University of Glasgow
Department of Computing Science Seminar Series

Challenge Question Authentication

25 February 2009

Mike Just
University of Edinburgh

(joint work with David Aspinall)

Introducing Your Speaker

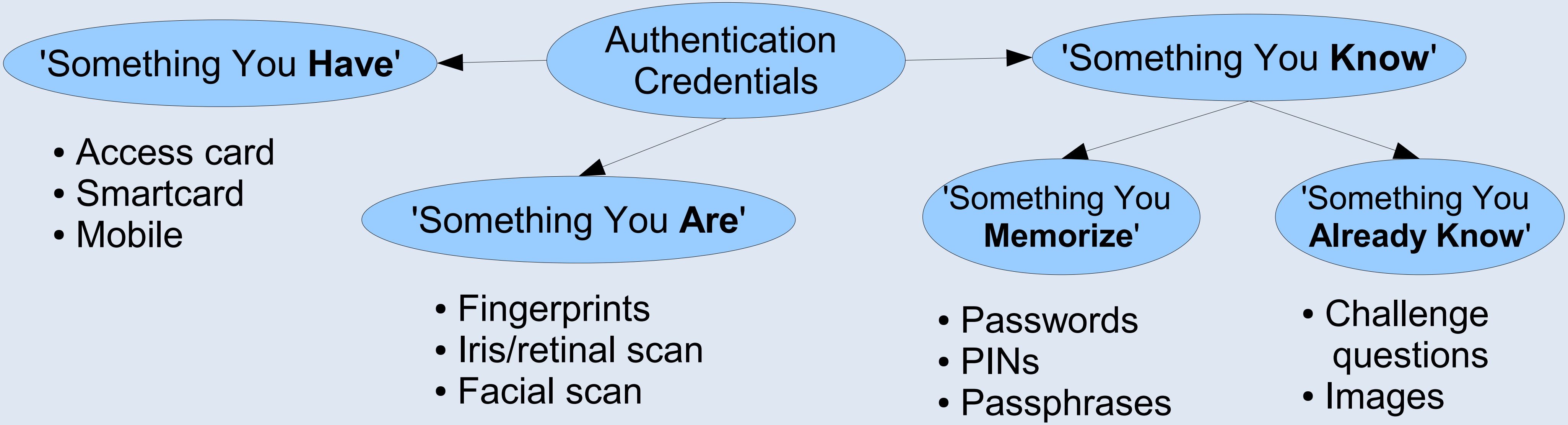
- Visiting Research Fellow till Sept 2009
  - EPSRC-funded project with David Aspinall
- Former Director of Innovation, and continue to work part-time, remotely for Canadian Government
- Worked in public and private sectors, and academia during past 10 years (focus on Applied Cryptography)
- In 2005, designed the Challenge Question Authentication Solution used by Canadian Government to authenticate appox 3 million citizens and businesses
- PhD, Carleton University, 1998

Outline of this Talk

- The Scenario
- Challenge Question Research
- Our Research
- Experiments
- Security and Usability Analysis
- What Does it all Mean?
- Further Information

The Scenario (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



Authentication Credentials

'Something You **Have**'

- Access card
- Smartcard
- Mobile

'Something You **Are**'

- Fingerprints
- Iris/retinal scan
- Facial scan

'Something You **Know**'

'Something You **Memorize**'

- Passwords
- PINs
- Passphrases

'Something You **Already Know**'

- Challenge questions
- Images

The Scenario (2 of 3)

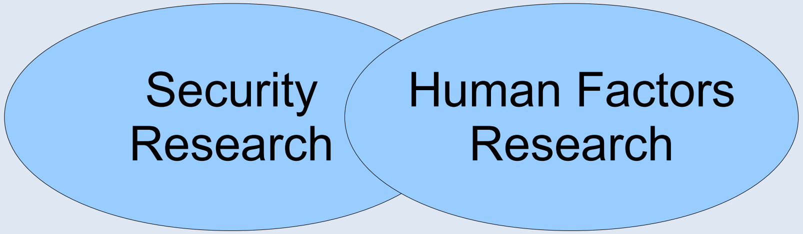
- Common Examples
  - 'What is my Mother's Maiden Name?'
  - 'What was the name of my first pet?'
  - 'What was the name of my primary school?'
- How do Challenge Questions support authentication?
  - The answers to the questions should be known only to the users that registered the questions, similar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

The Scenario (3 of 3)

- How and why do we use Challenge Questions?
  - Almost exclusively as secondary/fallback authentication in case of lost 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
  - (And yes, they could be used as a primary credential as well)

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)

- And while other forms of authentication have received more study, not all is transferable
  - 'Known' information risk is difficult to quantify
- A systematic analysis of the security and usability of challenge questions is lacking
- Basic facts regarding Challenge Questions aren't known

Our Research (1 of 2)

- Our goals are to answer the following:
  - Do users choose secure questions?
  - Do users choose memorable answers?
  - Can we lead *realistic* yet *ethical* authentication experiments?
- Investigation of security and usability of user-chosen challenge questions

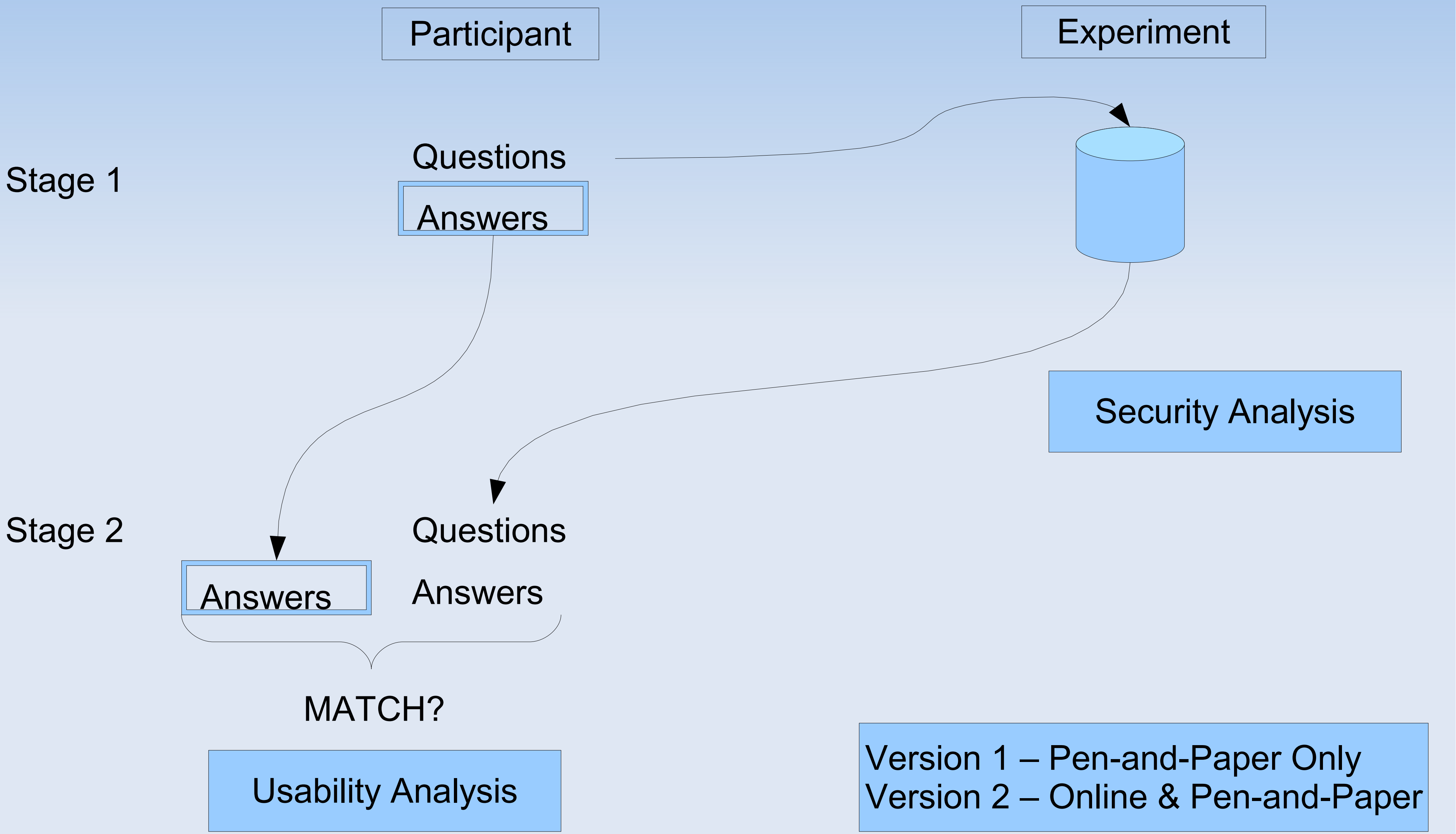
Our Research (2 of 2)

- Lead three experiments with classes at the University of Edinburgh
  - Human Computer Interaction (HCI) class (Oct/Nov 2008)
  - Computer Security class (Jan/Feb 2009)
  - Biology class (Jan/Feb 2009)
- 170 participants submitted 500 questions
- The remaining slides review our preliminary results

Experiments (1 of 3)

- Collecting authentication data can be tricky
  - Users are consistently told to not reveal their authentication information
  - For our analysis, we'd like to see this information
  - Ethically, we could ask for their information
  - But will users give use 'real' information?
- Our solution
  - Pen-and-paper experiments where participants retain their authentication credentials
  - Participant self-assessments

Experiments (2 of 3)



Participant

Experiment

Stage 1

Questions

Answers

Security Analysis

Stage 2

Answers

Questions

Answers

MATCH?

Usability Analysis

Version 1 – Pen-and-Paper Only
Version 2 – Online & Pen-and-Paper

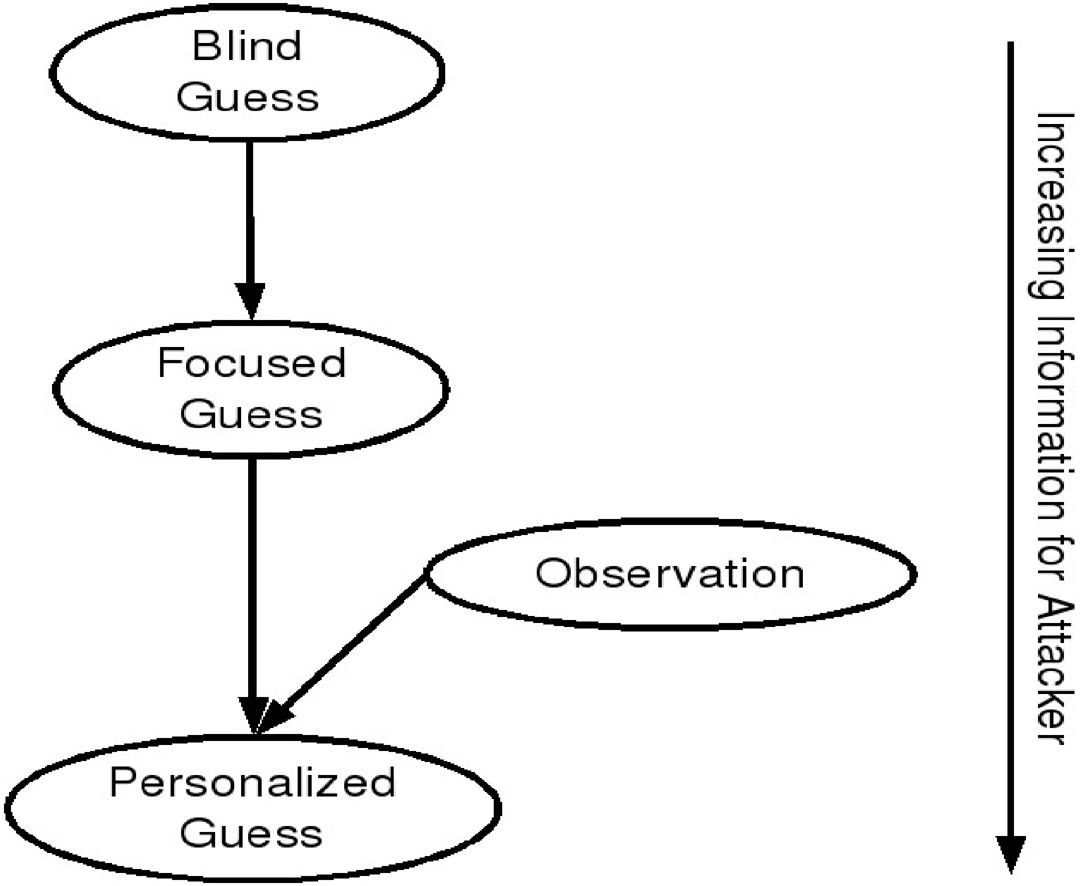
Experiments (3 of 3)

- 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 extremely ad hoc
- There are no clear guidelines for choosing 'good' questions and answers
- We're attempting to follow a more systematic approach that will either
  - Provide some guidance for secure design, or
  - Recommend abandonment of the concept

Security Analysis (2 of 7)



Blind Guess

Focused Guess

Observation

Personalized Guess

Increasing Information for Attacker

- Answer alphabet, distribution
- Common answer sets

- Questions
- Distribution of likely answers

- Likely user 'peer groups'
- Personal information about user

Security Analysis (3 of 7)

- Blind Guess
  - Based upon our prelimary experiment results the average answer length is 7.95 characters
  - Unlike passwords, the alphabet for answers is just 26 lowercase letters (plus 10 digits in some cases)
  - With uniformly distributed answers, we have entropy (uncertainty) of 4.7*8=37.6 bits for 8-character answer
  - According to Shannon, for answers from English lang. we can reduce to 2.3*8=18.4 bits of uncertainty (approximately 350,000 answers)
  - For comparison, a uniformly chosen password (upper and lowercase, numbers) has approx. 6*8=48 bits of uncertainty

Security Analysis (4 of 7)

- Blind Guess (cont'd)
  - Use of a single question seems to provide insufficient protection against the simplest attack (Blind Guess)
  - Conclusion: Without knowledge of the questions, or personal details, attacks will succeed
  - Why? It's a numbers game.
  - For a targetted attack (online), some attackers will succeed
  - For a random attack (online), some accounts will be compromised
  - For an offline attack, all attackers would succeed

Security Analysis (5 of 7)

- Focused Guess
  - Knowing question gives further reduction in uncertainty (and questions are effectively public)
  - E.g. ”What was my first pet's name? (http://www.babynames.com/Names/Pets/ gives the top 200 names for dogs & cats)
  - Most questions suggest a small target answer space (see Tables)
  - Some questions simply suggest very low entropy answers, e.g. ”What religion is my father?”, ”Favourite colour?”

Q Type	**%**
Proper Name	50%
Place	20%
Name	18%
Number	3%
Time/Date	3%
Ambiguous	6%

Proper Name	**%**
Last Name	48%
First Name	12%
First & Last	9%
Pet Name	30%
Other	1%

Security Analysis (6 of 7)

- Observations from many sources
  - Questions, User Identifier, Web Site, User, Social Networks, Published data, …
  - Gender, Age (range), Interests, Opinions, Relations, ...
- Personalized Guess
  - Typically involves more work (observation), but can contribute to a much-reduced number of guesses
  - E.g., ”Mother's Maiden Name” is often easy to determine from public records

Security Analysis (7 of 7)

- User Perceptions of Security
  - We asked participants how difficult they believed it would be for (i) strangers, or (ii) friends/family to determine the answers to their questions
- 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)

- Our initial results suggest some difficulty with perfect recall of answers
  - 15% of respondents in our first experiment gave either a completely different, or slightly different answer
  - Comments suggest that 'complicated answers' and allowance of free-form answers may be culprit
  - Further results indicate high incidences of recall (perhaps due to our participant population - students)

What Does it All Mean? (1 of 2)

- Our preliminary results indicate that relying upon only a single question-answer is insecure
- Some Candidate Recommendations
  - Require multiple questions at authentication
  - Dynamically assess Questions and Answers at registration
  - Use fixed-form answers (e.g., drop-down menus)

What Does it All Mean? (2 of 2)

- Next Steps
  - Complete our security assessment, aligned to Attack Model
  - Study the impact of our recommendations
  - Investigate use of 'more recent' information for authentication (not 'original' answers)
  - More study of Jakobsson's 'preferences' solution
  - Use of image, rather than textual, information
- Other 'Lessons Learned'
  - 'Prizes' not necessarily sufficient for participation
  - Require much larger groups for meaningful usability results

Further Information

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