# Knowledge-Based Authentication: Evaluating and Improving

## Case for Support – Description of Proposed Research and its Context

**David Aspinall and Michael Just**

# 1 Introduction

Authentication is central to computer security and almost every use of computerised systems. With the explosion of online e-commerce, banking, social web sites and governmental services, the problem of finding secure, usable and efficient authentication systems is more acute than ever. The risks of security failure are obvious, and unusable or inefficient systems additionally risk loss of customers or overly expensive support services managing password recovery.

Despite the obvious importance of everyday authentication and the widespread adoption of improved mechanisms such as challenge questions, there is a surprising lack of underpinning published research for these methods. Comparative studies, measures of usability and recoverability costs, scientifically justified guidelines for efficient implementation, are all lacking.

This research proposes to understand and assess existing practice, and make recommendations for improvement. We expect that the results will have a widespread impact across many sectors, both inside and outside of the UK. The research will be undertaken by a Visiting Fellow (Just), who is the world leader in Knowledge-Based Authentication (KBA), and responsible for the KBA mechanism used in the Government of Canada's online e-government solution, serving 3 million accounts. We anticipate that the results of our research will contribute positively to the security and usability of many applications, both inside and outside of the UK.

# 2 Background

Entity authentication supports the "corroboration of the identity of an entity" [HAC97]. Authenticating entities provide evidence from one of the following categories to aid successful corroboration:

- Something *known* to the entity. For example, a password, or Personal Identification Number (PIN).
- Something *inherent* to the entity. For example, a scan of a biometric such as a fingerprint or iris.
- Something *possessed* by the entity. For example, a smartcard or other physical token.

Success typically results in access to some information system, e.g., an individual's bank account. In addition to being one of the key areas of study in cryptographic and computer security research, entity authentication directly impacts many individuals on a daily basis as they will often authenticate using passwords to gain access to web sites, personal accounts and other information systems.

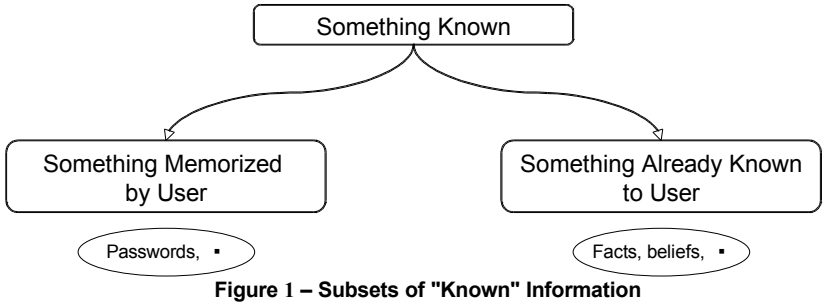



**Figure 1 – Subsets of "Known" Information**

While many techniques for authentication are studied in academia, the most common and often-used technique involves a simple password. Unfortunately, passwords have not provided a sufficient level of security for the last several years, and improving their security involves solutions that stretch human memory capacity (e.g., with longer passwords). In some cases, the threshold of human ability to "memorize more" is near, if not surpassed already. For this reason, alternate techniques for entity authentication are studied. Techniques based upon "something inherent" or "something possessed" have thus far met with limited, albeit increasing, success.

Our research will focus on a different aspect of "something known" referred to as *Knowledge-Based Authentication (KBA)*. KBA uses information *already known* to an individual, as opposed to *memorized information* (such as a password). With this distinction, the authentication research challenge moves from finding effective ways for users to *memorize information*, to effective ways to recall and use *already known*

*information*. This distinction is shown in Figure 1, and it is our conjecture that KBA may offer more usable and secure authentication systems by providing less strain on human recall abilities. Indeed, the wide deployment of such solutions in industry (in the form of "challenge questions" – see below) seems to suggest broad agreement with this conjecture. However, we also believe the current deployment of KBA solutions do not offer sufficient security (consider, for example, those that ask for only a single answer, such as "What is your mother's maiden name?", while numerous web sites claim to be able to easily provide answers to such questions).

"Challenge questions" are one form of KBA using something "already known" to an entity (as opposed to memorized for the purpose of authenticating). Their use involves registering an answer to a question, and later providing the answer as a means of authenticating. For example, at registration an individual might provide the answer to the question *"What was the name of my first childhood friend?"*, and then later asked to respond to the same question in order to authenticate. And because the information is known only to that entity, as opposed to memorized solely for the purpose of authentication, challenge questions (and more generally, KBA) offer the potential for better recall (and therefore, usability) at time of authentication. And with improved usability comes the potential for increasing security through techniques that do not necessarily over-tax human memory capacity.

However, despite the near ubiquity of KBA techniques in industrial applications for recovery from a forgotten password, there exists surprisingly little analysis with regard to either their *security* or *usability*. A review of peer-reviewed literature will uncover around a dozen articles on the subject of KBA. Early work in the area of challenge questions recognized the advantages of using more memorable items for successful user authentication, but this work has generally been sporadic and never brought together in a complete model. Several articles have focused on determining the usability of so-called *cognitive* and *associative passwords* [Haga91, Pond00, Sec94, Zvir93]. Haga *et al.* [Haga91] identified cognitive passwords as questions and answers that relate to the users' facts, opinions, or interests. They further classified them as either *fact-* or *opinion-based* questions. Associative passwords are similar to the game of word association. A word pair is used in which the first word prompts the answer (the second word). For the purpose of a question-and-answer framework, the first word serves as a question and the second as the answer. Usability studies revealed small variances in the usability of cognitive or associative passwords: where the former were generally easier to remember, they were also more susceptible to guessing by a close family member or colleague.

More recent work focused on cryptographic design issues. Ellison *et al.* [Elli00] examine several issues related to the secure use of challenge questions, and apply secret sharing principles to account for a forgetful user: users must be able to answer $t < n$ questions properly. Frykholm *et al.* [Fryk01] focus on improving usability by designing a system using error-correcting that tolerates the typing mistakes made when entering an answer. O'Gorman *et al.* [OGor04] focus on improved user experience, rather than cryptographic issues. They use "selectable questions" and "multiple-choice answers" and present several implementations related to this model.

The proposed Visiting Fellow, Dr. Michael Just was the first to provide some initial structure in which KBA solutions could be designed and evaluated, providing a taxonomy of questions and answers, as well as defining some measurable security, usability and privacy criteria [Just04, Just05]. Types of questions and answers were identified as fixed, open or controlled, generalizing on the "selectable" and "multiple choice" features of O'Gorman *et al.* above. Security criteria included not only the difficult of guessing answers, but also of "observing" answers, recognizing that KBA uses information that might be readily, even publicly, available to attackers. Usability criteria included applicability (suitability for the target population), memorability (recall ability), and repeatability (consistency over time of both the syntax and semantics of an answer). Privacy criteria were established to aid in limiting the collection and use of personal information beyond the requirements for authentication.

Just also focused on both technical and usable design issues. For the former, he considered issues related to the number of questions posed, as well as complementary security measures that would be implemented as part of the overall authentication solution. For the latter, small focus group analysis was performed so as to increase the usability of the overall solutions. The results of Just's model and techniques were incorporated into the Government of Canada's online authentication solution for citizens and businesses. Research suggested that 2 questions would not be sufficiently secure (or at least not perceived as such) so that this particular solution uses a combination of 3 questions. The first allows the user to select a fixed question from a drop-down list of 15 pre-defined questions so as to provide sufficient choice for users; one of the questions is selected from the list. The second and third are more generic, respectively asking the questions: "Who is a person that is memorable to you?"; and, "What is a date that is memorable to you?" For the second and third questions, users are allowed to submit "hints" (or "cues") that will be presented when later answering the questions.

For an evaluation of security, measures of the potential work involved for an attacker to guess an answer were computed. However, unlike passwords whose length is typically fixed with a small range, answers to questions can vary greatly, making accurate quantitative analysis difficult in the absense of empirical data. In addition, little analysis was performed as to the ability of an attacker to find the answer somewhere else (for example, if the answer to question 3 was the date of birth of the user, the answer might be posted on that user's Facebook page), or the ability of an attacker that has a relationship with the user (e.g. cousin) to determine the answer. Such attacks are more difficult to quantify, but are extremely important to providing complete information regarding the security of the solution.

For an evaluation of usability, the overall set of questions was validated by a focus group of 17 users. However, while providing some guidance regarding the suitability of some of the questions, the study did not validate use of the solution in practice. And more importantly, given that recovery answers are seldom used, there was no analysis as to the ability of user's to recall their answers after 6 months time.

Therefore, there remains many unanswered questions regarding the secure, usable and privacy-friendly use of knowledge-based authentication techniques. Security and privacy failure lead to compromised personal information and lack of confidence in the systems that the authentication solutions are intended to protect. Unusable or inefficient systems additionally risk loss of customers or overly expensive support services managing password recovery. This is especially important for knowledge-based authentication solutions whose primary purpose is to support recovery from forgotten passwords.

# 3 Programme and Methodology

Our research into Knowledge-Based Authentication (KBA) hopes to answer the major open questions left unresolved by previous work.

- Are currently deployed challenge question authentication solutions sufficiently **secure**? In particular, can an attacker spend less time and effort in determining an answer to a user's recovery question than attempting to determine their login password? It seems likely, given that most solutions rely upon only a single answer. And is the use of multiple questions sufficient to provide the desired security level, given that the attacker may not have to *guess* but rather *locate* the answers, and the attacker might have a relationship with a user, providing some a priori knowledge.
- Are currently deployed challenge question authentication solutions sufficiently **usable**? In particular: Are the questions applicable to the user base, especially if the user base crosses many cultures (e.g., asking for the year in which the user purchased their first car, is not terribly useful to user's that have never owned a car)? Are the questions supportive of easy recall, even after 6 months have elapsed since the answer was last provided? Are the questions supportive of a user consistently providing the same answer over time (e.g. Must a user recall whether they entered the word "Street", or the shortened "St."?).
- Are currently deployed challenge question authentication solutions sufficiently **privacy-friendly**? In today's climate of identify theft, and use of personal information beyond its intended purpose, there is great need for protecting a user's privacy. For example, responding to a question with the name of a user's personal physician might reveal unwanted information in case the doctor's specialty reveals personal information about the user's condition.
- Is it possible to **design authentication solutions** using "already known information" (introducing only limited memorization burdens on the user) that can meet the requisite security, usability and privacy requirements? For security reasons, the entire system design must be considered, beyond just "asking questions." For example, some password recovery solutions will, in addition to a challenge question, send an email to the user (to which the user must respond, or "click" an accompanying link) to complete the recovery process. This alone seems to suggest some security concerns with the use of challenge questions by themselves. Hence it would be important to investigate what other complementary techniques might raise the level of security. For usability reasons, additional design features that may support error-tolerant responses, as well as "t of n" schemes whereby only t answers need be provided in response to posing n questions. In this vein, a series of related or dependent questions might prove viable, in which after answering each question a user is provided a hint for the next answer.
- Can KBA solutions provide a viable **alternative to passwords**, and other primary authentication techniques used today, e.g. public key certificates? Till recently, deployments of KBA solutions were almost exclusively for recovery purposes. Some sites (e.g., banks) now use them in addition to passwords, presumably for added security (though this is yet another situation that has received insufficient analysis). Though given the security concerns over password use, it is interesting to determine the practical benefits of using KBA for a wider purpose (i.e., on their own, without passwords).

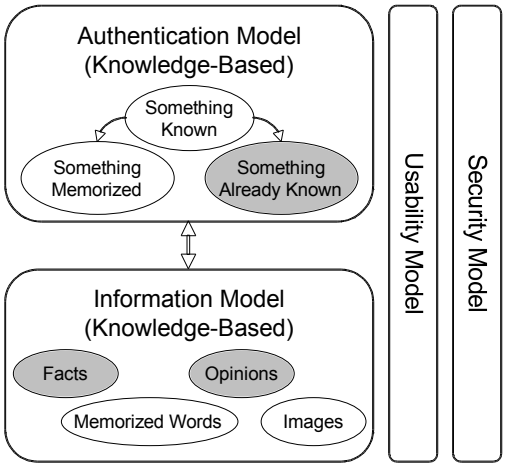
- Is it possible to provide a **"cookbook"** that would allow designers to develop a KBA solution that meets their needs with regard to security, usability and privacy. Previous work has only provided a "partial menu" – the "complete menu," and accompanying "cooking instructions" are required. This cookbook could, for example, identify requirements such as the number of questions that need to be used in order to achieve a particular level of security, and perhaps try to quantify a range of entropy for the answers to particular classes of questions.

As part of responding to these open questions, we will use several measurable objectives in order to assess the success of our work, namely

- The discovery of security, usability or privacy issues with existing solutions that would cause some consideration for the continued use of the solutions.
- The acceptance of our research work for presentation at a recognized security or usability conference, or publication in a related journal.
- The adoption of our models or solutions in a government or industrial application.

In support of our research, we will use the KBA Model depicted in Figure 2 (below). The Authentication Model component will define the types of systems that will support access of known information from individuals. This includes systems for challenge questions as well as complementary security and usability techniques. The Information Model component includes the types of information that can contribute to a secure, usable and privacy-friendly KBA solution, including information already known to an individual (including facts, opinions, etc.), but also any memorization requirements that might be required in order to facilitate effective recovery. The Security Model component includes parameters under which the security and privacy of our KBA techniques can be evaluated. The security modeling will include more traditional quantitative analysis, but also consider the secondary effects of additional information that might be known (or easily determined) by family, friends, strangers, co-workers, or other individuals who might have access to information known to an individual. The privacy modeling will take into account adherence to the OECD privacy principles. The Usability Model component will define the requirements to be met by a user-friendly KBA solution, and also begin to define the methods to be used in evaluating to ensure conformance to these requirements.

**Figure 2 – Components of a Knowledge-Based Authentication Model**




The bulk of the project work will be performed by the Visiting Fellow (Just), with project management (including quality assurance) responsibilities with the Principal Investigator. The workplan (see below, and attached) is decomposed into stages with the ends of each stage coinciding with a milestone for the project. Progress checkpoints will be held bi-weekly in order to ensure progress against these milestones, with formal checkpoints and deliverables delivered at the end of each stage. The results of each stage will be appropriately documented to support subsequent stages, to prepare for submission of publications, and also in support of general communication of our results. Expected durations for, and dependencies between, each stage are provided in the attached workplan.

**Stage 1 – Project Initiation**

The first stage of the programme of work will simply initiate and set-up the project. It will also include the establishment of baseline requirements for security, usability and privacy.

**Stage 2 – Analysis and Evaluation of Existing Solutions**

The second stage of the programme of work will involve the discovery and evaluation of existing KBA solutions. Particular sub-stages of work will include

- Research of Existing Industrial KBA Implementations – Several examples of 'real-life' implementations of KBA techniques (particularly in UK industry and government) will be selected,

reviewed and documented in support of subsequent analysis and evaluation. This task will involve the systematic registration at various web sites supporting authentication and password recovery. Details regarding the primary password authentication solution will be documented (e.g. to support later comparison between the entropy of the passwords versus the answers to the challenge questions), and additionally the details of the overall solution for password recovery.

- Analysis of Solutions Against Requirements – The solutions will be examined for their adherence to established security, usability and privacy requirements. This analysis will be performed independent of any empirical data. For the security analysis, attacks to the basic password authentication portion will be quantified and compared to similar quantification for the answers. Some analysis as to the likelihood of locating (as opposed to guessing) answers will also be made, perhaps measuring on a scale of low, medium or high risk. For the usability analysis, factors such as the normalization of answers will be considered – for example, if the answers are case-insensitive, there is less burden upon users to recall which characters were capitalized as part of their original response (though such a feature would negatively impact security as well). For the privacy analysis, consideration to the likelihood of answers revealing personal user information will be considered. For example, a user might provide their date-of-birth as an answer to a question. And if the application that is protected by the authentication system has no requirement to know such information, this risk will be highlighted.
- Human Computer Interaction (HCI) Evaluation – Extending the previous analysis, an evaluation will be performed that includes data provided by test subjects in an attempt to strengthen (or refute) our earlier analysis. The evaluation will involve selection of a small number of representative KBA authentication solutions, development of a prototype for each solution, and soliciting a user base (likely Students, Staff and Faculty at the University of Edinburgh) to participate in a experiment (likely involving periodic use over several months) gathering empirical data regarding their use of the prototype(s). Users will be asked to provide us with their information as used for authentication to the prototype(s), and also complete a survey their satisfaction with the usability, and concerns with the privacy of information used with the system. With the collected data, we hope to be able to refine our earlier analysis with empirical data.

**Stage 3 – Recommended Improvements**
The third stage of the programme of work will involve the development of improvements to the previously analyzed and evaluated KBA solutions. To potentially improve the security of the solutions, it will be important to identify questions, or classes of questions, that solicit answers with a sufficient level of entropy, but also answers whose values cannot be easily located. To potentially improve upon the usability of the solutions, options for error-tolerant answer entry will likely be considered, as well as options for responding to only a subset of the questions. It may also be possible to identify particular classes of knowledge-based information that supports improved recall for the user. To potentially improve the privacy of the solutions, it may be possible to develop criteria that would be identify the classes of questions that would avoid collection of personal, sensitive information.

Similar to Stage 2 above, our improved solutions will be analyzed and evaluated in order to provide demonstrable evidence of improvement.

**Stage 4 – Project Wrap-Up**
The fourth stage of the programme of work will consist of the completion of the final research and project management deliverables that will document the achievements of our research, and identify future work.

# 4 Beneficiaries and Dissemination of Results

The short-term impact of the proposed work will be to provide a clear validation of the level of security, usability and privacy for current KBA techniques used widely in industry. In response to shortcomings, our improvements will serve to provide and new solutions for which the levels of compliance to security, usability and privacy requirements will be better known. In addition, we hope to provide solutions to communities that might have, till now, rejected current KBA techniques (for example, for the purpose of account recovery).

The primary beneficiaries of the proposed research are

- Designers of KBA solutions, thereby supporting wealth creation based upon our analysis and designs.
- Users of information systems protected by KBA solutions, for example, to access information-based systems, so as to improve the quality of life of individuals, including UK citizens.
- Other researchers, such as the Human Centred Systems Group led by Professor Angela Sasse at University College London, in order to aid in the development of further, improved solutions.

We anticipate that our results will be of benefit to the owners of information systems in both government and industry. As part of our project, we hope to make contact with applicable UK government and industry

organizations and inform them of the intent, and draft results of our analysis.

Partly motivated by our assumption that the industry community has simply not published their analysis of the Knowledge-Based Authentication (KBA) techniques they use to protect their applications, we believe that our beneficiaries would be best served by a full disclosure and dissemination of our research in the public domain. This would not only allow organizations to develop their own secure, usable and privacy-friendly KBA solutions, but allow others to continue to evaluate and validate existing and newly deployed KBA solutions.

We expect that an initial dissemination of our results will be to fellow researchers that would be interested in reviewing our draft results. Such validation is important to validating our results, but also often stimulates collaboration and development of further results. In particular, the Visiting Fellow (Just) has been invited to present this research to the Human Centred Systems Group led by Professor Angela Sasse at University College London.

# 5 List of References


[Elli00] C. Ellison et al., “Protecting Secret Keys with Personal Entropy,” J. Future Generation Computer Systems, vol. 16, no. 4, 2000, pp. 311–318.

[Fryk01] N. Frykholm and A. Juels, “Error-Tolerant Password Recovery,” Proc. ACM Conf. Computer and Comm. Security (CCS’01), ACM Press, 2001, pp. 1–9.

[HAC97] A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.

[Haga91] W. Haga and M. Zviran, “Question-and-Answer Passwords: An Empirical Evaluation,” Information Systems, vol. 16, no. 3, 1991, pp. 335–343.

[Just03] M. Just, “An Overview of Public Key Certificate Support for Canada’s Government On-Line (GOL) Initiative,” *Proceedings of the 2nd Annual PKI Research Workshop,* National Institute of Standards and Technology (NIST), April 2003.

[Just04] M. Just, “Designing and Evaluating Challenge Question Systems”, in *IEEE Security & Privacy: Special Issue on Security and Usability*, September/October 2004 (L. Faith-Cranor, S. Garfinkel, editors), p. 32-39.

[Just05] M. Just, “Designing Authentication Systems with Challenge Questions”, in *Designing Secure Systems That People Can Use*, O’Reilly, Laurie Faith-Cranor, Simson Garfinkel, editors, 2005.

[OGor04] L. O’Gorman, S. Begga, and J. Bentley, “Call Center Customer Verification by Query-Directed Passwords,” Proc. Financial Cryptography 04, Int’l Financial Cryptography Assoc., 2004.

[Pone00] R. Pond et al., “Word Association Computer Passwords: The Effect of Formulation Techniques on Recall and Guessing Rates,” Computers and Security, Volume 19, Number 7, 1 November 2000, pp. 645-656.

[Spec94] Y. Spector and J. Ginzberg, “Pass-Sentence—A New Approach to Computer Code,” Computers and Security, vol. 13, no. 2, 1994, pp. 145–160.

[Zurk96] Zurko, M. E. and Simon, R. T. 1996. User-centered security. In Proceedings of the 1996 Workshop on New Security Paradigms (Lake Arrowhead, California, United States, September 17 - 20, 1996). NSPW '96. ACM, New York, NY, 27-33.

[Zvir93] M. Zviran and W. Haga, “A Comparison of Password Techniques for Multilevel Authentication Mechanisms,” The Computer J., vol. 36, no. 3, 1993, pp. 227–237.