Designing Authentication Systems with Challenge Questions

**Mike Just**

**Chapter Eight**

**What Is Your Mother’s Maiden Name?** “What is your date of birth?” Such questions are often used to authenticate an individual. The answers often represent information well known to the individual, but (one hopes) not so widely known so as to be available to a potential impersonator. These *challenge questions* require an individual to recall and present previously registered answers when authenticating.

In this chapter, I review the design and evaluation of authentication systems that use challenge questions and answers to identify or authenticate individuals. I pay particular attention to ensuring that the design satisfies the security, usability, and privacy requirements of the authentication system.

While systems today use challenge questions for recovering forgotten passwords, they can be used more broadly for other forms of authentication, such as routine user login. This chapter focuses on password recovery but considers other applications as appropriate.

**Challenge Questions as a Form of Authentication**

Most people are familiar with passwords as a form of authentication. Passwords or Personal Identification Numbers (PINs) are two examples of using “something you know”
in order to authenticate. Biometrics, such as a fingerprint or voice recognition, represent “something you are,” and a physical token, such as a bank card, represents “something you have.” These three “something” categories are the common means of classifying authentication techniques. Challenge questions (and their corresponding answers), like passwords, are a form of “something you know” because they represent information that is known to an individual.

When I refer to challenge questions in this chapter, I am referring to questions and answers that are deposited or registered by an individual. For the purposes of subsequent authentication, the questions are posed to the individual who is required to repeat the original answer as his response.

An alternative form of “challenge questions” does not require the explicit deposit of information by an individual. Rather, the individual responds to questions posed by the account manager. In the case of a financial institution, for example, the individual might be asked to provide the monetary amount of his most recent transaction. I refer to such information as shared secrets, as they represent information shared a priori with the account manager.

**Using Challenge Questions for Credential Recovery**

Security credentials such as passwords and physical security tokens are issued in a variety of situations. Sometimes an initial identification step takes place in which an individual must show a passport or other identification papers. In other cases—for example, when establishing an account to read a free online newspaper—individuals identify themselves, and no efforts are made to verify a claimed identity. If an individual should forget a password or lose a security token, he might be able to recover his lost credential by re-identifying himself to the credential issuer. However, this is often inconvenient and generally requires a physical rather than an online interaction. In cases where an individual initially identified himself without providing any identification papers, re-identification may not be possible without the use of a shared secret or challenge questions.

Differing somewhat from passwords that are often memorized by an individual in support of future, routine authentication, challenge questions are most often based upon information already known to the user. While password construction might similarly rely upon information known to an individual, password rules (e.g., requirements to include both alphabetic and numeric characters) typically necessitate some additional memorization. Thus, challenge questions are particularly well-suited for credential recovery, as they do not require individuals to memorize additional information that subsequently could be forgotten.
Using Challenge Questions for Routine Authentication

While offering some advantage for the purpose of credential recovery, challenge questions can also be used for the day-to-day authentication of an individual. However, challenge questions may be inconvenient for day-to-day authentication for a number of reasons:

- **A challenge question system may require an additional step to obtain the challenge questions.** Unlike a system whereby the individual submits his username and password in one step, when questions are specific to the individual, the username must be provided first and the appropriate questions retrieved and presented to the individual for his response. Such delay may be intolerable to individuals.

- **A challenge question system may choose not to obscure display of the answers.** To prevent “shoulder-surfing” attacks, credentials such as passwords are often obscured (e.g., each password character is replaced with a “*” when displayed on the screen). Because challenge questions may prompt answers that include varying capitalization and punctuation, challenge question systems often allow users to enter responses unobscured.

- **A challenge question system may use more than one question-answer pair.** In this case, the use of multiple questions will most certainly require more time for authentication, at least when compared to password authentication.

- **A challenge question system may make use of an “out-of-band” authentication step.** This might require, for example, sending mail to the individual’s address of record (e.g., his home address). Such a step may introduce unacceptable delay for routine authentication.

In addition, because the answers to challenge questions are not constructed in the same way as passwords (e.g., with no requirements for including punctuation, capitalization, etc.), the answers may be “dictionary searchable.” By using challenge questions for routine authentication, a system may give an attacker more opportunity to validate his answer guesses. A recovery process can be controlled more easily; because recovery attempts are less frequent, each such attempt can result in a notice to the individual account owner. A poorly designed challenge question system can dramatically weaken the security of an otherwise strong password system.

Criteria for Building and Evaluating a Challenge Question System

We begin our introduction to the design of challenge question systems by introducing criteria that are helpful in both their design and evaluation. These criteria relate to the privacy, security, and usability of the challenge question system.
Privacy Criteria

In environments that use personal information, it should be common practice to follow recognized privacy principles to protect answers to challenge questions. For the use of challenge questions and answers to authenticate users, one principle in particular seems relevant: collection limitation. This criterion serves to limit the collection of personal user information to what is necessary for the purpose of authenticating an individual. Adherence to this principle helps to ensure that only information necessary to support a suitable level of security and usability is maintained.

Designers should give particular caution to using questions that ask for personal information, such as “What is your mother’s maiden name?”, because the answer, while possibly obscured (hashed), will be stored at the account server. Preference should be given to asking nonpersonal questions, provided that they offer sufficient security and usability.

In addition, answers to challenge questions should be used only for the purpose of recovering user access to one’s account—conforming to a use limitation principle. If challenge questions are to be used for other purposes, individuals should be notified and their consent obtained. Furthermore, care should be taken when asking for answers that users may find sensitive. Therefore, best practice involves offering as much choice as possible (while maintaining a suitable security level) to individuals for question selection, allowing individual control over the answers that are provided.

Security Criteria

The security of a challenge question system is related directly to the confidentiality of the challenge question answers. Other properties such as integrity and availability are also important to the security of the overall system, but are not the focus of our framework. The following security criteria apply primarily to the content of individual questions and answers:

Guessing difficulty

Answers should be difficult to guess and have an answer space with a fairly uniform distribution. Questions that can be guessed successfully in a small number of attempts (for example, “What is your eye color?”) do not make good challenge questions.

Observation difficulty

The answers to challenge questions should be difficult for an attacker to retrieve or observe easily. In particular, the answers should not be available from public sources. Questions that individuals are often asked to answer, such as “What is your mother’s maiden name?”, do not pose much observation difficulty. Unlike guessing difficulty, a determination of a question’s observation difficulty is more subjective, as the difficulty

---

of determining the answer is dependent upon a number of factors (e.g., the availability of the answer). In addition, observation difficulty will differ for individuals that have different relationships with the user—for example, family, friends, acquaintances, colleagues, or strangers.

For an authentication system that consists of multiple questions, additional criteria should be considered, including the total guessing and observation difficulty for the entire set of questions. In addition, answers should be unrelated so that both their availability and entropy can be maintained independently when multiple questions are used. One way to support answer independence is to use independent questions (questions that would encourage the submission of independent answers).

**Usability Criteria**

The usability of a challenge question system is concerned with providing a user-friendly experience at the stages of both answer registration and subsequent answer presentation. The following usability criteria should be used when evaluating a challenge question system:

**Applicability**

The applicability criterion attempts to characterize the size of the target population for which a question might be applicable. For example, a question about pets would not apply to those individuals who have never owned a pet. Attempts should be made to support highly applicable questions, although not at the expense of other criteria. A sufficiently large quantity of possible questions can be used to ensure enough coverage across the population of individuals.

**Memorability**

An answer is memorable as long as the user is able to recall the answer. This generally implies that the answer would be personally significant. Information that is used frequently will be more memorable, indicating that answers reflecting the habits, activities, or practices of users provide suitable answers. For an answer with high recall, only the likelihood of recalling an answer rather than the likelihood of knowing some answer is considered for the memorability criterion. For example, although many people may not know their high-school locker combination, those who do know the combination are likely to be able to continue to recall the answer. However, such a question would be applicable (as discussed for the applicability criterion) to only a smaller set of individuals that would know of an answer.

**Repeatability**

There are at least two aspects of answer repeatability to consider. First, answers should have few syntactic representations. For example, a question involving an address might be answered with “St.” or “Street,” and the word contractions (e.g., “St.” versus “Street”) may cause a discrepancy. Second, answers should have a semantic value that remains the same over time. For example, questions about favorites may be susceptible to the answer changing over time. For this reason, questions asking for “favorites” should be avoided in favor of “first time” or perhaps “memorable” qualifiers.
Additional usability issues include the number of questions and answers stored and the number of answers required to authenticate. These issues are discussed further in the next section.

**Types of Questions and Answers**

In this section, we present a candidate classification of questions and answers. For both questions and answers, three different types are described: fixed, open, and controlled. Depending on the system needs, various combinations of questions and answers can be used.

**Question Types**

The two types of questions that are likely to be most familiar are fixed questions and open questions. A *fixed question* provides a list of preset questions to a user, where the user’s choice of question can be taken only “as is” from this list. At the other extreme is an *open question*, where a user has complete choice and control over the question; guidance as to the question construction may be provided to the user, but the user enters the question in free-form text.

A *controlled question* lies between the extremes of a fixed question and an open question; it is a question whose content is partially fixed, although modifiable by the user. Two potential variations of a controlled question follow:

- The fixed question might allow for additional text to be added, forming a modification of the original question. The modified question would be presented subsequently to the individual for authentication. For example, consider the following question that supports customized addition of a name by an individual to the original question. In this case, the individual would choose an appropriate Name along with his answer to the question, and at subsequent authentication, the question would be asked along with the individual-provided name.

  What is Name's middle name?

- The fixed question might support a combination with an optional user-provided hint, where the hint would be presented to the individual for authentication. For example, to the following original fixed question the user might provide the following hint: “Dog.” The question and the hint would be provided to the user at authentication time. For the user, this hint might indicate a special date associated with his pet dog, such as its date of purchase. In this case, during answer registration, the individual is asked to provide a hint, and during answer presentation, the hint is provided for the user as an aid to recalling his answer.

  **Answer Registration**
  What is a memorable date for you? Date
  Hint: Hint

  **Answer Presentation**
  What is a memorable date for you? Date
  Hint: Dog
Answer Types

A similar distinction applies for fixed answers, controlled answers, and open answers. A fixed answer set involves user selection of an answer from a preset list of answers. At the other extreme, an open answer involves a user manually entering his response. Guidance may be provided as part of answer registration, but the answer is entered in free-form by the user.

A subtle variation is a controlled answer, where the answer space is neither fixed nor open. Some ways in which this might be achieved are:

- Providing a fixed set of answers where the answer space is large enough so that most potential answers are allowed. For example, in case an individual answers with a geographic location, the answer may be entered using drop-down menus listing all possible cities, states/provinces, and countries for some region.
- The individual is able to enter an answer, but the format of the answer is controlled—answers that do not conform are rejected. For example, an individual might be asked to provide a memorable numeric value so that alphabetic and punctuation characters would not be permitted for inclusion in the answer text.

Designing a Challenge Question Authentication System

Previous sections presented options for question and answer types and criteria upon which to evaluate a challenge question system design. This section discusses the design of a complete authentication system.

Determining the Number of Questions to Use

Usability tends toward requiring fewer questions and answers. This lessens the recall requirements for an individual, and also introduces fewer repeatability mistakes. For reasons of security, however, it is often necessary that more than one question-answer pair be registered by a user. This is to ensure a sufficient difficulty for either guessing or observing the answer. To ensure a sufficient level of protection against guessing, the entropy for the answers should provide a level of security similar to that for routine authentication (that may be performed with a password). In situations in which no complementary security measures are used (see the later section, “Complementary Security Techniques”), the entropy for the answers should be at least that for the routine authentication. In terms of guessability considerations, the strength of a challenge question system can be measured explicitly against password-based authentication. For example, an 8-character password constructed from the set of 52 upper- and lowercase characters, 10 numbers, and 32 punctuation characters, has approximately $2^{52}$ possible passwords. An 8-character answer to a question that uses only lowercase characters has only $2^{38}$ possibilities.

Unfortunately, this number is misleading. Answers to questions cannot be expected to conform to the same, strict rules as for passwords; otherwise, the answers effectively
become passwords. Instead, we would expect many answers to be dictionary words. This can be a problem, as most dictionaries have only between $2^{16}$ and $2^{20}$ words, while studies show that many adults have vocabularies between $2^{13}$ and $2^{17}$ words. Thus, even with these extremely optimistic values, at least two questions would need to be asked to ensure security similar to that provided by an 8-character password. In addition, as discussed earlier under “Security Criteria,” observability is important, but unfortunately is less quantifiable.

When more than one question is asked, both the interface and the administrative storage for the answers should ensure that multiple answer attempts are all validated before an indication of success or failure is given. Through the interface, for example, if two questions are asked, an indication of success or failure should be given only when both questions have been answered. If not, an attacker can guess answers to one question at a time, even though the entropy level of only one question might not provide a sufficient level of security. Similarly, as with the storage of passwords, answers should be obscured (hashed), but additionally, when multiple answers are used, they should be obscured in such a way that if the obscured answers are compromised, an attacker would have to guess all answers before determining the success of his guess. This can be achieved, for example, by inputting all answers to a single hash rather than by separately hashing each answer.

Variations exist where the number of questions presented at recovery is less than the number of questions registered. There are at least two models:

- The user registers $n$ questions, but is presented only $t \leq n$ questions upon recovery. All $t$ questions must be answered properly in order for the recovery process to continue.
- The user registers $n$ questions, and is presented $t \leq n$ questions upon recovery. Differing from the previous model, only $r < t$ questions must be answered in order for the recovery process to continue.

The first model is an attempt to offer a level of security equivalent to that of $n$ questions, but to provide a usability benefit at the time of recovery, with fewer questions to the user. However, the usability benefits appear only to reduce the time required for recovery and do not affect the arguably more important concerns of memorability and repeatability (the user still has to remember the answers for $n$ questions, as it is not known what questions will be posed at recovery). Yet there is some benefit for users who register $n$ questions, but after a period of time happen to forget the answers to some of these questions. The purpose of the second option is to tolerate mistakes upon answer presentation. However, it seems that an additional question is being used to tolerate such mistakes whereas a more usable system might attempt to reduce the number of questions used.

For a set of candidate questions, some form of question grouping might be beneficial. For example, supposing that three questions are to be registered, it may be advantageous to require that one fixed and two controlled questions be selected, and for these questions,
that a combination of fact-based and opinion-based questions be used. Alternatively, questions might be classified based on their topic so that users might have to select one question that required them to enter a “date” response, while the second might require a numeric response and the third, an alphabetic response. Finally, if one can classify users’ questions based on their security strength, the system could offer multiple classes where a user must select one question from each class as part of registration.

**Determining the Types of Questions and Answers to Use**

The types of questions and answers used contribute to both the security and the usability of the challenge question authentication system.

**Determining the appropriate question type**

With fixed questions, individuals are not required to conceive of their own questions at registration, perhaps offering an advantage to some. An open question has potential for improved memorability and improved applicability for individuals who are better able to recognize information that is more memorable to them, and to construct an appropriate question from this information. However, asking for a completely open question might require too much novelty. A controlled question seems to support a reasonable compromise whereby only part of the question development is delegated to the individual. For example, a question may be as simple as “Enter a number that is memorable for you” (giving some content control and guidance for the individual), and the individual can provide the hint, “Grade 8 locker,” thereby providing some equivalence to an open question. However, controlled questions also share the weaknesses of open questions, as the question or hint entered can be insecure by providing too much guidance for the answer to an attacker. Notice, however, that the repeatability and the memorability of the hint are not a concern because the hint is shown to the user upon answer presentation.

With a fixed question, individuals are prevented from a potentially insecure question selection (e.g., “What color are my eyes?”) whose answer space is exhausted easily, thereby providing a security advantage to an attacker. With an open question, individuals might select a question that is potentially insecure, although capable individuals are able to select more secure questions—for example, individuals are able to customize questions directly related and meaningful to their childhood. In addition, with open questions, individuals can form associative word pairs (e.g., the word “cat” might associate with the word “my pet,” or possibly with “shedding”).

When developing questions for a challenge question system, further distinctions can be made. One such distinction is that of fact-based versus opinion-based questions.\(^2\) Fact-based questions relate to factual statements regarding an individual. Such questions might be

---

expected to have less varying answers over time, and can be constructed as such (e.g., by asking for the first place the individual lived rather than his most recent residence). Care must be taken, though, as the answers to such questions (involving factual information about a user) might be more readily available to an attacker. Opinion-based questions relate to beliefs an individual has, and thus may be more susceptible to change over time. However, they should be less pervasive than the answers to fact-based questions, as opinions might be less frequently presented and recorded as part of the individual’s day-to-day activities.

**Determining the appropriate answer type**

Individuals can be prevented from selecting insecure answers if the system requires choosing an answer from a set of fixed answers. Such systems must be designed to disallow answers that would be very common and thus easily guessed by an attacker. However, memorability and repeatability may be hampered if there is no unique answer to satisfy an individual’s preference (either the individual’s first choice is not available, or more than one satisfactory choice is available). With open answers, larger variation in the answer space is provided, although for certain questions, a user would be able to select highly probable answers. Memorability may be better than with fixed answers, although repeatability can be problematic if the registered answer is ambiguous (e.g., “St.” versus “Street”). Controlled answers offer an alternative whereby a large answer space can be used, but control over the possible values improves repeatability. There do not seem to be any significant security advantages offered by using a controlled answer instead of by supporting a large answer space.

An interesting option is supported with answers whereby the answer registered and the answer presented need not be of the same type. Two such options are:

- **Fixed answer at registration; open answer at authentication.** When registering his question and answer, the individual is provided a fixed answer set corresponding to the question. However, at subsequent authentication, an open answer is designed, allowing the individual to enter his response, rather than choosing from a list. Still, as noted earlier, fixed answers at registration are problematic.

- **Open answer at registration; fixed answer at authentication.** When registering his question and answer, the individual provides a free-form response. However, at subsequent authentication, a fixed list of answers is provided, one of which is the correct answer originally chosen by the individual. This option offers improved repeatability and can be an advantage to individuals with poor memories.

Expanding upon the open-fixed option, a likely implementation might involve the storage of a set of “fake answers” along with the user’s given answer upon registration. At answer presentation, the user’s answer would consistently be presented along with the same set of fake answers. There are numerous issues to consider regarding the secure implementation of such a system. In particular:
• The “fake answer” set must not be repeated across users; otherwise, an attacker could easily determine the fake answer sets (and thus eliminate and recover the user’s submitted answer) by attempting to recover two or more users.

• The fake answer set must be consistent from one recovery attempt to the next; otherwise, an attacker could identify the user’s answer as the only consistent answer across a number of recovery attempts.

• The fake answer set must be changed should the user choose to modify his submitted answer; otherwise, an attacker (aware of a potential answer update) could determine the user’s answer from the variance in the answer sets from before and after the update.

Care must be taken in the selection of the fake answer sets for each user so that the user’s submitted answer is sufficiently concealed by the fake answers. For example, suppose that the user is asked the question “What is your favorite fruit?” but answers with the word “mushroom.” In this case, if only fruits were provided as part of the fake answer set, the user’s submitted answer would be easily distinguishable. Optionally, “incorrect” fake answers might be provided in order to anticipate any user variance and serve to confuse would-be attackers. Finally, two security problems present themselves with this scenario:

• The size of the fake answer set should be large enough to resist exhaustive guessing attacks against the individual user.

• The user’s submitted answer must not be hashed, as it must be presented to the user as part of answer presentation. Thus, while great care must be taken for this solution, it does offer an interesting variation.

**Complementary Security Techniques**

In addition to the construction, evaluation, and grouping of questions, additional techniques can be used for authenticating individuals, some of which are more suited to a recovery system than to general user authentication. Most notably, mailing to an address of record is a useful tool. For example, if the address of record is an email address, then as part of the recovery process an appropriate message can be emailed giving instructions. Some recovery systems will even choose to rely only upon a mailing, and not include any additional authentication (e.g., with a challenge question). While it is certainly possible for an attacker to intercept unprotected email (if an individual needs to recover, he likely won’t have key material to support a protected email message), the decision to use an additional factor is a risk management decision. When combined with a challenge question recovery system, an additional factor is used, and an email might be sent immediately after the user has answered the challenge questions successfully. By using an address of record, additional security is provided, as other security precautions are typically in place to control access to that address. Adding another communication step does impact usability; however, the extent to which this is true depends primarily upon

---

the amount of time required for this step to be completed. For some accounts, such latency may not be tolerable.

In addition, there are additional security measures that can greatly improve the usability of a challenge question system by reducing the security rigor that is applied to each question and possibly reducing the number of questions. These include:

• A system lockout feature whereby access to the recovery functionality would be reduced or removed after a number of failed attempts.

• A “graduated lockout” feature that would reduce access over time, perhaps locking out recovery for a fixed period of time after some number of failed recovery attempts, and fully blocking the recovery after some number of temporary lockouts.

Of course, the denial-of-service implications of using such features must be considered carefully. Reverse Turing Tests (e.g., CAPTCHA) help reduce the likelihood of success for automated attacks. Client puzzles offer a variation for limiting the effectiveness of denial-of-service attacks, whereby the client is required to perform additional computations before his request can be processed.

Some Examples of Current Practice

Challenge questions are used at a variety of web sites, often in combination with additional protections such as mailing to an address of record (typically an email address). For example, web email sites such as Yahoo! and Hotmail, and e-commerce sites such as Amazon, eBay, Chapters, and FutureShop, each use challenge questions in support of account recovery. In addition, online banking services similarly support a challenge question system.

From a privacy point of view, personal information is sometimes used as part of identification during recovery for some of these systems. Several banking sites use personal information (shared secrets) as part of account recovery. This is perhaps not too surprising, as the personal information used was related directly to information already retained by the banks. However, some of the web email sites, for example, do collect additional personal information (such as a date of birth) with the apparent, sole purpose of recovery.

From a security point of view, of those solutions in which a user registers a recovery question, only one such question is registered. In most cases, the use of personal

---

4 The CAPTCHA Project; http://www.captcha.net/.


A candidate challenge question system, based upon the framework in this chapter, was recently designed in support of Canada’s Government OnLine solution.\(^a\) Input to some of the design decisions came from a focus group consisting of 17 individuals from the general population that had Internet experience. Compared to a previous five-question system that was perceived negatively, participants of the current group appreciated the following three-question system:

- **Question 1.** Consists of 15 fixed questions, where the focus group input was used to determine several of these questions. The corresponding answer is open, both at registration and recovery. Some of the fixed questions proposed for this fixed list include: "What was my first pet’s name?" "Where did I first meet my significant other?" and "What was the last name of my childhood best friend?"

- **Question 2.** Consists of a controlled question, "Please choose a person who is memorable to you," and an open hint. Originally, a fixed hint was used, but participants were not comfortable with the choices it offered, as they had difficulty mapping their desired hint to a single selection of a fixed hint.

- **Question 3.** Consists of a controlled question, "Please choose a date that is memorable to you," and an open hint. The corresponding answer is controlled at both registration and recovery, consisting of drop-down selections for each of year, month, and day.

Free-form answers are normalized, removing whitespace, some punctuation, and capitalization. A confirmation page is displayed to confirm the user’s answers. Some additional lessons learned from the focus group include the following:

- Although questions related to “first-time” events are good for repeatability, they can be more difficult for older users to recall.

- Regarding questions with calendar date answers, participants indicated an inability to recall more than a half-dozen dates. However, even in this situation, such a question offers strength against a random attack, while being more susceptible to a targeted attack. Thus, additional questions and/or complementary security techniques should also be used.

- Although participants indicated a preference for open questions, the candidate list of questions they provided did confirm the designers’ assumptions that an insufficient level of security would be attained for open questions.

information, or mailing to an address of record, is used to provide additional security. In cases where recovery questions are used, users are asked to choose from a list of questions (an option described earlier as a “fixed question”).

About the Author

Mike Just is a policy and business strategist with the Canadian Federal Government. He is also an adjunct professor at Carleton University. His interest is in ensuring the delivery of secure yet usable online solutions for government. Prior work includes federal government IT security policy development, and work as an information security specialist at Entrust. He holds a Ph.D. in computer science from Carleton University and is active in the computer-security community.

http://www.scs.carleton.ca/~just/