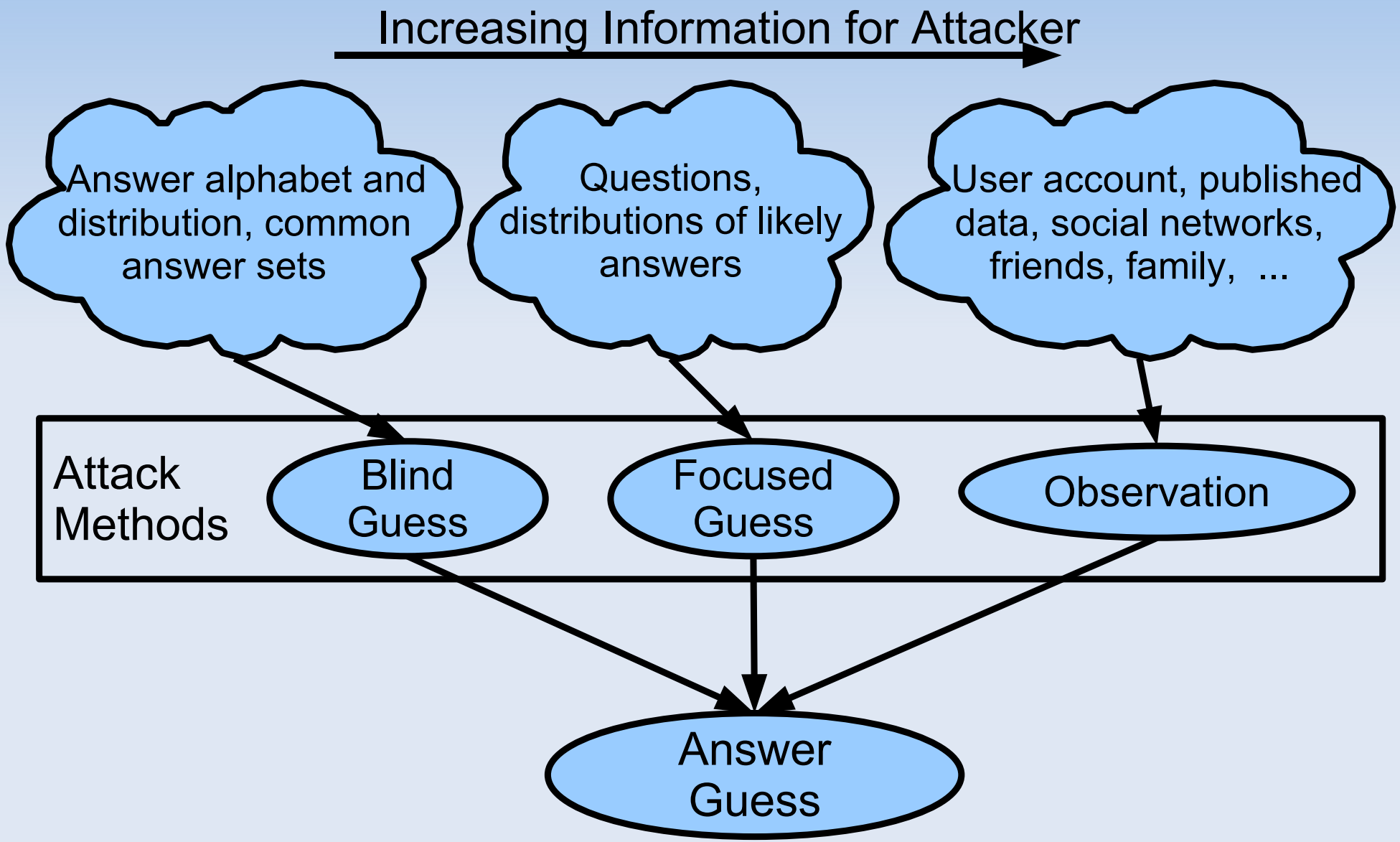
Collecting Data (3 of 3)

- 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 re-use 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

Security Analysis (1 of 7)

- 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)



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})

Low (2^{34}) Med (2^{48}) High

- 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

Security Analysis – Blind Guess (4 of 7)

- 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 (targetted and random). Offline attacks would require more security (2^{80})

Security Analysis – Focused Guess (5 of 7)

- 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'

Q Type	%	Space
Proper Name	50%	$10^4 – 10^5$
Place	20%	$10^2 – 10^5$
Name	18%	$10^3 – 10^7$
Number	3%	$10^1 – 10^4$
Time/Date	3%	$10^2 – 10^5$
Ambiguous	6%	$10^8 – 10^{15}$

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'

Usability Analysis (3 of 3)

- 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

What Does it All Mean? (1 of 2)

- 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
  - http://homepages.inf.ed.ac.uk/mjust/KBA.html
  - Includes some recent publications
- Email
  - mike.just@ed.ac.uk

Additional Slides

Usability Results

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