A Comprehensive Analysis of Security Questions in Web Authentication

XIN SUN

A THESIS

IN

THE DEPARTMENT OF

CONCORDIA INSTITUTE FOR INFORMATION SYSTEMS ENGINEERING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF APPLIED SCIENCE

INFORMATION SYSTEMS SECURITY

AT CONCORDIA UNIVERSITY

MONTRÉAL, QUÉBEC, CANADA

DECEMBER 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Xin Sun**

Entitled: **A Comprehensive Analysis of Security Questions in Web Authentication**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

Information Systems Security

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

Dr. Jeremy Clark _____ Chair

Dr. Mohammad Mannan _____ Supervisor

Dr. Amr Youssef _____ Supervisor

Dr. Jeremy Clark _____ Examiner

Dr. M. Zahangir Kabir _____ Examiner

Approved by _____

Dr. Chun Wang, Director
Concordia Institute for Information Systems Engineering

_____ 2024

_____

Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

A Comprehensive Analysis of Security Questions in Web Authentication

Xin Sun

With the growing prevalence and sophistication of Internet services, user account security has become a critical concern. Security questions, widely adopted as a secondary authentication method, play a pivotal role in various online services. Although research on security questions has a long history, key gaps remain, particularly concerning user perceptions about security questions and the requirements used by websites for selecting and answering security questions. In this thesis, we address these gaps through a two-part study: (1) a comprehensive user survey ($N = 292$) that captures insights from a diverse and largely representative sample of the US population and (2) an analysis of an extensive set of 26 security requirements across 73 websites, also aiming to uncover security practices and weaknesses in their authentication systems (i.e., answer length restrictions). Additionally, we gather and analyze common online security questions (totaling 1913 questions) across several dimensions, including memorability, consistency, applicability, confidentiality, and specificity.

Our findings reveal previously unreported user misconceptions, such as users' believing that websites already possess correct answers to personal security questions. We also find that many websites allow insecure practices, such as accepting single-character, offering limited question choices, or identical answers for multiple security questions. By addressing both user perceptions and website security requirements, we provide a comprehensive understanding of the weaknesses in current security question practices and contribute to the discourse on improving authentication methods.

Acknowledgments

Throughout the process of writing this thesis, I have come to deeply appreciate the invaluable support behind every researcher. It is with heartfelt gratitude that I extend my sincere thanks to those who have supported and inspired me throughout my research journey.

First and foremost, I am immensely grateful to my advisors, Dr. Mohammad Mannan and Dr. Amr Youssef. Their expertise, patient guidance, and unwavering encouragement have been indispensable to my research path. From experimental design and data analysis to the final stages of writing, their insightful feedback and steadfast support have been invaluable. I am especially thankful for the financial support they provided, which allowed me to complete my experiments and research smoothly.

I also wish to thank all members of the Madiba Security Research Group. In this vibrant and innovative team, I found endless inspiration and assistance. Their professional insights and generous collaboration played a crucial role throughout my research.

Lastly, I must extend my deepest gratitude to my family. Their love, encouragement, and support have been my driving force. They have steadfastly stood by me in my pursuit of my academic dreams, offering both emotional support and material backing.

Thank you to everyone who contributed to and supported this research. Your help has been instrumental in bringing this work to fruition.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As digital life becomes more integrated into our daily routines, personal sensitive information and assets are increasingly stored and processed online. Whether it is online banking, government services, or travel-related services (e.g., airlines), protecting user accounts and the information behind them has become crucial. There exist multiple ways to do so, the most popular is the use of passwords, despite the repeated attempts to replace them with alternatives (most recently with passkeys) [32]. Security questions (SQs), also known as personal knowledge questions (PKQs), is another way to authenticate users. Generally, during account registration, users are required to select a few questions and set their answers and then, during usage (e.g., login/account recovery), to answer one or multiple of these questions. The core idea behind SQs is to confirm the user's identity by asking questions that only the legitimate user would know the answers to. This approach offers simplicity and convenience, allowing users to regain access to their accounts without remembering complex passwords.

Motivation. Some major online services (e.g., Google, since 2014) dropped the use of SQs primarily for security reasons.[1] However, many websites still rely on questions, e.g.,

[1] According to an industry survey, 22% of US adults suffered online account takeover, and among the victims, 36% of them had SQs set on the compromised accounts; see https://www.security.org/digital-safety/account-takeover-prevention/.

"What is your mother's maiden name?", which is easy to obtain or guess through public sources. Additionally, users tend to choose simple and easy-to-remember answers, further increasing the risk of unauthorized access to their accounts. In 2017, the US NIST issued a recommendation against the use of SQs for specific use cases, including login and account recovery (Special Publication 800-63-3). The design and evaluation of SQs, from both security and usability perspectives, have been studied in prior work (e.g., see the 1991 paper by Haga and Zviran [14], and several other publications since then [3, 15, 29, 27, 17, 19]). However, as apparent from our literature review, several important research questions remained unaddressed. There is a lack of comprehensive collection and systematic analysis of the SQs used by online services (e.g., types of questions), of the types of services that use such questions (e.g., banking and transportation), and of their usage of these questions (e.g., number of questions). Such data collection could make it easier to explore the security (e.g., entropy/guessability) of SQ answers and the usability of SQs. Even more critical for security are the requirements, as imposed by websites, of SQ answers that largely remained under-explored. Existing studies also lack connecting user perceptions of SQs with real-world question-and-answer requirements that could identify important gaps in the security provided by SQs vs. users' mental models. We believe all these aspects demand a closer look, as many users are still faced with SQs in their day-to-day use of the web.

1.1 Problem Statement and Analysis Overview

Problem statement. In short, we address three critical problems associated with SQs: (1) the lack of a comprehensive and diverse repository of security questions; (2) the lack of a comprehensive and systematic analysis of real-world requirements for both security questions and their corresponding answers; and (3) the limited understanding of user perceptions regarding security questions, particularly from a representative user population.

Overview of our work. To provide a comprehensive understanding of SQs, we explore the above-mentioned research gaps, through the combination of a manual collection, categorization, and requirements analysis of 212 websites that offer SQs and of an online survey with a representative sample of the US population (N = 292). A significant dataset of SQs is compiled from survey responses and manual web searches, facilitating detailed categorization and evaluation. We base our requirement analysis on the recommendations provided by the OWASP Foundation [25] (for legacy use cases for SQs, as OWASP also recommends following NIST guidelines). In particular, we assessed whether the SQs possess the desired characteristics (i.e., memorable, consistent, applicable, confidential, specific). In addition to OWASP's recommendations, we included a few additional requirements–some borrowed from password research [34], notably how special characters are handled and how responses to SQs are validated (e.g., case sensitivity, handling of spaces, punctuation, and symbols). Besides collecting names of websites that use SQs, our survey delves into users' perceptions of and insights about SQs. More specifically, we collect data about (1) users' experiences and practices (e.g., prevalence and purpose of SQs, strategies for selecting SQs); (2) users' strategies for answering (e.g., providing truthful vs. untruthful answers, attitude towards answering privacy-sensitive questions); (3) answer recall (e.g., mechanisms used for recall, forgetting answers); (4) perceived effectiveness of the protection level provided by SQs and the difficulty of guessing answers by various adversaries; and (5) preference of security mechanisms (e.g., comparing SMS vs. SQs, different types of SQs). Including both closed and open-ended questions, the survey provides the most comprehensive understanding of users' perceptions and insights on SQs compared to existing work.

1.2 Contributions

We make several key contributions to the understanding and improvement of SQs as an authentication mechanism:

(1) We represent the evaluation of 1913 SQs, providing a rich data foundation for understanding the diversity and effectiveness of these questions. SQs have been meticulously categorized and analyzed across five key dimensions, revealing common issues and challenges in the design and implementation of SQs.

(2) A detailed analysis of security question requirements across websites is conducted, which highlights the disparities in the application of SQs across different websites, providing an empirical basis for optimizing security strategies in online services. Additionally, we demonstrates the significant security risks caused by website administrators not following established guidelines, such as those from OWASP, and the continued use of easily guessed or overly personal questions (e.g., "mother's maiden name").

(3) Our findings highlight critical weaknesses in the current use of SQs, such as users' tendencies to provide consistent answers across services and misconceptions about service providers possessing the correct answers. We address a gap in prior research by surveying a diverse and representative sample, thus offering new insights into user perceptions, practices, and misconceptions about SQs.

These findings highlight the urgent need for stronger security measures in authentication processes and provide actionable insights for improving both user practices and the design of SQs on websites.

A note on authorship. A large part of the content from this thesis is currently under submission, which is the result of a collaboration with colleagues from the University of

Lausanne (Dr. Kévin Huguenin and Dr. Kavous Salehzadeh Niksirat), who primarily contributed in the design and analysis of the user survey.

1.3 Thesis Organization

The remaining chapters of this thesis are organized as follows. Chapter 2 provides background information about authentication mechanism, and related work. Chapter 3 and Chapter 4 respectively contain methodology and results of security question categorization and requirement analysis. Chapter 5 presents cooperative user study within methodology and results. Finally, in Chapter 6, we conclude with key takeaways, limitations, recommendations and potential future work.

Chapter 2

Background and Related Work

In this chapter, we briefly review the evolution of authentication methods, with a concise overview of multi-factor authentication and SQs. We also discuss prior research on SQs as an authentication method, examining their role in account recovery and verification, their security challenges, and users' perspectives on them. This chapter aims to provide background context for subsequent chapters, helping readers to better understand the security aspects of SQs and the value of research on improvement methods.

2.1 Development of Authentication Methods

With the rapid advancement of internet technology, user authentication methods have evolved significantly. Starting with simple password-based mechanisms, this method has shown vulnerabilities such as password leaks and susceptibility to brute-force attacks,

while it is convenient [18]. Digital certificates provide a Public Key Infrastructure (PKI)-based solution for online authentication, enhancing security against threats (e.g., man-in-the-middle attacks and data breaches) [5]. Biometric authentication, using inherent physical characteristics (e.g., fingerprints and facial recognition), offers users a secure and convenient verification method. While these technologies have greatly strengthened authentication security, practical challenges and limitations remain in their implementation.

2.2 Multi-Factor Authentication

Multi-factor Authentication (MFA) is an authentication method that requires the user to provide two or more verification factors to gain access to a resource such as an application, online account, or a VPN. MFA is a core component of a strong identity and access management (IAM) policy. Unlike single-factor authentication, which typically relies on a password alone, MFA incorporates two or more distinct factors—often categorized as something the user knows (e.g., a password), something the user has (e.g., a smartphone or security token), and something the user is (e.g., biometric data) [30]. This layered approach significantly reduces the risk of unauthorized access, as compromising one factor alone is insufficient to breach the system.

2.3 Security Questions

Security question is a security mechanism helping online service to confirm user's identity. According to answer sensitivity, it is classified into two parts, sensitive SQs and personal SQs. The questions including highly sensitive information, e.g., driver license number, are categorized as sensitive SQs. In contrast, questions related to personal history or family background, e.g., mother's maiden name, are personal SQs. Additionally, OWASP[25] categorizes SQs into two main types. For user-defined SQs, users select a

question from a provided list and supply an answer. This approach offers users some flexibility in question choice but depends on their ability to select details that are both memorable and secure. For system-defined SQs, the system leverages information already known about the user, reducing the risk of weak or easily guessed responses. This method bypasses the need for users to generate unique answers; however, it requires the system to maintain sufficient user data and assumes this information is difficult for an attacker to access.

2.4 Related Work

Academic research in SQs has a long history (e.g., see the 1991 study by Haga and Zviran[14]). However, compared to passwords, they have been studied far less extensively [7, 4]. Researchers from various communities use different terminologies to refer to, such as "SQs" [12, 15], "secret questions" [29], "personal knowledge questions" [3, 27], "question-and-answer passwords" [14], and "password choice and challenge questions" [17]. Hence, for consistency, we will use the term "Security questions (SQs)", regardless of the term used in the papers we describe.

2.4.1 Security Aspects and Users' Perspectives

We first summarize, in chronological order, the studies that focus on the analysis of SQs, from security and users' perspectives. Rabkin [27] focused on banking/financial websites in the US and categorized 215 SQs from 11 such websites. "Names" of friends and family and "Favorites" were the most common categories. Rabkin additionally discussed the use of personal knowledge questions as a fallback authentication method in online systems and 11 banking sites relied on SQs with serious usability and security weaknesses, including:

memorability issues, and unsuitability of some questions due to a lack of relevant user experience. The author analyzed how SQs were chosen both at setup time and at authentication time, including the number of questions offered by the services, the number of questions users set up, the number of questions answered at authentication time, and whether users could write their own questions. Furthermore, the author designed a questionnaire to ask 46 respondents about their habits in selecting and answering SQs. Only a small number of respondents (7%) were very concerned about the security of these questions. 70% respondents considered memorability very important instead of security when they selected questions. Also, 38% respondents claimed to "always" answer these questions truthfully, and 13% usually provided untruthful answers.

Just and Aspinall [17] collected 282 questions from the inputs of respondents (rather than from real-world websites). Analyzing possible answers by using a novel attacker model based on the knowledge level of potential attackers, they conclude that the answers to the majority of the questions would fail to resist attacks from strangers (in contrast to the beliefs of the respondents).

Schechter et al. [29] conducted a user study to measure the reliability and security of SQs used by four major webmail providers (AOL, Google, Microsoft, and Yahoo!). They asked respondents ($N = 130$) to answer these questions and then asked them to guess their partners' answers to these questions. They find that acquaintances (e.g., spouse, relatives, and friends) guessed 17% of the answers; respondents also forgot 20% of their own answers within six months. The author suggests that webmail providers should reconsider the use of personal authentication questions as they may not be a sufficiently secure authenticator.

Just and Aspinall [16] designed an experimental study to analyze the usability and security of user-chosen challenge questions used for authentication. They categorized answers by type (e.g., names, dates) and analyzed answer lengths for security assessment. Approximately 75% of answers were recalled accurately, but a significant portion showed issues,

e.g., misspellings or substitutions, indicating potential memorability challenges. The authors found that answers, often less than eight characters, exhibited low entropy, making them vulnerable to brute-force attacks. Over 88% of respondents believed their questions would be difficult for strangers to guess, but entropy calculations showed this confidence was misplaced, especially given the predictable nature of common question types.

Ullah et al. [35] conducted a study ($N = 23$) to evaluated the efficiency and effectiveness of profile-based SQs in an online examination setting. Their findings show that specific question types influence usability metrics (e.g., response time and accuracy). Academic and personal questions demonstrated higher effectiveness due to clarity, while ambiguous or complex questions, such as contact-related items, reduced efficiency and led to authentication errors. The authors also indicate that certain personal and public knowledge-based questions are vulnerable to guessing attacks.

Analyzing a large real-world data set on security and memorability of personal knowledge questions from Google and 11 million account recovery claims, Bonneau et al. [3] found that SQs have poor security and memorability. They also find that several answers expected to be unique (e.g., phone numbers) were identical across different users, significantly compromising the security of the questions. A survey of 1,000 respondents revealed that the primary reason for providing untruthful answers was that respondents believed the answers would be difficult to guess (37%), did not contain their privacy information (31.9%), and remembered answers more easily (15%). However, they also reported untruthful answers not only severely compromise security but pose significant challenges to memory issues. Additionally, the authors reported that 62.8% respondents never considered the possibility of SQs being compromised.

Golla and Dürmuth [13] analyzed a leaked database of 3.9 million answers to SQs from AshleyMadison.com (Aug. 2015). They find that the question "What is Your Favorite Sports Team?" is the least secure, using statistical entropy to measure the success rate of

guessing answers for four different questions as used by the website. Furthermore, they find the security of knowledge questions is slightly higher than the four-digit user-chosen PINs.

Golbeck and Li [12] collected screenshots of SQs via crowd-sourcing, and identified 283 unique SQs (total 607) from 52 unique websites. They summarized a set of categories and themes: "Locations", "Firsts", "Time and Dates", "Numbers", "Dreams and Aspirations", and "Other", in addition to "People's Names" and "Favorites" (cf. Rabkin[27]).

Dhekane [6] conducted a survey ($N = 153$) to analyze the security and usability of SQs, and they find that many respondents prefer customized questions over those offered by the services. To further study the memorability of both types of questions, the author surveyed students with knowledge in security ($N = 48$). The majority of the students found it difficult to recall the answers.

Micallef and Arachchilage [24] developed a survey ($N = 30$) to study how users chose SQs, the memorization strategies employed for answers, and user perceptions of the security and memorability of their answers. The respondents were divided into two groups. One group was required to answer by themselves, and the other group was required to select answers from a system-generated document. Respondents who answered by themselves (as opposed to system-generated answers) chose questions (1) whose answers were based on respondents' lives; and (2) whose answers were fixed. According to respondents using answers generated by the system, they employed questions due to answers, e.g., not being public, difficult to guess, based on related objects in the life. Memorable answers were the same factor in both groups. The authors find respondents selecting their own answers are confident to remember the answers without any strategy. In contrast, those answering with the system-generated answers not only share this confidence but also frequently repeat the answers or wrote them down on paper for memorization. 24 out of 30 respondents stated they used the same questions in different online servers. Additionally, 28 out of 30

respondents in the survey indicated that they selected the same answers to the questions in various services because of answers memorability, answers uniqueness, answers being truthful, and answers being fixed.

More recently, Lassak et al. [19] conducted a user study ($N = 97$) over a period of 18 months to compare various fallback authentication mechanisms (email, SMS, SQs, and trusted contacts). The authors found that the percentage of respondents (57%) who used SQs to successfully recover passwords was the lowest, compared with those who employed SMS (92%), email (100%), and trusted contacts (83%). They also reported that some respondents (not specified in the paper) used untruthful answers as a strategy to boost security of their answers and protect privacy.

In another survey ($N = 281$), Höltervennhoff et al. [15] reported that respondents indicated that SQs have poor performance, both in terms of security and usability, compared to other account recovery methods (e.g., recovery code, e-mail recovery, and trusted devices).

2.4.2 Improvements of Security Questions

To improve the security and usability of knowledge-based authentication systems, Senarathet al.[31] designed an interface with mnemonics to help users set up secure answers. They measured answer strength via five parameters, similar to passwords : capital letter, digit, special character, letter, and character length. They also implemented a strength checker to offer feedback to users about the strength of their answers. Also, AlHusain and Alkhalifah [2] introduced SQs based on user behavior and personal preferences. Through an experiment ($N = 23$), they evaluated the security and usability of answers to the questions they designed. They reported that using user behavioral details in selecting topics for SQs are effective, where the questions achieve both high recall rates and resistance to guessing.

2.4.3 Research Gap

Compared to previous work, our survey is more comprehensive ($N = 292$, with over 30 questions), and the only one with a largely representative sample of the US population (w.r.t. location, age, sex, and ethnicity). We not only collected the respondents' perceptions of SQ memorability [29] and their strategies for answering questions ([24] and [3]) but also their insights of guessing attacks from more (nine) entities than [17] (three entities), SQ effectiveness and security mechanism preferences, and the evaluation of their general security behavior. We also analyzed a comprehensive set of security requirements (26) of SQs (following OWASP, and common requirements for passwords), for a set of 73 websites. Additionally, we gathered more questions (1913) from 194 websites with more logical and understandable categories than [12] (607 questions from 52 websites), [27] (215 questions from 11 websites) and [22] (8 questions from 3 websites).

Authors	Year▲	# Websites	# Security questions	Question-based requirements	Answer-based requirements	Comparison with password	User study coverage	# Respondents	Representative sample
Rabkin [27]	2008	11	215	✔	-	-	◔	46	-
Just and Aspinall [17]	2009	-	282-180	-	-	✔	◑	94-60	-
Schechter et al. [29]	2009	4	29	✔	✔	-	◑	130	-
Bonneau et al. [3]	2015	-	12	-	✔	✔	◑	1000	-
Golla and Dürmuth [13]	2016	1	4	-	✔	✔	○	-	-
Senarath et al. [31]	2017	-	-	✔	✔	✔	○	-	-
Dhekane [6]	2020	-	7	✔	-	✔	◑	153-48	-
Golbeck and Li [12]	2020	52	607	-	-	-	○	-	-
Micallef and Arachchilage [24]	2021	-	15	-	-	-	◕	30	-
AlHusain and Alkhalifah [2]	2022	-	10	-	✔	-	◕	23	-
Höltervennhoff et al. [15]	2024	-	-	-	-	-	◕	281	-
Lassak et al. [19]	2024	-	8	-	-	-	◕	97	-
Our work	-	212 †	1913	✔	✔	-	●	292	✔

Table 2.1: High-level comparison with the state-of-the-art. Notation and notes include: "-": not applicable/addressed; "∗": the number in a follow-up survey; "†": the number of unique websites we analyzed (194 websites for SQ categorization and 73 for SQ requirement analysis); "Question-based requirements": the features/requirements about the questions offered (e.g., the number of SQs); "Answer-based requirements": the requirements that users must fulfill for the answers (e.g., the minimum length of answers); "User insights": the author(s) collected user-related perceptions, behaviors, and insights, related to SQs. "○", "◔", "◑", "◕", and "●" represent the extent of user study coverage in the paper, based on the following scale: "○": *No user involvement*; "◔": *Minimal user involvement*; "◑": *Some user involvement, limited in scope*; "◕": *Strong user involvement, detailed insights*; and "●": *Comprehensive user involvement, covering multiple aspects*.

Chapter 3

Security Question Analysis Requirements: Methodology

In this chapter, we provide a detailed overview of the research methodology for our SQ requirements analysis. We begin by outlining the process of manually searching websites that utilize SQs. Next, we describe the criteria and steps used to categorize and evaluate the collected SQs. Finally, we conduct an in-depth analysis of the security requirements employed by various websites in implementing SQs.

3.1 Website Collection

We built our dataset of websites, which rely on SQs, from two sources: (1) the survey and (2) manual web searches. For the web search, we relied on Google, using the following templated search string: "`$term` `$category` websites online services", where `$term` was one of terms used to denote SQs (e.g., security question, personal knowledge question; five in total), and `$category` was either empty or picked from the list of typical services that rely on SQs (e.g., insurance, air travel; 32 in total; see Table 3.1). We manually handled results within the first five pages for each search string, checking whether we could

Category	Frequency
Banking Credit and Lending	n = 39, 26.7%
Law and Government	n = 16, 11.0%
Insurance	n = 8, 5.5%
Investing	n = 8, 5.5%
Financial Planning and Management	n = 6, 4.1%
Telecommunications	n = 6, 4.1%
Finance	n = 6, 4.1%
Health	n = 5, 3.4%
Computers Electronics and Technology	n = 5, 3.4%
Universities and Colleges	n = 5, 3.4%
Video Games Consoles and Accessories	n = 5, 3.4%
Government	n = 5, 3.4%
Air Travel	n = 4, 2.7%
News and Media	n = 3, 2.1%
Energy Industry	n = 3, 2.1%
Medicine	n = 2, 1.4%
Education	n = 2, 1.4%
Consumer Electronics	n = 2, 1.4%
Jobs and Employment	n = 2, 1.4%
Shipping and Logistics	n = 2, 1.4%
Food and Drink	n = 1, 0.7%
Accounting and Auditing	n = 1, 0.7%
Search Engines	n = 1, 0.7%
Faith and Beliefs	n = 1, 0.7%
Programming and Developer Software	n = 1, 0.7%
Web Hosting and Domain Names	n = 1, 0.7%
Fashion and Apparel	n = 1, 0.7%
Home and Garden	n = 1, 0.7%
Immigration and Visas	n = 1, 0.7%
Coupons and Rebates	n = 1, 0.7%
Social Networks and Online Communities	n = 1, 0.7%
National Security	n = 1, 0.7%

Table 3.1: All categories of websites (according to Similarweb).

set up SQs after creating accounts. For website categorization, we used Similarweb [33]. However, Similarweb did not provide categories for 23 websites, hence we had to manually categorize them. For searching, we used a fresh instance of Firefox v125.0.2 for Windows 11 with no plugins/extensions, using a commercial VPN service in the US. For each search, we cleared the state (e.g., cookies) of the browser and used a new private window. We used the same pseudo-information (e.g., name, date of birth)for all registrations.

For each identified website, we tried to register an account in order to collect data on the ways SQs are used on the website. We excluded websites that require real-world IDs (e.g., US SSN, client ID) or that do not involve SQs during registration. From the websites where we could register an account, we collected the complete list of SQs they used and other relevant aspects.

In the end, we processed our requirements analysis (26, see Table 3.2 and Table 3.3) at a total of 73 websites (including 72.6% websites ($n = 53$) from Google Search only, 47.9% ($n = 35$) from the survey only, and 20.5% ($n = 15$) from both); the rest of the websites required real-world IDs or other account details.

3.2 Security Question Categorization and Evaluation

We adopted eight question categories designed by Golbeck and Li [12] (e.g., “Firsts”, “Favorites”, and “Locations”) and also added a new category, “Seconds”.

Before categorizing questions, we rewrote each question (when needed), and merged those with the same phrasing. For example, we replaced “Name of first pet?” and “What is the name of your first pet?” with “What was the name of your first pet?”.

We came up with the following categories of questions.

(1) Names: This category includes names of people, pets, groups, and things produced or created by people; e.g., “What is your best friend’s name?”.

(2) Firsts: includes first entities, and experiences; e.g., “What was your first job?”.

(3) Favorites: includes favorite or least favorite items; e.g., “What is your favourite hobby?”.

(4) Locations: includes questions about cities, towns, or countries; e.g., “What city were you born in?”.

(5) Times: includes questions about times or dates; e.g., “In what month is your best friend’s birthday?”.

(6) Numbers: questions about digits only; e.g., “What was your childhood phone number?”.

(7) Aspirations: about people’s dreams and wishes; e.g., “What is your dream job?”.

(8) Seconds: questions about second items or experiences; e.g., “What was the name of your second pet?”.

(9) Others: includes the questions that did not belong to the above categories; e.g. “What is your password unlock code?”.

Based on the suggestions from OWASP, we evaluated collected SQs with five characteristics:

(1) Memorable: the user should be able to recall the answer, even years after creating their account.

(2) Consistent: the answer should remain unchanged over time.

(3) Applicable: the user must be able to provide an answer to the question.

(4) Confidential: the answer should be challenging for an attacker to uncover.

(5) Specific: the answer should be unambiguous and easily understood by the user.

To quantify the extent to which these characteristics are met, we established three evaluation criteria:

- Fully Satifies. It indicates that the characteristic fully satisfies all users, which means the question or its answer are universally appropriate across the user base.
- Partially Satifies. It indicates that the characteristic only applies to a subset of users, acknowledging that some SQs may be well-suited for specific user groups.
- Does Not Satisfy. It indicates the the characteristic is inadequate for all users.

For example, for the question, "What was the first name of your father", it fully satisfies *Memorable* and *Consistent*, as most users are unlikely to forget their father's name, even over time, which is generally a constant part of personal knowledge and is unlikely to be forgotten over time. Additionally, the question fully satisfies *Applicable*, given that nearly all users will know this information, and *Specific*, as the question is clear and unambiguous. However, it does not satisfy the *Confidential* characteristic. A father's first name is generally easy to discover through social media or public records.

3.3 Security Question Requirement Analysis

This analysis employed a multi-step method to investigate the characteristics and settings of SQs in online registration and usage contexts. After completing the registration process for 73 websites, we proceeded to log in and access the settings of each website.

In total, we examined 26 requirements for the answers to SQs. We selected these requirements from past analysis of passwords ([21]), relevant descriptions (e.g., *Your answers must contain at least three characters and contain no special characters (for example: %,*

Requirements (i.e., criteria)	Description
Allowance of customized questions	Does the website allow users to write their own questions?
Allowance of digits	Does the website allow users to input digits in an answer?
Allowance of punctuation marks	Does the website allow users to input punctuation marks?
Allowance of same answer	Does the website allow users to set same answer if multiple questions needed?
Allowance of same question	Does the website allow users to set same question if multiple questions needed?
Allowance of space characters in the settings	Does the website allow users to input space characters on the settings page (registration)?
Allowance of space characters in the usage	Does the website allow users to input space characters on the usage page (e.g. login page)?
Allowance of symbols in the settings	Does the website allow users to input symbols on the settings page (registration)?
Allowance of symbols in the usage	Does the website allow users to input symbols on the usage page (e.g. login page)?
Authentication flow	Does the website follow login flow in OWASP guidance?
Case-sensitive	Does the website treat uppercase and lowercase letters as distinct?
Fixed questions	Do the questions remain consistent until users complete verification if multiple questions are set?
Forgotten password flow	Does the website follow forgotten password flow in OWASP guidance?
Maximum length	What is the maximum number of characters in an answer a user can input?
Minimum length	What is the minimum number of characters in an answer a user can input?

Table 3.2: Description of the criteria used in the SQ requirement analysis.

Requirements (i.e., criteria)	Description
Minimum number of digits	What is the minimum number of digits in an answer a user can input?
Minimum number of letters	What is the minimum number of letters in an answer a user can input?
Minimum number of letters or digits	What is the minimum number of letters or digits in an answer a user can input?
Minimum number of lower case letters	What is the minimum number of lower case letters in an answer a user can input?
Minimum number of upper case letters	What is the minimum number of upper case letters in an answer a user can input?
Number of the questions offered in the source	How many questions does the website offer for users to select?
Number of the questions in the settings	How many questions does the website allow users to set up on the settings page (registration)?
Number of the questions in the usage	How many questions do users answer on the usage page (e.g. login page)?
Re-authentication methods	What method does the website implement for authentication when users edit their security questions?
Tries threshold	How many attempts does the website allow for an answer?
Usage	What is the usage of security questions on the website?

Table 3.3: Description of the criteria used in the SQ requirement analysis (cont.).

#, @)) from online services on registration pages, and the factors that can affect the complexity of answers.

We assumed that the answer fields on a website impose the same requirements irrespective of the chosen questions (from our manual observation of website code/behavior for several websites, this assumption appears to hold).

3.3.1 Security Questions: Quantity and Customization

We enumerated the SQs from the account settings page of each website; we also noted the number of required questions that a user must set, and whether user-generated customized questions are allowed. If multiple (predefined or customized) questions are allowed, we checked if the user can set the same question multiple times. Furthermore, we recorded the number of questions that users answered for authentication on the usage page (e.g., during login).

3.3.2 Requirements for Answers to Security Questions

We input a combination of letters, digits, spaces, and special characters (punctuation marks and symbols) to test the following requirements for answers:

Minimum and maximum answer lengths. We checked for both minimum and maximum length requirements by manual inputs (also guided by length parameters from registration or settings page code, if available). For the minimum length, we input just a single letter (‘ a’) first, and if needed, increased the length one-by-one (i.e., adding more ‘a’) until the answer was accepted. For the maximum length, we used a Python script to generate strings (with random letters) of a specific length and pasted them into the answer box to check answer acceptability. We employed binary search to determine appropriate length ranges, initially checking between 0 and 200, and then between 201 and 1000.

Input constraints for setting answers. We checked the following constraints: the minimum number of (1) letters, (2) digits, (3) characters (letters or digits), (4) upper and lower case letters; (5) punctuation marks; (6) allowing the same answer; and (7) allowing the digits. We tested all punctuation marks (˜!@#$%ˆ&*()_+{}|:"<>?`-=[]\;',./). If the combined marks exceeded the maximum length, they were divided into segments based on the maximum length. For the websites that did not allow punctuation marks alone, we added the letter 'a' and repeated it, if needed; e.g., we used 'aa?' to test '?' for a website where the minimum answer length is 3. Additionally, each mark was individually checked with the letter to fulfill the minimum length requirement. If a website allows multiple questions, we checked if the same answer was accepted for all questions. To verify if a website allows users to input digits, we repeatedly entered "1" as the answer, matching the minimum length requirement.

Answer constraints between setup and usage pages. We pre-set specific answers within the settings page and proceeded to the SQ usage page (such as the login page or password recovery page) to verify whether the answers set in the settings page corresponded with those in usage. We performed the following tests: (1) case sensitivity, (2) the use of spaces, (3) the use of symbols, and (4) the number of allowed attempts for an answer. To test case-sensitivity, we simply switched upper case letters with lower case ones (e.g., 'A' in place of 'a'). If such a modified answer was accepted in the usage page (or between different questions in the configuration page if the website allows multiple questions without same answers), we concluded that the website does not enforce case-sensitivity. To test the use of spaces, we set an answer string with letters (e.g., repeated 'a's), and then subsequently added a space at the beginning, end, and at the middle; in the usage page, we checked the modified answer with the spaces removed (each space placement was tested separately). To test the use of symbols, we set an answer with "■▲♥◆" symbols (letters are added to meet the minimum length requirement if needed); we then tested such an answer with

symbols in the usage page. To determine the number of allowed attempts for an answer, we repeatedly input an incorrect input answer until the service indicated an account lockout (e.g., *"Your user account has been locked temporarily. It will be automatically unlocked in 30 minutes."*), or provided a relevant message (e.g., with the threshold for wrong answers), or we made more than 10 failed attempts.

Additionally, we examined whether users are required to authenticate their identity before modifying security questions, aiming to prevent attackers from updating questions after gaining temporary access to user accounts. This assessment includes verifying whether websites ensure secure identity authentication through strong passwords, two-step verification, or multi-factor authentication, as well as whether systems implement account lockout strategies in response to consecutive authentication failures. Furthermore, we examined the login and password recovery mechanisms of websites. On the login page, we verified whether the authentication process follows recommended steps: first, the user enters their username and password; if correct, the system then prompts the user with a security question; finally, upon a correct answer, the user is logged in. For the password recovery interface, we similarly evaluated the process, specifically checking whether websites implement both email and security question verification. In cases where multiple security questions are configured but only one question requires answering, we examined whether the website restricts question modification until the user provides a correct answer on the usage page.

Chapter 4

Security Question Analysis Requirements: Results

In this chapter, we comprehensively present the research findings, which include the classification and evaluation of SQs, as well as an analysis of various websites. We begin by providing a detailed classification and evaluation of SQs. Subsequently, we analyze the diverse requirements employed by websites in the implementation of SQs, with a focus on the usability and security of the requirements set by each website. This analysis offers deep insights into the current practices surrounding the use of SQs.

We collected 146 unique websites from our user survey (from the 292 URLs provided by the respondents), and 86 unique websites from our web search; thus totaling 212 unique websites. We were able to register accounts on a total of 73 websites, on which we based our analysis. As we included website categories in our web searches, thus biasing the results, we reported only on the categories of the 146 websites collected through our user survey. The five most frequent categories for the collected websites are "Banking Credit and Lending" (n = 39, 26.7%), "Law and Government" (n = 16, 11.0%), "Insurance" (n = 8, 5.5%), "Investing" (n = 8, 5.5%), "Financial Planning and Management" (n = 6, 4.1%), "Telecommunications" (n = 6, 4.1%), and "Finance" (n = 6, 4.1%).

Category	Frequency (all)	Frequency (unique)
Aspirations	n = 32, 1.7%	n = 12, 1.9%
Favorites	n = 439, 22.9%	n = 160, 25.5%
Firsts	n = 605, 31.6%	n = 175, 27.9%
Locations	n = 329, 17.2%	n = 103, 16.4%
Names	n = 1373, 71.8%	n = 420, 67.0%
Numbers	n = 50, 2.6%	n = 29, 4.6%
Others	n = 23, 1.2%	n = 19, 3.0%
Seconds	n = 5, 0.3%	n = 3, 0.5%
Times	n = 44, 2.3%	n = 25, 4.0%

Table 4.1: Categories of the collected SQs.

	Frequency (all)		
Characteristic	Fully satifies	Partially satifies	Does not satisfy
Memorable	n = 789, 41.2%	n = 1119, 58.5%	n = 5, 0.3%
Consistent	n = 841, 44.0%	n = 1067, 55.8%	n = 5, 0.3%
Applicable	n = 1190, 62.2%	n = 723, 37.8%	n = 0, 0%
Confidential	n = 39, 2.0%	n = 1548, 80.9%	n = 326, 17.0%
Specific	n = 539, 28.2%	n = 1122, 58.7%	n = 252, 13.2%

Table 4.2: Five desired OWASP characteristics analyzed in all collected SQs.

4.1 Security Question Categorization and Evaluation

As for SQs, we collected a total of 1913 questions; after merging equivalent questions, we obtained 627 unique questions. More specifically, the most frequent questions are "What was the name of your first pet?" (n = 65, 10.4%), "What was the maiden name of your mother?" (n = 46, 7.3%), and "What was the middle name of your father?" (n = 39, 6.2%).

Original security questions. We labeled 71.8% questions (n = 1373) as "Names", 31.6% (n = 605) as "Firsts", 22.9% (n = 439) as "Favorites", 17.2% (n = 329) as "Locations", 2.3% (n = 44) as "Times", 2.6% (n = 50) as "Numbers", 1.7% (n = 32) as "Aspirations", 1.2% (n = 23) as "Others", and 0.3% (n = 5) as "Seconds", shown in Table 4.1. For the desired characteristics of the SQs (see Table 4.2), while 41.2% of SQs (n = 789) are easily memorable and 44.0% fully consistent (n = 841), 62.2% are broadly applicable (n = 1190). However, only 2.0% (n = 39) meet full confidentiality standards, raising

	Frequency (unique)		
Characteristic	Fully satifies	Partially satifies	Does not satisfy
Memorable	n = 181, 28.9%	n = 441, 70.3%	n = 5, 0.8%
Consistent	n = 227, 36.2%	n = 395, 63.0%	n = 5, 0.8%
Applicable	n = 388, 61.9%	n = 239, 38.1%	n = 0, 0%
Confidential	n = 24, 3.8%	n = 532, 84.8%	n = 71, 11.3%
Specific	n = 100, 15.9%	n = 416, 66.3%	n = 111, 17.7%

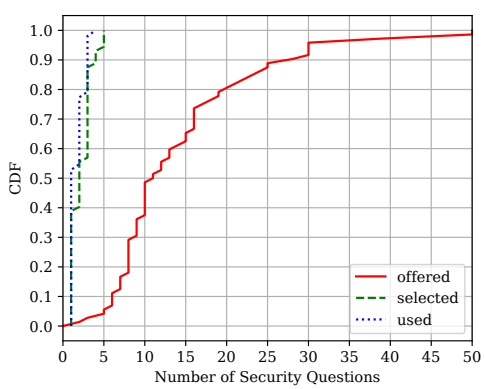
Table 4.3: Five desired OWASP characteristics analyzed in unique collected SQs.

significant security concerns. Additionally, approximately 28.2% of SQs (n = 539) were considered in specificity.

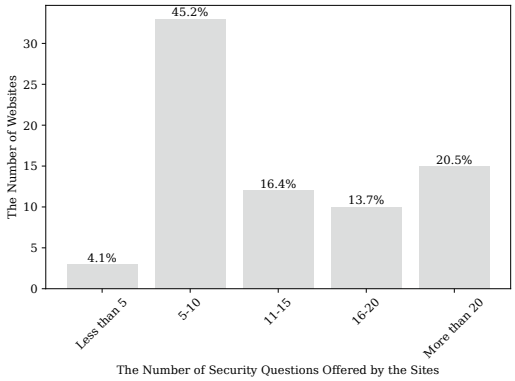
Rewritten security questions. After rewriting all the questions, we recorded 67.0% questions (n = 420) labeled as "Names", 27.9% (n = 175) as "Firsts", 25.5% (n = 160) as "Favorites", 16.4% (n = 103) as "Locations", 4.0% (n = 25) as "Times", 4.6% (n = 29) as "Numbers", 1.9% (n = 12) as "Aspirations", 3.0% (n = 19) as "Others" and 0.5% (n = 3) as "Seconds". For the desired characteristics, the proportion of unique questions deemed memorable and consistent had declined, with 28.9% (n = 181) and 36.2% (n = 227), respectively; see Table 4.3. Whereas 61.9% (n = 388) were considered applicable, only 3.8% were confidential (n = 24). This indicates a significant reliance on easily discoverable information. Furthermore, only 15.9% of unique questions (n = 100) were deemed specific, thus highlighting a potential weakness in the design of SQs.

4.2 Security Question Requirement Analysis

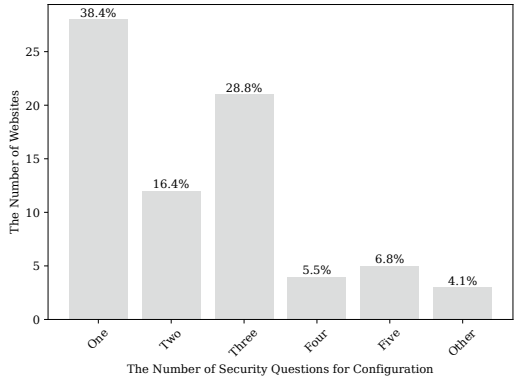
We reported our manual analysis of online services' SQ requirements with 73 websites.



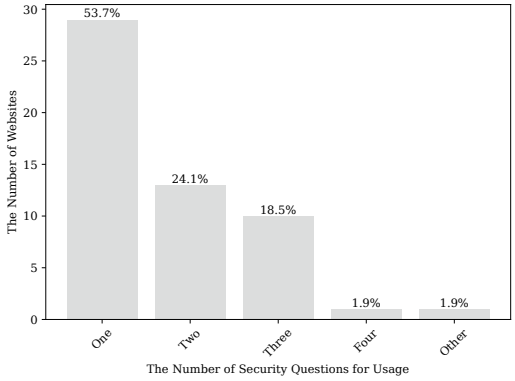
(a) Cumulative distribution function (CDF) of the numbers of SQs offered, selected, or used.



(b) The number of SQs offered.



(c) The number of SQs selected.



(d) The number of SQs used.

Figure 4.1: Distribution of the numbers of SQs.

4.2.1 Security Questions: Quantity and Customization

Most websites (n = 33, 45.2%) offered between five to ten questions to choose from, while only 4.1% websites (n = 3) provided fewer than five questions. Regarding the number of SQs required during registration, the most websites allowed users to set one question (n = 28, 38.4%), followed by three (n = 21, 28.8%) and two (n = 12, 16.4%). We found that 26.0% of the websites (n = 19) allowed setting SQs but did not use SQs in usage (i.e., login and password reset operations). Thus, we eventually analyzed 54 websites for testing the number of questions required for usage; the distribution is as follows: answering one question is required by 53.7% of the websites (n = 29), two questions by 24.1% (n = 13),

three questions by 18.5% (n = 10), and four questions by 1.9% (n = 1); see Figure 4.1d.

Three websites were different than others for the required number of questions during setup and usage. The website of Emerald Coast Utilities Authority[1] allowed users to set between three to ten questions, but during usage, the website asked users to answer all the questions set during registration (i.e., not a subset). The website of Australian Immigration and Citizenship[2] required users to set between three to five questions, but the number of questions presented during usage is always fixed to two. The website of Bell[3] allowed users to set between one to three questions[4].

Additionally, only 11.0% websites (n = 8) allowed customized questions. For the websites that offered setting multiple SQs, we found that 55.6% websites (n = 25) allowed users to configure the same answers to different questions. Similarly, we checked whether users could configure different answers to the same questions and found that no websites restricted this requirement.

4.2.2 Answers to Security Questions

Due to our analysis of 73 websites, we found *United Airlines*[5] predefined a list of answers to each question, unlike other websites. It offered 15 questions with 564 unique answers. The most categories of questions (11/15) were based on "Favorites" (e.g., "What is your favorite type of music?").

All websites accept answers of short length (under six characters); the only exception is United Airlines that does not allow user-typed answers. The majority of websites allows users to set answers with one character (e.g., a letter, digit, punctuation mark, or a symbol) (n = 36, 49.3%), three characters (n = 17, 23.3%), or four characters (n = 13, 17.8%). The

[1] `https://ecua-egov.aspgov.com/Click2GovCX/index.html`
[2] `https://online.immi.gov.au/lusc/login`
[3] `https://www.bell.ca/`
[4] No usage page for login and password recovery
[5] `https://www.united.com/en/us`

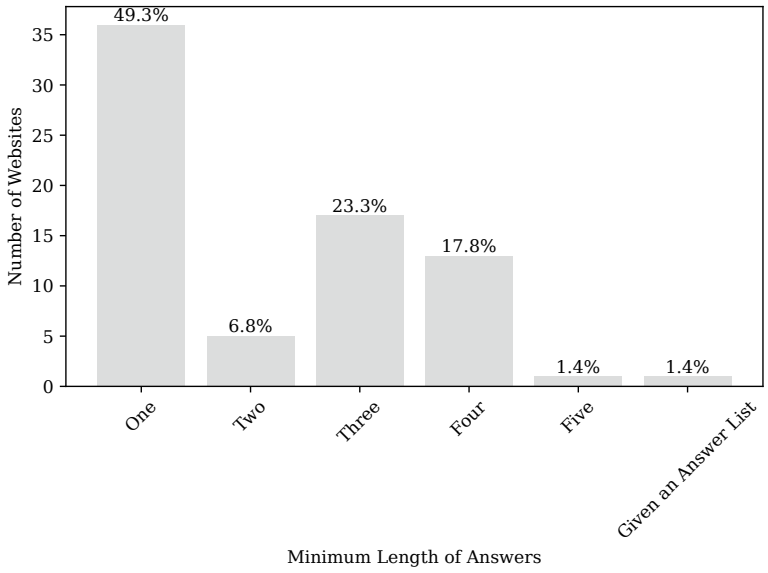


Figure 4.2: The minimum allowed number of characters.

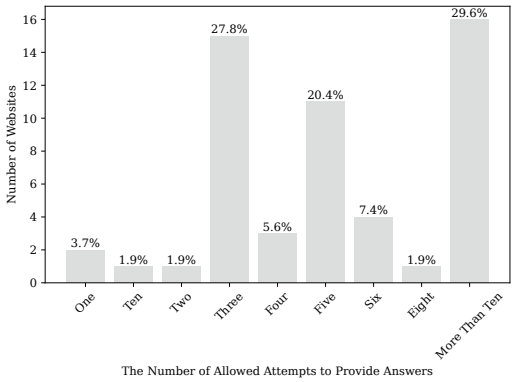
maximum allowed length for answers varied between 17 to 1000+ characters.

For the 32 punctuation marks that we tested, 13.7% websites ($n = 10$) did not allow users to use any of them; 57.5% ($n = 42$) accepted all of them; and the remaining websites accepted a subset of the punctuation marks.

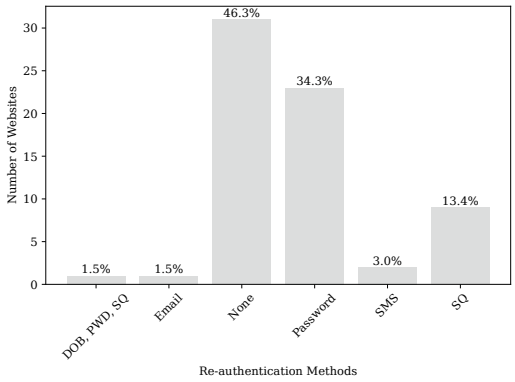
All websites except United Airlines and e.visa[6] permitted users to input digits. Only 8.9% websites ($n = 5$) allowed case-sensitive answers. 91.1% websites ($n = 51$) ignored case and match values regardless of their lower or upper case letters. 61.6% websites ($n = 45$) allowed users to input symbols when configuring the answers; however, among these, 10% websites ($n = 3$) did not accept symbols on the usage page (mainly due to encoding issues).

Since some websites did not offer SQs on the login or password reset page, we checked 74.0% websites ($n = 54$) for checking the number of allowed attempts to provide an answer (see Figure 4.3a). The most frequent number was three attempts ($n = 15$, 27.8%), followed by more than ten ($n = 16$, 29.6%), and five ($n = 11$, 20.4%).

[6] `https://visa.visitsaudi.com/`



(a) The number of allowed attempts to provide answers.



(b) The re-authentication methods when users change SQs.

Figure 4.3: Overview of security mechanism about SQs.

Similarly, according to the minimum cases in passwords, such as upper case letters, lower case letters and digits, we additionally tested the minimum cases of letters and characters (letters or digits). The number of all minimum cases on most websites (n = 61, 83.6%) was zero. A few websites did not allow users to type only punctuation marks or symbols, so the minimum case of characters was based on the minimum length of the answers.

Most websites (n = 69, 94.5%) allow users to input space during registration. After checking in the usage, 84.8% websites (n = 39) allow spaces[7] in the answers (Table 4.4), including 43.5% of the websites (n = 20) not allowing leading or trailing space and 4.3% (n = 2) not allowing leading space.

The majority of websites require users to answer SQs to verify their identity during password recovery (n = 47, 79.7%), followed by verification during the login process (n = 12, 20.3%). The rest websites apply SQs during customer service (n = 5, 8.5%) and when making changes to account settings (n = 2, 3.4%).

We found that, aside from 46.3% websites (n = 31) that did not implement re-authentication, passwords (n = 24, 35.8%) were the most common method (Figure 4.3b). A website also

[7]If websites disregard the space, treating answers with spaces as equivalent to those without spaces, we define they do not allow spaces.

Allowance of customized questions	Frequency
Not allowed	n = 65, 89.0%
Allowed	n = 8, 11.0%
Allowance of same answer (multiple questions needed)	
Not allowed	n = 20, 44.4%
Allowed	n = 25, 55.6%
Allowance of digits	
Not allowed	n = 2, 2.7%
Allowed	n = 71, 97.3%
Allowance of Case-sensitive	
Not allowed	n = 51, 91.1%
Allowed	n = 5, 8.9%
Allowance of symbols in the settings	
Not allowed	n = 28, 38.4%
Allowed	n = 45, 61.6%
Allowance of symbols in the usage	
Not allowed	n = 3, 10%
Allowed	n = 27, 90%
Allowance of space characters in the settings	
Not allowed	n = 4, 5.5%
Allowed	n = 69, 94.5%
Allowance of space characters in the usage	
Not allowed	n = 7, 15.2%
Allowed	n = 39, 84.8%

Table 4.4: Allowance of different features based on requirement analysis.

required simultaneously current SQs and date of birth. Other methods were used less, such as requiring the answer to the current SQs (n = 9, 13.4%) or sending a one-time password via SMS (n = 2, 3.0%). Notably, all websites (n = 11) designed a security flow in login authentication based on OWASP. In password recovery procedures, the majority of websites (n = 21, 58.3%) followed OWASP guidelines by combining security questions with email verification to authenticate users. Opposite the suggestion that the user's question on the usage page should remain the same until they answer it correctly, users can answer different questions if they input incorrect answers on 69.2% websites (n = 9).

As of November 13, 2024, four out of 212 websites have removed security questions from their authentication methods, while three out of four websites continue to retain security question in account settings.

4.3 Discussion

Notably, United Airlines provides a fixed list of answers for each predefined question and does not allow users to customize their answers. It is able to restrict user choice, increasing the likelihood of answer repetition among users. This limitation makes it easier for attackers to guess answers through brute force or social engineering techniques, raising the risk of SQs being compromised. Additionally, with fixed and limited answer choices, attackers may employ dictionary attacks, trying all possible options to bypass security. The use of preset answer lists can simplify the process for users, as they don't have to remember complex, customized answers, potentially enhancing usability and recall. For some users, particularly those who struggle to remember custom answers, this approach could make it easier to recall the answer when needed. However, not allowing users to customize answers limits their ability to choose information closely aligned with personal memories, which could affect applicability. For example, users may find that none of the provided options have clear memory associations, making the answer more difficult to remember and diminishing the user experience. Moreover, preset answer lists may fail to meet diverse user needs. If the options provided do not match a user's real-life experience, they may have to select irrelevant answers, impacting their experience and possibly leading to choices that prioritize memorability over security. To balance security with usability, websites could consider offering users a preset list along with partial customization options, which would allow for personalization and enhanced security, addressing the varied needs of users without compromising ease of use.

Our findings indicate that name-based SQs are the most prevalent type, constituting 71.8% of the options provided by websites. This popularity probably stems from the intuitive and familiar nature of questions involving names (e.g., "mother's name" or "pet's name"). However, these questions generally have low confidentiality, as names can often be obtained through public information or indirect means. In comparison, questions

about “firsts” (e.g., first job, first car) and “favorites” (e.g., favorite color, favorite movie) appear less frequently, at 31.6% and 22.9%, respectively. Although these types are more varied, they still pose security risks, as an attacker could potentially gather answers from publicly accessible data or social engineering techniques. Additionally, we observed that some websites offer a very limited selection of SQs, sometimes fewer than five. This restricted question pool limits user choice, increasing the likelihood of repetitive answers across accounts, which heightens the risk of answers being guessed or attacked. A small pool also limits the variety of question types available, reducing effectiveness in catering to users’ diverse backgrounds and recall preferences. To improve the usability and security of SQs, websites should expand their question pool by increasing both the quantity and variety of questions. Providing a wider range of unique questions allows users to select options that best align with their lives and memories, enhancing personalization and reducing the likelihood of common answers. An optimized question pool would better support users in safeguarding account security while maintaining a balance of usability and convenience.

Our findings on the possibility of guessing answers in the manual analysis show that the minimum length of answers is the main factor, additionally among considering the minimum number of (1) letters, (2) digits, (3) characters (letters or digits) (4) upper and lower case letters; (5) punctuation marks; and (6) the number of questions in the usage. Generally, SQs do not mandate minimum uppercase or lowercase characters, which reduces answer complexity and leaves them more weak to brute-force attacks, especially when compared to passwords that often have stricter complexity requirements. Additionally, increasing the complexity of answers to enhance security is impractical, as SQs typically rely on real information (names, places, dates), which do not match strict character-type requirements (cf. passwords). The main advantage of SQs lies in their simplicity and memorability. Imposing stricter complexity requirements on answers could undermine these qualities, making answers more difficult to recall and negatively affecting user experience. To strengthen SQs

without compromising usability, service providers could consider limiting the number of SQs required in the account verification process while also raising the minimum answer length. This approach would enhance the protective value of SQs without adding a significant memory burden for users, achieving a balance between security and ease of use. This finding emphasizes the trade-off between security and usability in the use of SQs. Although password managers can help with storing complex answers securely, their use also introduces reliance on external tools, which complicates the balance between security and usability.

Our analysis of website SQ requirements reveals that many websites fail to adhere to established guidelines, such as those from OWASP, thus increasing the risk of brute-force attacks. We also suggests that policymakers and data-protection authorities should play a more active role in enforcing best practices for SQs. Guidelines such as those from OWASP are widely available, yet they remain under-implemented by website administrators. Regulatory frameworks, such as the General Data Protection Regulation (GDPR) in the EU and the California Consumer Privacy Act (CCPA) in the US, already emphasize the need for strong security measures for protecting personal data. This could be extended to include stricter enforcement of SQ standards. Websites should be required to adhere to minimum standards for answer length and the avoidance of overly personal questions.

Responsible Disclosure. We conducted responsible disclosure for any identified security concerns with the websites involved, following best practices for ethical research. Finally, we contacted 44 websites to report SQs weaknesses but received four non-automated and 17 automatic responses, none of which addressed the reported weaknesses. For instance, American Airlines completely disregarded the reported weaknesses, responding with (*"I understand that you want to know more about our website"*) and providing a phone number and a link to their FAQ instead. Similarly, Australian Immigration and Citizenship referred us to the 2023-2030 Australian Cyber Security Strategy but avoided discussing

the weaknesses, citing concerns about disclosing sensitive information. Upwork merely redirected us to its bug bounty program.

Chapter 5

User Survey of Security Questions

In this chapter, we lay out the key steps and findings of the user study, beginning with its methodology. We describe the recruitment process, the structure and design of the questionnaire, the procedure followed, and the steps taken to ensure data reliability. Next, we present the study's results, highlighting respondents' insights and behaviors on SQs. Finally, we offer a brief discussion of these findings.

5.1 Methodology

To understand people's perceptions of SQs, we conducted an online user survey (N = 292); Figure 5.1 depicts the general study procedure with the key numbers regarding the

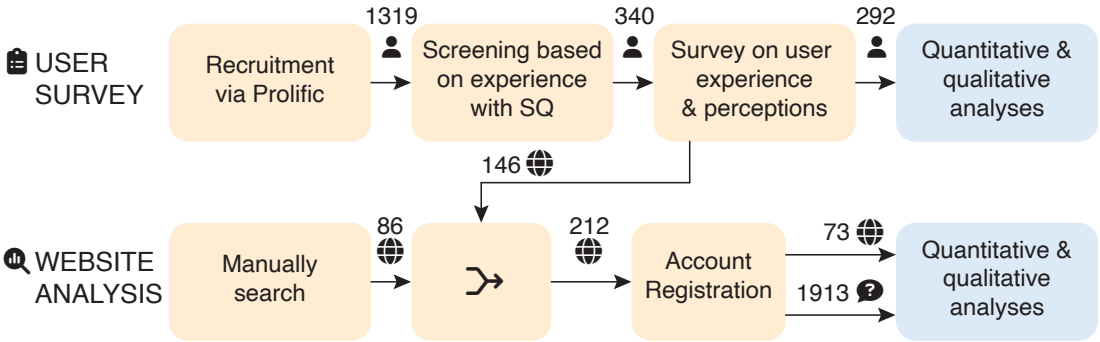


Figure 5.1: Flow chart of the study procedure.

collected data.

Ethics. The survey was approved by the Institutional Review Board (IRB) of the University of Lausanne. All aspects of this study were conducted with the utmost attention to respondent privacy and data protection. We informed the respondents about the nature of the survey and obtained their consent. The respondents were informed of their right to withdraw at any time, with a guarantee that their data would be deleted upon withdrawal. No personally identifiable information was collected. Respondents were also reminded to not upload screenshots of the SQs with their responses, thus further ensuring confidentiality.

5.1.1 Recruitment

We relied on the Prolific platform, considered a reliable source of respondents for scientific research [26]. We requested a representative sample of the US population (w.r.t. location, age, sex, and ethnicity, as per the US Census Bureau). We chose the US as Prolific has a large user base in the country and the English language used in the article is a commonly used language in the US. After two weeks of data collection, we obtained the targeted number of respondents in all categories except for the two respondents in the 55-64 age group, four in the 65-74, and 25 in the 75-100. Following Prolific's recommendation, we re-assigned the unfilled quotas to the other age groups. Table 5.1 provides the summary statistics about the respondents in our final sample (i.e., after our data-reliability procedure described next in Procedure Section). Our survey was advertised to the respondents as a 15-minute questionnaire about people's perceptions of and responses to SQs on online services. In the advertisement, we specified that (eligible) respondents should have at least one account on an online service that uses SQs. We also asked prospective respondents to choose one such online service in preparation for the survey. In accordance with Prolific's recommendations (i.e., £9.0 per hour), we compensated the respondents with £2.75 for their participation.

	n	%		*n*	%
Sex			**Age**		
Male	146	50%	18-24	36	12.3%
Female	146	50%	25-34	60	20.5%
Ethnicity			35-44	64	21.9%
Asian	16	5.5%	45-54	54	18.5%
Black	30	10.3%	55-64	45	15.4%
Mixed	29	9.9%	65-74	31	10.6%
Other	19	6.5%	75-100	2	0.7%
White	198	67.8%	**SeBIS [8]**	$M = 29.8, SD = 5.4$	

Table 5.1: Demographics of the survey respondents (final dataset, $N = 292$).

5.1.2 Questionnaire

Depending on the survey logic, respondents were asked 36 to 38 questions, distributed across eight parts. Almost all questions were close-ended, but four were open-ended in order to collect deeper responses about the more complex aspects (e.g., describe the respondent's rationale behind a certain behavior). The questionnaire covered various topics, including users' experiences, their strategies for answering SQs, and their perceptions of security mechanisms. We divide our questionnaire into eight parts (see the survey transcript in Appendix A.1).

Part A. Introduction and Consent. The questionnaire started with a consent form that detailed the topic and purpose of the study, the entities involved in conducting the study, the inclusion/exclusion criteria, as well as the respondents' rights in the research process. Next, respondents were required to provide explicit consent to partake in the survey.

Part B. Sample Website that use Security Question. We first asked the respondents if they use SQs and further asked them to provide the URL and a screenshot of a website that uses SQs. The respondents were instructed that the screenshot should provide evidence that the website indeed uses SQs but that it should *not* include their answers to the SQs (for obvious security reasons). The goal of this question was two-fold: (1) to make sure that the respondents know about SQs (and have some experience with them); and (2) to gather

names of websites that use SQs for subsequent analysis (see Figure 5.1). We warned the respondents that only accurate URLs and relevant screenshots will be accepted (otherwise, their data would be discarded and they would not be compensated).

Part C. User Experiences and Practices. We assessed the respondents' experiences and practices regarding the use of SQs in online services. We covered topics such as the prevalence of SQs, the purpose of their use, the number of SQs offered by the service providers, the number of SQs required by the service providers to be answered for registration and login, and respondents' strategies for selecting the SQs offered by the service providers. Lastly, we asked the respondents the proportion of SQs that rely on their usage of a given service (e.g., the phone number they call most often, in the case of a mobile operator).

Part D. User Strategies for Answering. We investigated respondents' strategies on how to set out answers to the SQs and recall them. We first asked about the frequency of providing consistent answers to the same SQ, and later assessed the respondents' attitudes towards providing truthful vs. untruthful answersto SQs. Respondents were asked about the reasons for doing so. We also assessed their comfort level about sharing their personal information with the service provider when answering SQs.

Part E. Answer Recall. We assessed respondents' perceptions and practices regarding the recalling of their answers to the SQs. We asked how difficult or easy it is for them to remember their answers, the mechanisms they use to facilitate the recalling process, and the extent which these mechanisms are utilized across online services relying on SQs. We also asked about the frequency of forgetting SQ responses.

Part F. Perceived Effectiveness. We assessed respondents' perceptions of the effectiveness of the protection provided by SQs. First, we asked the respondents to compare and explain the level of protection provided by SQs compared to passwords. The insights indicated whether there was a prevailing opinion among respondents regarding the relative effectiveness of passwords vs. SQs in providing protection. Next, we asked about the perceived

effectiveness of the protection provided by the SQs. Furthermore, we asked respondents to select the difficulty level of their answers being guessed by different adversaries such as strangers, close relatives, and mobile operators.

Part G. Preference of Security Mechanisms. We asked respondents to take into account both security and usability, and rank various security mechanisms (e.g., SMS, SQs, email [1, 36, 11]) as a second factor of authentication, from the most preferred to the least preferred. Additionally, we asked the respondents to rank different categories of SQs that they would prefer to be queried about. We followed the study of Golbeck and Li [12] to determine the following categories: Aspirations, Favourites, Firsts, Locations, Names, and Numbers. To mitigate potential bias stemming from the order in which items are presented, we randomized these categories.

Part H. General Security Behavior. We used the 8-item Security Behavior Intentions Scale (SeBIS) [8] to evaluate individuals' intentions and behaviors concerning online security practices. Through these questions, we aimed to gain insights into several key aspects: (i) awareness of security measures; (ii) behaviors in response to security incidents; (iii) strategies for managing security; and (iv) consistency in security practices.

5.1.3 Procedure

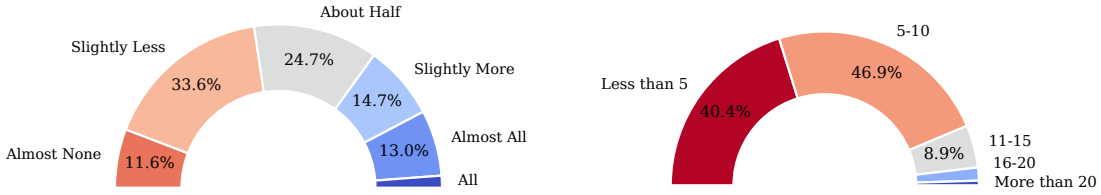
Prior to survey deployment, we conducted cognitive pretests and a soft launch to proactively identify and address potential issues in the questionnaire design. For the survey, we asked (via e-mail) seven master's students, from Concordia University, with backgrounds in computer security and (in-person) two doctoral students, from the University of Lausanne, with backgrounds in HCI. None of these students were part of the research team of this work. Using their feedback, we refined the question wording to enhance clarity and optimized the answer formats to improve efficiency and comprehension. For questions that respondents found confusing or difficult to understand, we rephrased them to make them

more understandable and highlighted certain key elements in bold to draw respondents' attention. Taking into account their suggestions, we modified some options by changing the adverbs. Subsequently, before deployment, we published a soft launch with a sample size of 20 respondents to assess the survey's comprehensiveness and to identify any necessary modifications for enhancement. During the soft launch, we discovered that one question was never displayed due to an error in the condition; we later changed this for the final survey. We did not use the soft launch data for the final analysis.

5.1.4 Data Reliability

Although Prolific serves as a valuable resource for recruiting respondents in order to facilitate high-quality data for academic research, it might not entirely prevent distracted, unfocused, or impatient respondents. Consequently, we implemented several strategies to mitigate these issues and ensure data quality. More specifically, we removed the respondents who (1) did not fully complete the questionnaire (907 removed, 412 kept); (2) disagreed with the consent (3 removed, 409 kept); (3) selected "no online service relied on SQs" or "none" for the proportion of online services where their accounts rely on SQs (69 removed, 340 kept); (4) selected the wrong answer in the attention check question (6 removed, 334 kept). We also (5) attempted to remove the responses of *speeders* who were *exceptionally fast*. However, following the guidelines of Matsuura et al. [23], we decided to keep them as their response looked reliable: they submitted a relevant screenshot and provided meaningful answers to open-ended questions, with consistent answers overall (2 suspected speeders identified, but 0 removed, 334 kept); (6) discarded the individuals whose screenshots did not display *original* SQs[1] (42 removed, 292 kept); (7) contacted the 17 respondents who misunderstood specific questions (e.g., frequency of truthful answers) and provided inconsistent answers in order to give them the opportunity to correct their

[1]We used the Google Image search engine to identify the images copied from the Web.



(a) Proportion of websites that rely on SQs.

(b) Number of SQs offered by websites upon registration.

Figure 5.2: Practices of websites regarding SQs (reported by the survey respondents).

responses: 10 of them did so. For instance, a respondent indicated "frequently" in response to the question about the frequency of providing truthful answers to SQs, but subsequently stated in an open-ended question that they never provided untruthful answers. We still retained the data of the seven other respondents as they provided meaningful responses to the open-ended questions, thus demonstrating their commitment. Ultimately, we collected $N = 292$ responses. The median time to complete the survey was 17:28.

5.2 Results

In this section, we present the findings derived from our survey[2].

5.2.1 User Practices and Experiences

Regarding the prevalence of SQs, the majority of the respondents reported that less than half of the online services, where they have accounts, rely on SQs (see Figure 5.2a). Note, however, that people who do not have any accounts on a service that relies on SQs were excluded from our pool. As for the evolution of the prevalence of SQs, overall, respondents had the impression that the proportion of online services that rely on SQs decreased over the last years (see Figure 5.3).

In terms of use cases, the majority of the respondents ($n = 267$, 91.4%) reported that

[2]In this section, we focus solely on the contributions we made in the survey.

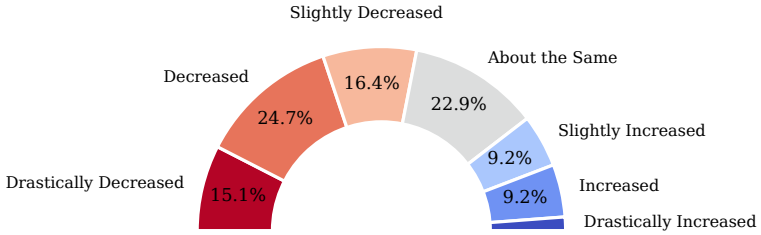
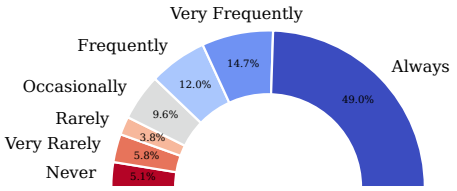


Figure 5.3: Evolution of the proportion of websites that rely on SQs (perceived and reported by the survey respondents).

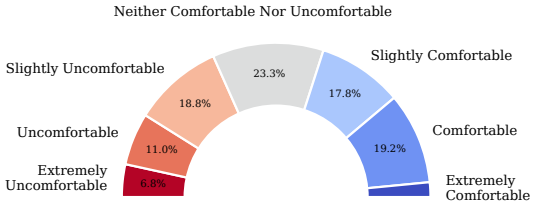
SQs are used by online services for letting users regain access to their account (e.g., password recovery). The other two use-cases were for accessing their account, e.g., login (n = 144, 49.3%), and for performing sensitive operations, e.g., money transfer (n = 104, 35.6%).

The vast majority of the respondents reported that the services, where they have accounts, provided less than ten SQs to choose from when registering their accounts; see Figure 5.2b. Regarding the number of SQs they were required to set up upon registration, the most frequent response was one of three questions (n = 157, 53.8%), followed by two (n = 85, 29.1%) and one (n = 33, 11.3%). As for the number of SQs they were required to answer on the usage page, e.g., login and password recovery, the vast majority reported only one (n = 156, 53.4%) or two (n = 82, 28.1%).

Regarding the type of SQs used by online services, the respondents were asked about the prevalence of questions based on their *use* of that service, e.g., "What is the approximate amount of the last transaction on your bank account?". Apparently, such questions are not widely used in practice with the large majority of respondents reporting that none (n = 116, 39.7%) or almost none (n = 87, 29.8%) of the services rely on such questions.



(a) Frequency of providing truthful answers to SQs.



(b) Comfort wrt giving away personal information in answers to SQs.

Figure 5.4: Respondents' strategies for answering SQs.

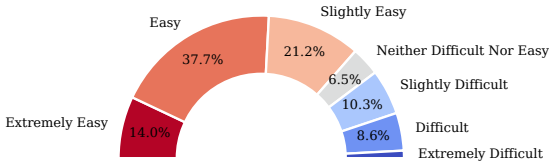
Mechanism	Frequency
Password manager	n = 107, 36.6%
Regular written notes (e.g., a physical notebook)	n = 67, 22.9%
Regular digital notes (e.g., simple file)	n = 53, 18.2%
Secure digital notes (e.g., Standard Notes, Joplin)	n = 34, 11.6%
Secure written notes (e.g., a physical note in a vault)	n = 26, 8.9%
None	n = 67, 22.9%

Table 5.2: Mechanisms used by the respondents for recalling their answers to SQs.

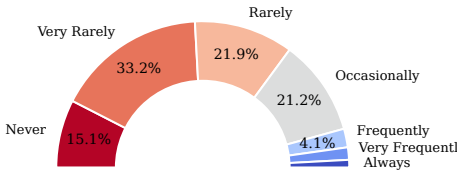
5.2.2 User Strategies for Answering

When asked whether they provide different responses to the same SQs across different services, only 16.1% (n = 47) of the respondents reported at least *"frequently"* providing different answers, whereas the majority mentioned that they never do so (n = 113, 38.7%), or do so very rarely (n = 61, 20.9%). In terms of answering truthfully vs. untruthfully when respondents registered their accounts, as shown in Figure 5.4a, the majority of the respondents usually provide truthful responses (*"always"*: n = 143, 49.0%; *"very frequently"*: n = 43, 14.7%; *"frequently"*: n = 35, 12.0%).

We asked the respondents how comfortable they would be sharing personal information when answering SQs. Overall, the largest group of respondents (n = 117, 40.1%) expressed some level of comfort, yet a substantial proportion (n = 107, 36.6%) declared being uncomfortable, as shown in Figure 5.4b. Neutral responses accounted for a substantial portion of the group (n = 68, 23.3%), indicating a mixed overall sentiment.



(a) Difficulty of remembering answers to SQs.



(b) Frequency of forgetting answers to SQs.

Figure 5.5: Memorability of SQs (reported by the survey respondents).

5.2.3 Answer Recall

About 20.5% (n = 60) of the respondents thought that it was difficult to remember their responses to SQs, with the majority finding it *"easy"* (n = 110, 37.7%), or *"extremely easy"* (n = 41, 14.0%). We then examined the frequency with which respondents forget their responses to SQs. Note that people might remember the answer to their SQs but not their exact response because of the variety of ways to spell it (e.g., upper/lower cases, spaces, hyphens, spelling of non-Latin characters in Latin). Note that we did not distinguish between the case where people forget the answer (i.e., the true response) and the case where they forget *their* response. The results indicated that only a few respondents forget the responses to SQs (*"always"*: n = 5, 1.7%; *"very frequently"*: n = 8, 2.7%; *"frequently"*: n = 12, 4.1%).

We asked the respondents whether they use any mechanism to help them recall their responses, and how frequently they do so; Table 5.2 summarizes these findings. Password managers (n = 107, 36.6%) and regular written notes (n = 67, 22.9%) are the two most common mechanisms. To the best of our knowledge, though password managers often offer the option to store generic notes (that can accommodate answers to SQs), none of the popular ones appear to offer the possibility to automatically save and/or fill the answers to SQs directly on webpages (as done for passwords). This is probably because there is no web standard associated with SQs (unlike for passwords, for which HTML elements such as '`<input type=''password''>`' can be used). Notably, 22.9% (n = 67) of respondents

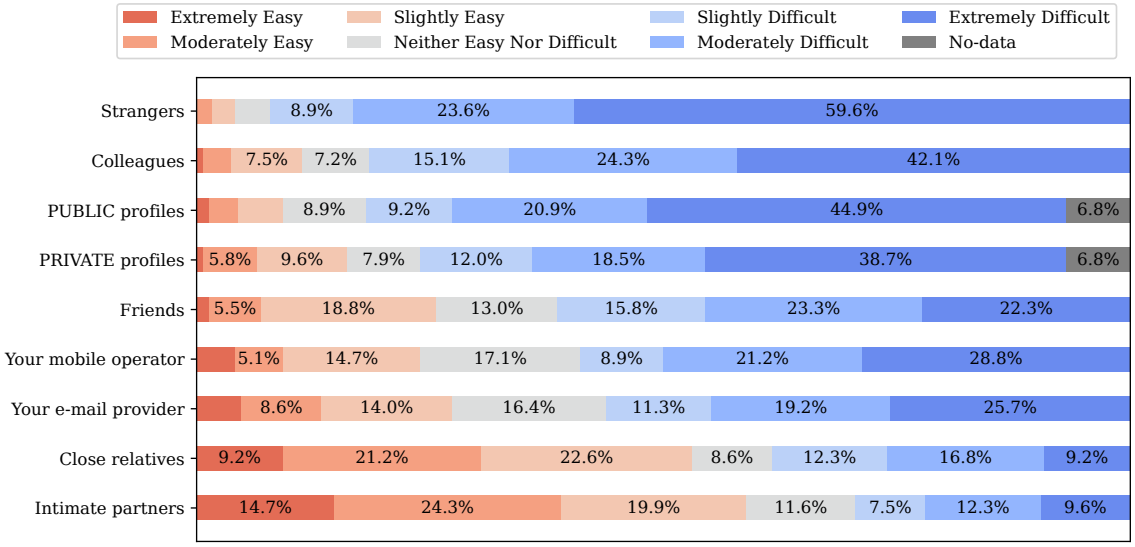


Figure 5.6: Levels of difficulty for different entities to guess someone's answers to SQs (assessed by the respondents).

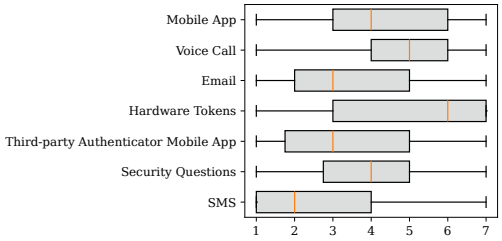
did not use any specific mechanism. Regarding the extent to which they rely on such mechanisms, we observed a tendency toward the lower end of the scale (*"none"*: $n = 21$, 9.3%; *"almost none"*: $n = 40$, 17.8%; *"slightly less"*: $n = 38$, 16.9%).

5.2.4 Perceived Effectiveness and Preferences

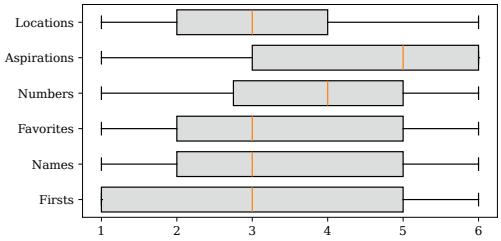
We assessed the respondents' perceptions of the protection level provided by SQs compared to passwords. More respondents ($n = 117$, 40.1%) indicated that SQs provided better protection than passwords, whereas 31.5% ($n = 92$) believed that passwords offered better protection (the rest perceived them as equivalent).

We asked respondents how effective they find SQs for protecting their online accounts. Most respondents ($n = 190$, 65.1%) believed that SQs were effective, with *"slightly effective"* being the most selected option ($n = 88$, 30.1%).

Regarding the difficulty, for different entities, to guess someone's answers, results shown in Figure 5.6), respondents reported that it would not be too difficult for their close relatives and intimate partners. Yet, interestingly, they reported similar levels of difficulty



(a) Ranking of security mechanisms (on top of a password); A rank of "1" denotes that a mechanism is *"the most preferred"* mechanism. A rank of "7" denotes the *"least preferred"* mechanism.



(b) Ranking of SQs types. A rank of "1" denotes that a question type is *"the most preferred"* type. A rank of "6" denotes *"least preferred"* type.

Figure 5.7: Respondents' preferences regarding different security mechanisms, considering both security and usability. The orange lines indicate the median values.

for friends and for e-mail/SMS providers, and higher levels for people with access to their social profiles.

Finally, we asked respondents to consider both security and usability and rank the security mechanisms when used as a second factor of authentication—on top of a password. The results are shown in Figure 5.7a (1 is the most preferred). SMS was the most preferred authentication method with a mean rank of 2.8 ($Mdn = 2$), while Hardware Tokens were the least preferred with a mean rank of 5.0 ($Mdn = 6$). SQs are ranked after authenticator mobile apps and emails with a mean rank of 3.9 ($Mdn = 4$). SMS was the most preferred authentication method, likely due to its excellent usability, particularly the seamless autofill feature that automatically pastes the verification code, making the process quick and convenient. However, despite its usability advantages, SMS has security issues that respondents may not be fully aware of, such as susceptibility to shoulder surfing attacks [9], SIM swap attacks [20], and the lack of encryption, which can allow adversaries to read the messages.

Similarly, we inquired about the kinds of SQs respondents preferred to be asked about (see Figure 5.7b). While Firsts were the most preferred type with having the lowest mean rank at 3.1, overall, Firsts, Names, Favorites, and Locations were all preferred with having the same median rank ($Mdn = 3$), possibly suggesting that these types of questions are

generally perceived as easier to remember and answer. Aspirations were the least preferred type with the highest mean rank at 4.3 ($Mdn = 5$), likely because they involve more subjective and potentially changing information, which may introduce uncertainty or difficulty in providing consistent answers over time.

5.3 Discussion

Our findings reveals that users can feel pressured to reveal truthful information, unaware that providing untruthful answers can be a valid and effective security strategy. This pressure also increases their exposure to *social engineering attacks*, e.g., via casual conversations or phishing attempts, particularly from friends and relatives, and even strangers. To mitigate these risks, users should be informed about the importance of varying their answers across services and be encouraged to use password managers for securely storing answers, especially when using more complex responses. Educating users about the flexibility of SQs and promoting these practices could help reduce their weaknesses.

Our findings also highlight *insider threats*, such as intimate-partner violence [10], as a significant concern. SQs often rely on personal information accessible to those closest to the user, making them vulnerable to exploitation by individuals with social proximity. Protecting users from these insider threats is especially important in cultural contexts where shared-device use is common. For example, cultural norms in regions such as South Asia often require women to share their devices with household members [28], which can create unexpected challenges. To address these vulnerabilities, SQs might need to be redesigned, potentially incorporating multi-factor authentication that cannot be easily bypassed by someone familiar with the user's personal details. Lastly, websites should consider the type of personal information users share publicly on their social media profiles and should propose SQs that rely on non-public behavioral information rather than static easily accessible details.

Chapter 6

Concluding Remarks

In this chapter, we delve into the key findings on SQs, outline the limitations, and offer targeted recommendations. Additionally, we provide an overview of future research directions to guide and inspire subsequent researchers.

6.1 Key Takeaways

Through detailed categorization of widely collected security questions and analysis of websites implementing them, this thesis reveals several key findings.

The Findings reveal that questions about "Names" make up the majority, accounting for 71.8% of all collected questions—significantly higher than other types, such as those related to "Firsts" (31.6%), "Favorites" (22.9%), and "Locations" (17.2%). When further analyzing unique questions, name-related questions remain dominant at 67.0%, while "Firsts", "Favorites", and "Locations" questions comprised 27.9%, 25.5%, and 16.4%, respectively.

Evaluation of these security questions based on the five key characteristics—memorability, consistency, applicability, confidentiality, and specificity—shows underwhelming results. Only 41.2% of questions fully meet the memorability, 44.0% satisfy consistency, and

62.2% are applicable. Confidentiality and specificity are met by just 2.0% and 28.2% of questions, respectively. For unique questions, these proportions decline further, highlighting poorer performance across these five attributes.

Analysis of current website practices also uncovers concerning trends. The majority of sites lacks minimum length requirements for security question answers, with 49.3% sites allowing answers as short as one character. Additionally, over half (55.6%) of the sites permit users to reuse the same answer for multiple security questions, and 91.1% do not enforce case sensitivity for answers. While these practices may enhance user convenience, they clearly compromise security.

Additionally, manual analysis of websites shows that many websites do not follow OWASP's recommendations when building their SQ authentication systems. For example, 31 out of 67 websites (46.3%) do not re-authenticate the user's identity before allowing them to set SQs, and 15 out of 36 websites (41.7%) rely solely on SQs during password recovery, without combining them with e-mail verification.

Our survey findings reveal important insights into users' perceptions and behaviors regarding SQs. Notably, users report a declining reliance on SQs for authentication, yet they remain frequently used for account recovery. Respondents commonly select questions based on ease of recall, yet a substantial proportion prioritizes privacy and security by choosing less predictable answers and answers that are not publicly available (e.g., from users' social media profiles). The survey highlights a significant gap in user awareness, with some believing that services already possess the correct answers to personal SQs. Although most users provide truthful answers, some choose untruthful responses for added security. The findings show that though users generally find SQs easy to recall, they express mixed feelings about the security and privacy implications of sharing personal information.

6.2 Limitations

Despite efforts to ensure broad coverage and objectivity in the evaluation process, this thesis has certain limitations.

Due to the recommendations by security authorities (such as those of the US NIST) against the use of SQs for online authentication, some websites have phased out this method, e.g., Google (since 2014). Additionally, we are not able to fully analyze the requirements of SQs on many websites because they require real-world identification, such as U.S. Social Security Numbers or client IDs. These pose a significant barrier to accessing websites for a thorough analysis of the requirements. Although we identified 146 websites offering SQs from the survey, only 35 out of 146 websites were ultimately suitable for detailed analysis. Thus, we extend the dataset of websites via the Google search engine, but it is not an efficient method. OWASP has outlined that SQs used for resetting forgotten passwords should possess the following characteristics: “memorable”, “consistent”, “applicable”, “confidential”, and “specific”. However, the lack of a standardized assessment method introduces bias, possibly leading to less precise or debatable assessments of certain security questions, thus affecting the accuracy and reliability of the conclusions. For instance, the memorability criterion implies that a security question’s answer should be easy for the user to recall. However, what qualifies as “easy to remember” largely depends on individual memory habits, adding complexity to consistent evaluation.

The limitations of the questionnaire study stem from several key factors that influence its generalizability and scope. Firstly, the respondent pool comprises 292 individuals, which, while providing meaningful insights, may not be sufficiently large to capture diverse perspectives comprehensively. Secondly, all respondents are drawn from a single country, potentially limiting the applicability of findings across varied cultural or geographical contexts. Furthermore, a majority of participants with university-level education and strong technical expertise, introducing a bias that excludes individuals with different educational

or technical backgrounds.

6.3 Recommendations

Here are some recommendations for online services to enhance the security and usability of security questions:

- Enhancing the diversity of security questions. Expanding the categories, numbers and specificity of security question options allows users to select questions that best suit their preferences, improving both usability and security.

- Optimizing security question design. It is essential to consider the five characteristics of memorability, consistency, applicability, confidentiality, and specificity. (1) Questions and answers should focus on memorable cues users can easily recall; (2) Questions and answers should be keep consistent to prevent users from forgetting or misinterpreting responses due to time or situational changes. (3) Questions should suit users of various ages, professions, and cultural backgrounds to ensure broad applicability. (4) Questions and answers should avoid including details that could easily be discovered through social media or other publicly accessible sources. (5) Questions and answers should be specific and clear to minimize user confusion or misinterpretation, avoiding ambiguous or overly broad language.

- Implementing strict security standards. By following guidelines from trusted organizations (e.g., OWASP), websites can establish stringent requirements for security question authentication mechanism, including length, complexity, and guess attempts, to strengthen account protection.

For users, here are some suggestions:

- Choose unique and difficult-to-guess answers. Users should avoid simple or easily guessed answers and select the responses that are known only to users and difficult for others to obtain.

- Handle personal information carefully. Users should protect their personal information by avoiding sharing details that may be linked to their security answers in public settings.

- Use memory tools for complex answers. For answers that are difficult to remember, users should consider using a password manager to securely store them, ensuring they can accurately recall them when needed.

6.4 Future Work

According to the findings and limitations in this thesis, there are some potential future works:

- Security questions in mobile apps. It would investigate various types of mobile apps (also increase the number) to analyze the security questions they employ, the design patterns used, and how these elements relate to user privacy and data security. It would evaluate the real-world effectiveness of security questions in these apps, considering factors such as their memorability and resilience against guessing attacks. Then it would compare with such mechanism in the websites to find the variations.

- Comprehensive review and optimization of the security question verification process. This includes, but is not limited to, re-evaluating the effectiveness of existing security questions, improving the mechanisms for validating user responses, and ensuring the process is as resilient as possible against automated attacks and human guessing.

For example, implementing more sophisticated logic or integrating historical user behavior data could significantly enhance the security of the verification process.

- Protections for user answers in online services. This may involve employing advanced encryption techniques to securely store answers, enforcing stricter data access control policies, and developing systems capable of real-time monitoring and blocking suspicious activities. Furthermore, it could focus on designing more personalized and less predictable security questions to increase the difficulty for attackers attempting to breach accounts.

- Security questions generation. It would research a novel security question answer management mechanism that allows users to dynamically update their security question answers periodically or based on specific events, such as forgotten passwords or device changes, enhancing account security. Additionally, leveraging machine learning and data analytics to customize security questions based on user behavior patterns could make these questions more difficult to guess while simultaneously improving the user experience.

Bibliography

[1] A. Al Abdulwahid, N. Clarke, S. Furnell, I. Stengel, and C. Reich. The Current Use of Authentication Technologies: An Investigative Review. In *2015 International Conference on Cloud Computing*, ICCC'15, pages 1–8, Riyadh, Saudi Arabia, 2015. IEEE.

[2] R. AlHusain and A. Alkhalifah. Evaluating knowledge-based security questions for fallback authentication. *PeerJ Computer Science*, 8:e903, 2022.

[3] J. Bonneau, E. Bursztein, I. Caron, R. Jackson, and M. Williamson. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 141–150, Florence, Italy, 2015. International World Wide Web Conferences Steering Committee.

[4] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *2012 IEEE Symposium on Security and Privacy*, S&P'12, pages 553–567, San Francisco, CA, USA, 2012. IEEE. ISSN: 2375-1207.

[5] J. Chia, S.-H. Heng, J.-J. Chin, S.-Y. Tan, and W.-C. Yau. An implementation suite for a hybrid public key infrastructure. 13(8):1535.

[6] R. Dhekane. *Towards a usable fallback authentication mechanism*. PhD thesis, California State University, Sacramento, CA, USA, 2020.

[7] V. Distler, M. Fassl, H. Habib, K. Krombholz, G. Lenzini, C. Lallemand, L. F. Cranor, and V. Koenig. A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research. *ACM Trans. Comput.-Hum. Interact.*, 28(6):43:1–43:50, 2021.

[8] S. Egelman and E. Peer. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI'15, pages 2873–2882, Seoul, Republic of Korea, 2015. ACM.

[9] M. Eiband, M. Khamis, E. von Zezschwitz, H. Hussmann, and F. Alt. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 4254–4265, Denver, CO, USA, 2017. Association for Computing Machinery.

[10] D. Freed, J. Palmer, D. Minchala, K. Levy, T. Ristenpart, and N. Dell. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–13, Montréal, Canada, 2018. Association for Computing Machinery.

[11] C. Gilsenan, F. Shakir, N. Alomar, and S. Egelman. Security and Privacy Failures in Popular 2FA Apps. In *32nd USENIX Security Symposium*, USENIX Security'23, pages 2079–2096, Anaheim, CA, USA, 2023. USENIX Association.

[12] J. Golbeck and S. Li. A Large Crowdsourced Dataset of Security Questions. In *Proc. of Who Are You?! Adventures in Authentication Workshop*, WAY'20, page 4, Virtual Event, 2020.

[13] M. Golla and M. Dürmuth. Analyzing 4 Million Real-World Personal Knowledge Questions (Short Paper). In *Technology and Practice of Passwords*, volume 9551 of *PASSWORDS'15*, pages 39–44, Cambridge, UK, 2016. Springer International Publishing.

[14] W. J. Haga and M. Zviran. Question-and-answer passwords: An empirical evaluation. *Information Systems*, 16(3):335–343, 1991.

[15] S. Höltervennhoff, N. Wöhler, A. Möhle, M. Oltrogge, Y. Acar, O. Wiese, and S. Fahl. A Mixed-Methods Study on User Experiences and Challenges of Recovery Codes for an {End-to-End} Encrypted Service. In *33rd USENIX Security Symposium*, USENIX Security'24, pages 7267–7284, Philadelphia, PA, USA, 2024. USENIX Association.

[16] M. Just and D. Aspinall. Challenging challenge questions: An experimental analysis of authentication technologies and user behaviour. 2(1):99–115.

[17] M. Just and D. Aspinall. Personal choice and challenge questions: a security and usability assessment. In *Proc. of the Symp. on Usable Privacy and Security*, SOUPS'09, pages 1–11, Santa Clara, CA, USA, 2009. Association for Computing Machinery.

[18] G. Kontaxis, E. Athanasopoulos, G. Portokalidis, and A. D. Keromytis. SAuth: protecting user accounts from password database leaks. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS '13*, pages 187–198. ACM Press.

[19] L. Lassak, P. Markert, M. Golla, E. Stobert, and M. Dürmuth. A Comparative Long-Term Study of Fallback Authentication Schemes. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–19, Honolulu, HI, USA, 2024. Association for Computing Machinery.

[20] K. Lee, B. Kaiser, J. Mayer, and A. Narayanan. An Empirical Study of Wireless Carrier Authentication for SIM Swaps. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, SOUPS'20, pages 61–79, Virtual Event, 2020. USENIX Association.

[21] W. Ma, J. Campbell, D. Tran, and D. Kleeman. Password Entropy and Password Quality. In *Proceedings of the 2010 Fourth International Conference on Network and System Security*, NSS'2010, pages 583–587, NW Washington, DC, USA, 2010. IEEE.

[22] M. Mannan and P. C. van Oorschot. Security and usability: the gap in real-world online banking. In *Proceedings of the 2007 Workshop on New Security Paradigms*, NSPW '07, pages 1–14, North Conway, NH, USA, 2008. Association for Computing Machinery.

[23] T. Matsuura, A. A. Hasegawa, M. Akiyama, and T. Mori. Careless Participants Are Essential for Our Phishing Study: Understanding the Impact of Screening Methods. In *Proceedings of the 2021 European Symposium on Usable Security*, EuroUSEC '21, pages 36–47, Karlsruhe Germany, 2021. Association for Computing Machinery.

[24] N. Micallef and N. A. G. Arachchilage. Understanding users' perceptions to improve fallback authentication. *Personal and Ubiquitous Computing*, 25(5):893–910, 2021.

[25] OWASP. Choosing and Using Security Questions, 2018.

[26] S. Palan and C. Schitter. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[27] A. Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of Facebook. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS'08, pages 13–23, Pittsburgh, PA, USA, 2008. ACM.

[28] N. Sambasivan, G. Checkley, A. Batool, N. Ahmed, D. Nemer, L. S. Gaytán-Lugo, T. Matthews, S. Consolvo, and E. Churchill. "Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, SOUPS'18, pages 127–142, Baltimore, MD, USA, 2018. USENIX Association.

[29] S. Schechter, A. B. Brush, and S. Egelman. It's No Secret. Measuring the Security and Reliability of Authentication via "Secret" Questions. In *Proc. of the IEEE Symp. on Security and Privacy*, S&P'09, pages 375–390, Oakland, CA, USA, May 2009. IEEE.

[30] E. M. Scheidt and E. Domangue. Multiple factor-based user identification and authentication, Jan. 18 2005. US Patent 6,845,453.

[31] A. Senarath, N. A. G. Arachchilage, and B. B. Gupta. Security Strength Indicator in Fallback Authentication: Nudging Users for Better Answers in Secret Questions, 2017. arXiv:1701.03229 [cs].

[32] S. W. Shah and S. S. Kanhere. Recent Trends in User Authentication – A Survey. *IEEE Access*, 7:112505–112519, 2019.

[33] Similarweb. All Categories of Similarweb Website Ranking, 2024.

[34] J. Tan, L. Bauer, N. Christin, and L. F. Cranor. Practical Recommendations for Stronger, More Usable Passwords Combining Minimum-strength, Minimum-length, and Blocklist Requirements. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, pages 1407–1426, Virtual Event, 2020. ACM.

[35] A. Ullah, H. Xiao, T. Barker, and M. Lilley. Evaluating security and usability of profile based challenge questions authentication in online examinations. 5(1):2.

[36] X. Wang, Z. Yan, R. Zhang, and P. Zhang. Attacks and defenses in user authentication systems: A survey. *Journal of Network and Computer Applications*, 188:103080, 2021.

Appendix A

A.1 Survey Transcript

Note: Coding rules are colored in gray (not visible to respondents)

Sec. A. Introduction and Consent

Q1. [Text Entry.] What is your Prolific ID? Please note that this response should **auto-fill** with the correct ID. [ ] (text)

STUDY

You are invited to participate in a research study about people's perceptions of and responses to security questions on online services (e.g., "What is your mother's maiden name?", which could be used to secure access to users' accounts). This study is conducted by researchers from [Redacted for anonymous review]. We ask you to take a survey that takes approximately **15 minutes to complete**.

PARTICIPATION CRITERIA

To be eligible for this study, you must be a user of at least one online service that relies on security questions.

YOUR RIGHTS

You will be paid **£2.75 (USD 3.5)** for your complete participation in the study. You may choose to terminate your participation in this study at any time and for any reason. However, if you choose to terminate your participation, you will not be compensated, and your data will be deleted. If you participate, your answers will be kept **confidential**. Also, we do not collect any personally identifying information, such as your name and e-mail address. All data will be **stored** on a **secured** server, and only the researchers participating in this study can access it.

The results of this research study will be used solely for academic research and might be published in scientific journals and conferences. Any published information will be **aggregated** and/or **de-identified**.

As a precautionary measure, in your responses to this survey, please refrain from providing any responses that you usually give to the security questions you receive online.

Q2. [Single Selection.] **CONSENT**

If you wish to participate in this research study and you meet the participation criteria, please select the "**Agree**" option to continue. It will indicate that you fulfill the participation criteria, that you will answer all questions truthfully, and that you consent to our use of the collected data under the conditions stated above. If you select the "**Disagree**" option, you will not participate in this research study and will not be paid.

(a) Agree

(b) Disagree [Terminate.]

Sec. B. Sample Website that use Security Question

Q3. [Single Selection.] Does any of the online services, where you have an account, rely on **security questions** (e.g., "What is your mother's maiden name?")?

(a) Yes

(b) No [Terminate.]

Q4. [Single Selection] Approximately what proportion of the online services where you have an account rely on security questions?

(a) 1 - None [Terminate.]

(b) 2 - Almost None

(c) 3 - Slightly Less

(d) 4 - About Half

(e) 5 - Slightly More

(f) 6 - Almost All

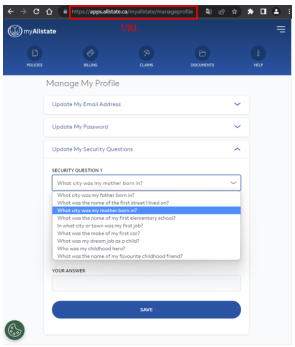
(g) 7 - All

Please note: For the following questions, only accurate URLs and relevant screenshots of security questions will be accepted. Respondents who submit incorrect or irrelevant URLs/screenshots **will not be compensated**.

Q5. [Text Entry.] Please provide the URL (link) of the website of one such online service where you have an account. [ ] (text)

Q6. [File Upload.] Please provide a screenshot showing the security questions (**but not your responses**) for one such online service where you have an account.

Only PDF and graphic files (e.g., PNG, JPG) are accepted. Here is an example (from Allstate).



For instructions on how to take a screenshot on your device, click on the appropriate link: Android, iOS, Windows, macOS, Linux.

[Drop files or click here to upload] (file upload)

Sec. C. User Practices and Experiences

Q7. [Single Selection.] Based on your experience, the proportion of online services that rely on security questions has recently

(a) 1 - Drastically Decreased

(b) 2 - Decreased

(c) 3 - Slightly Decreased

(d) 4 - About the Same

(e) 5 - Slightly Increased

(f) 6 - Increased

(g) 7 - Drastically Increased

Q8. [Multiple Selection.] For what purpose do these online services, where you have an account, rely on security questions?

(Select all that apply)

(a) To access my account (e.g., login)

(b) To regain access my account (e.g., password recovery)

(c) To perform sensitive operations (e.g., transferring large amounts of money, modifying important account settings)

(d) Other (please specify) [ ] (text)

Q9. [Single Selection.] Approximately how many different security questions do the online services you use typically **offer you to select from, when registering your account**?

(a) Less than 5

(b) 5-10

(c) 11-15

(d) 16-20

(e) More than 20

Q10. [Single Selection.] Approximately how many security questions do your online services typically **require you to select when registering your account**?

(a) 1

(b) 2

(c) 3

(d) 4

(e) More than 4

Q11. [Open-ended.] What is your **general approach for selecting** security questions when **registering** your account? [ ] (text)

Q12. [Single Selection.] Approximately how many security questions are typically **required when you use your online services** (e.g., to login)?

(a) 1

(b) 2 [Display this option if Q10$=2 or Q10$=3 or Q10$=4 or v$=More than 4]

(c) 3 [Display this option if Q10$=3 or Q10$=4 or Q10$=More than 4]

(d) 4 [Display this option if Q10$=4 or Q10$=More than 4]

(e) More than 4 [Display this option if Q10$=More than 4]

Q13. [Single Selection.] **Among your services that rely on security questions**, approximately what proportion of them rely on **questions based on your use of that service**?

Examples: from your bank: ”approximate amount of the last transaction on your bank account”, from your mobile operator: ”phone number you call most often”, etc.

(a) 1 - None

(b) 2 - Almost None

(c) 3 - Slightly Less

(d) 4 - About Half

(e) 5 - Slightly More

(f) 6 - Almost All

(g) 7 - All

Sec. D. User Strategies for Answering

Q14. [Single Selection.] How often do you use **different responses** to the **same security questions** across multiple online services (e.g., declaring your favorite food as ”Pizza” on one service and as ”Pasta” on another)?

(a) 1 - Never

(b) 2 - Very Rarely

(c) 3 - Rarely

(d) 4 - Occasionally

(e) 5 - Frequently

(f) 6 - Very Frequently

(g) 7 - Always

Q15. [Single Selection.] In general, how frequently do you provide the **truthful answers** (e.g., the **actual** maiden name of your mother) to security questions **when registering your account**?

(a) 1 - Never

(b) 2 - Very Rarely

(c) 3 - Rarely

(d) 4 - Occasionally

(e) 5 - Frequently

(f) 6 - Very Frequently

(g) 7 - Always

[Display Q16 if Q15$!=1 - Never]

Q16. [Open-ended.] Please briefly describe your reasons for providing **truthful** answers to security questions, when registering your account. [ ] (text)

[Display Q17 if Q15$!=7 - Always]

Q17. [Open-ended.] Please briefly describe your reasons for providing **untruthful** answers to security questions, when registering your account. [ ] (text)

Q18. [Single Selection.] How comfortable are you with giving away personal information to an online service when providing responses for the security questions?

(a) 1 - Extremely Uncomfortable

(b) 2 - Uncomfortable

(c) 3 - Slightly Uncomfortable

(d) 4 - Neither Comfortable Nor Uncomfortable

(e) 5 - Slightly Comfortable

(f) 6 - Comfortable

(g) 7 - Extremely Comfortable

Sec. E. Answer Recall

Q19. [Single Selection.] How difficult/easy do you find it to remember your responses to security questions?

(a) 1 - Extremely Difficult

(b) 2 - Difficult

(c) 3 - Slightly Difficult

(d) 4 - Neither Difficult Nor Easy

(e) 5 - Slightly Easy

(f) 6 - Easy

(g) 7 - Extremely Easy

Q20. [Multiple Selection.] Which of the following mechanisms do you use to help you recall the responses (to the security questions) you provided when registering the account?

Select all that apply.

(a) Password manager

(b) Regular digital notes (e.g., simple file)

(c) Secure digital notes (e.g., Standard Notes, Joplin)

(d) Regular written notes (e.g., a physical notebook)

(e) Secure written notes (e.g., a physical notebook stored in a vault)

(f) Other (please specify) [ ] (text)

(g) None

[Display Q21 if Q20$!=None]

Q21. [Single Selection.] To what extent do you use these mechanisms (i.e., to help you recall the responses) on the online services that rely on security questions?

(a) 1 - None

(b) 2 - Almost None

(c) 3 - Slightly Less

(d) 4 - About Half

(e) 5 - Slightly More

(f) 6 - Almost All

(g) 7 - All

Q22. [Single Selection.] How frequently do you forget the response you provided to a security question (when registering the account)?

(a) 1 - Never

(b) 2 - Very Rarely

(c) 3 - Rarely

(d) 4 - Occasionally

(e) 5 - Frequently

(f) 6 - Very Frequently

(g) 7 - Always

Sec. F. Perceived Effectiveness

Q23. [Single Selection.] In your opinion, **compared to** the level of protection provided by **passwords**, the level of protection provided by **security questions** is:

(a) Much Worse

(b) Worse

(c) Slightly Worse

(d) About the Same

(e) Slightly Better

(f) Better

(g) Much Better

Q24. [Open-ended.] Please briefly describe why you think the level of protection provided by security questions is [selected choice at Q23] than that provided by passwords. [ ] (text)

Q25. [Single Selection.] How effective do you find security questions for protecting your online accounts?

(a) 1 - Strongly Ineffective

(b) 2 - Ineffective

(c) 3 - Slightly Ineffective

(d) 4 - Neither Effective Nor Ineffective

(e) 5 - Slightly Effective

(f) 6 - Effective

(g) 7 - Strongly Effective

Q26. [Single Selection.] Please select "3 - Slightly Disagree" to show you are paying attention to this question. [Attention check]

(a) 1 - Strongly Disagree

(b) 2 - Disagree

(c) 3 - Slightly Disagree

(d) 4 - Neither Agree Nor Disagree

(e) 5 - Slightly Agree

(f) 6 - Agree

(g) 7 - Strongly Agree

Q27. [Single Selection.] Do you have any online social network accounts (e.g., Facebook, Twitter)?

(a) Yes

(b) No

When answering the following question, please note that a **public** social media profile means that anyone can see your posts/stories, and a **private** social media profile means that only your friends/contacts/followers can see them.

Q28. [Grid question.] How difficult/easy do you think it is for the following entities to guess — in a few tries (1-3) — your responses to security questions?

For entities that do not apply, please imagine and let us know.

Row options:

- Strangers
- Close relatives (parents, siblings)
- Intimate partners
- Colleagues
- Friends
- Anyone with access to your **PUBLIC** social network profiles (information such as gender and age), posts, photos, etc. [Display this option if Q27$=Yes]
- Anyone with access to your **PRIVATE** social network profiles (information such as gender and age), posts, photos, etc. [Display this option if Q27$=Yes]
- Your e-mail provider (e.g., Gmail)
- Your mobile operator (e.g., AT&T)

(a) Extremely Difficult

(b) Moderately Difficult

(c) Slightly Difficult

(d) Neither Easy Nor Difficult

(e) Slightly Easy

(f) Moderately Easy

(g) Extremely Easy

Sec. G. Preference of Security Mechanisms

Q29. [Rank Order. Order randomized.] Taking into account **both security and usability**, please **rank** the following security mechanisms — when used as a **second factor of authentication (on top of a password) — from the most preferred to the least preferred**:

(a) SMS

(b) Security questions

(c) Third-party authenticator mobile app (e.g., Google Authenticator)

(d) Hardware Tokens (e.g., bank smartcard and HSBC Secure Key)

(e) E-mail

(f) Voice call

(g) Mobile app (i.e., the mobile app provided by the online service provider, such as HSBC Mobile Banking App)

Q30. [Rank Order. Order randomized.] Considering **both security and usability**, which kind of questions would you prefer to be asked about? Please **rank** them from **the most preferred** to **the least preferred** one.

(a) Firsts (e.g., What was the first movie you saw in the theater?)

(b) Names (e.g., What is your mother's maiden name?)

(c) Favorites (e.g., What is your favorite music?)

(d) Numbers (e.g., What was your childhood phone number, including area code?)

(e) Aspirations (e.g., What is your dream job?)

(f) Locations (e.g., In which city did you meet your spouse/significant other?)

Sec. H. General Security Behavior (SeBIS) [8]

Please indicate how often you use the following technology-related practices.

Q31. [Single Selection.] I set my computer screen to automatically lock if I don't use it for a prolonged period of time.

Q32. [Single Selection.] I use a password/passcode to unlock my laptop or tablet.

Q33. [Single Selection.] I manually lock my computer screen when I step away from it.

Q34. [Single Selection.] I use a PIN or passcode to unlock my mobile phone.

Q35. [Single Selection.] I do not change my passwords, unless I have to.

Q36. [Single Selection.] I use different passwords for different accounts that I have.

Q37. [Single Selection.] When I create a new online account, I try to use a password that goes beyond the site's minimum requirements.

Q38. [Single Selection.] I do not include special characters in my password if it's not required.

(a) 1 - Never

(b) 2 - Rarely

(c) 3 - Sometimes

(d) 4 - Often

(e) 5 - Always