

An analysis of frequently asked security questions and their use for account security.

Year 4 Final Project Research Document

by Matas Aleksiejus

C00239630

Contents

Abstract	2
Introduction.....	2
Security Questions, Shared secrets and their use for authentication.....	3
History of security questions and their applications.	4
The Emigrant Industrial Savings Banks.....	4
The Banking Law Journal	4
Modern applications and ideas for security questions.....	5
Generating security stories for better protection of user privacy.	5
Propp’s theory	5
Main issues surrounding security questions.....	7
Guessability	7
Memorability	8
Geographical factors	9
Technologies.....	10
Existing applications.....	12
Proposed solution	12
Development.....	13
Technology Stack.....	13
Android Studio.....	13
MIT App Inventor.....	13
Emulation	14
SQLite	14
Chrome Browser	14
Languages	14
Java	14
Kotlin	15
XML.....	15
SQL.....	15
Other Possible Languages	15
HTML and JavaScript	15
Conclusion.....	16
Bibliography.....	17

Abstract

Since the 19th century security questions have been utilised with the intent to secure personal accounts however their place in modern online security is highly contested. This paper analyses their historical, current and possible future uses. Exploring the history of security questions it was discovered that their early uses were in financial institutions. The research carried out identified two main issues with security questions, memorability and guessability. Future proposals for security questions outline the possibility of using shared secrets that are randomly generated for use in account recovery. Based on the research a solution is proposed to build an application the generate and store secure answers.

Introduction

Forgot your password? No problem right? You go to account recovery only to be greeted by "Please answer the question 'What is your favourite food?' " for an account you made two years ago when you were really into a niche type of cuisine. Or maybe you're just not sure whether you wrote pizza or burger. Or worse yet received an email asking to verify a sign in that most definitely wasn't you. Everyone has gone through some form of this and security questions and their use in today's online security are to blame. Security questions are used for account recovery by many websites even though previous research indicates that the same types of account security and even similar questions were used to secure accounts over a hundred years ago. So do security questions have a place in our online activity now or in the future?

Account security has never been more important than it is right now. Due to the global pandemic forcing everyone to stay home and work, shop and basically live online. Our accounts for social media, shopping platforms video conferencing sites and everything else have become a trove of sensitive personal information and losing access to an account is inconvenient at the best of times but having your account and all the information in it stolen or defaced is much worse. Accounts secured by security questions do nothing to safeguard against this happening and if anything they make it easier to gain access. This is why this research and proposed solution is so important.

This paper will cover what security questions are, how they work and what their purpose is. It will also cover the origins of security questions and shared secrets, how they came to be and their earliest uses.

Completed research will then focus on a possible modern solution for security questions in the form of generated stories. This will identify the biggest problems with security questions and why it is so important that they are addressed. After that the research will explore some of the technologies that still use security questions, some solutions that people have made from necessity and my proposed solution to this problem including possible technologies that may be use to implement the solution. Finally the paper is concluded with recommendations and closing statements.

Security Questions, Shared secrets and their use for authentication.

The idea of using a shared secret between client and company has been around from the 19th century and has been used for all sorts of authentication purposes. At first it was used in banks to identify that a person was who they said they were when making a withdrawal and now it is more commonly used to recover usernames and accounts when a user has forgotten either one. A shared secret is a piece of information that only the customer/user knows and they share this information in confidence with the company they're opening an account with so that the company may use this confidential information to authenticate the user. A security questions works very much in the same way. When creating an account on a website a user will be asked a question and these questions can range a lot, some of the most common ones are:

- In what city were you born?
- What is the name of your favourite pet?
- What is your mother's maiden name?
- What high school did you attend?
- What is the name of your first school?
- What was the make of your first car?
- What was your favourite food as a child?
- Where did you meet your spouse?

(Haber, 2020)

When the user answers this question they are passing on personal information or a "shared secret" to the company that their making their account for. In the case of a user forgetting their password or username they can then enter their email address and the answer to their security questions to let the company know that it is indeed them and they want to reset their password.

History of security questions and their applications.

The Emigrant Industrial Savings Banks

The most prominent use of security questions historically was by various financial institutions dating as far back as 1850. The Emigrant Industrial Savings Bank opened its doors to Irish immigrants in New York due to many immigrants facing discrimination at other establishments. At the time majority of banks used signatures to identify a customer but since many immigrants could not read or write the bank kept an extensive record of personal information in a document called a "*Test book*" that customers could answer to the clerk to authenticate themselves. (Ruberg, 2017) The Emigrant Savings Bank later donated all the personal information to the New York public library as extensive genealogical resources.

The Banking Law Journal

Writing about security questions was first published in 1906 by "The Banking Law Journal" where at a meeting of the American Banker Association in Baltimore a banker named William M. Hayden described his banks use of security questions. The security questions were asked with the rest of the information used to open an account back then and the signature card had headings such as "*Birthplace*", "*Residence*", "*Mother's maiden name*", "*Occupation*" and "*Age*". Hayden described that a lot of these fields would often be left blank however "*Mother's maiden name*" was very useful as a test of identity as it was not commonly known outside the family and would be used as extra identification when making a withdrawal. (Hayden, 1906)

Modern applications and ideas for security questions

Despite the fact that security questions are an outdated and insecure type of authentication there is still research being carried out in how could shared secrets be used as a suitable means of account recovery or authentication. Other authentications methods might be more suitable for account recovery such as email-based authentication or mobile phone based authentication however these forms out security while at a glance are more secure also have their own set of problems and challenges for example email accounts are an issue as they are susceptible to phishing attacks while mobile phones are susceptible to malware and as personal affects are often broken or lost.

Generating security stories for better protection of user privacy.

In 2020 Armin Anvari, Lei Pan and Xi Zheng released an article in the “International Journal of Computers and Applications” about the common pitfalls of security questions and other traditional means of authentication for account recovery. In this article they carried out a study on fallback authentication that utilized Propp’s theory to generate unique and interesting stories and have users answer questions based on the story to be used as the shared secret.

Propp’s theory

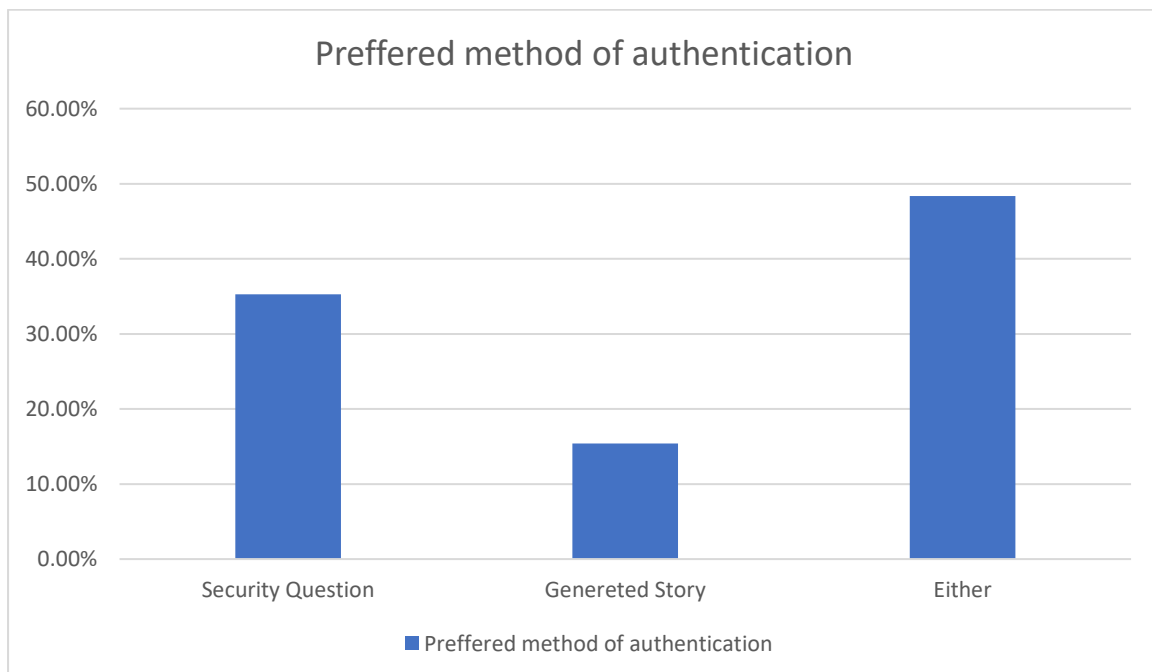
Vladimir Propp’s theory is based on his dissection of over a hundred Russian folk tales to discern seven main archetypes for every story that follow a select sequence, the seven archetypes and their sequence are:

- The Villain. At the start of the story the villain causes some type of evil or damage for their own personal gain.
- The Hero, reacts to the dispatcher. The hero can be of two different types a seeker and a victim, a seeker is someone who agrees to help defeat evil because someone else is suffering and the victim is someone who directly suffers from the action of the villain.
- The Donor is someone who provides necessary things/guidance to the hero that is necessary to defeat the evil they will face later.
- The Helper is someone who helps the hero by helping them acquire an object or break a spell usually using a power or trait that the hero doesn’t possess.

- The princess or Prize. The princess or sought after item is usually why the hero pursues the villain in the first place, the story usually ends when the hero finds who or what they are looking for.
- The Dispatcher is someone who sends the hero on their quest. If it's a princess who has been stolen by the villain, the dispatcher will be the king. If the hero themselves were the victim then someone who can point them in the direction of the villain might be the dispatcher.
- The false hero is someone who will seem good and just until the very end where they will try to take credit for the heroes work in which case they will show themselves to be actually corrupt.

(Vladimir Propp, An introduction to Propp's 7 character types, spheres of action and narrative functions, n.d.)

Based on this theory the researchers conducted a study by creating 3 stories and a set of questions for each and having a large group of people read through the story and answer the questions. The stories that were created followed Propp's theory and each had 4 multiple choice questions to answer. In this study they concluded that stories were a time consuming way for fallback authentication as people who spent more time reading over the generated story had answered more answers correctly. The researchers also found that the more predictable the story the less memorable it is. At the end of the study the researchers asked the participants to indicate their favoured method of authentication. They found that only 16.39% people wanted to use generated stories with 35.23% percent stating that they would use security questions and 48.36% who would use either.



(Armin Anvari, 2017)

Main issues surrounding security questions.

Guessability

Armin Anvari, Lei Pan and Xi Zheng summarized that security questions greatly suffer from two major factors: guessability and memorability. The probability of guessing an answer to a security question by someone else is one of their biggest weaknesses. By common knowledge or general information the answer could be predictable, while more personal questions like “what’s your mother’s maiden name” could be guessed using census estimates and general statistics. Memorability is another thing that security questions suffer from. In fact many users tend to forget their answers to security questions. There is an inverse relationship between memorability and the power of security, the more secure a question is the less likely people are to remember the answer. Time also affects memorability, the longer the time period between account creation the less likely people are going to remember the answer. Both great downsides, memorability and guessability, contradict each other as well. By choosing a more secure less guessable answer or giving a fake answer the less likely you are to remember it. (Armin Anvari, 2017)

The security of an answer highly depends on how well it is prepared against the threat model for which the following metrics are:

- Questions with common answers. A lot of common knowledge questions have are often picked by people to answer as they are easy to remember however this causes the answer to be the same as lot of other peoples which makes it insecure and easy to guess.
- Questions with few plausible answers. These are questions where the pool of answers to choose from is very small. For example, The questions “Who is your favorite superhero?” has very few answers considering that majority of the population would use a very mainstream and pop culture inspired answer such as “Iron man”, This makes it easy to guess the security answers of a large number of people.
- Questions with publicly available answers. These are answers that can be publicly found online such as where someone went to school listed on social media or the name of their pet being written in a picture post.
- Social guessing attacks. A lot of personal security answers may be known by family and friend, even acquaintances from school or work. It was found in a laboratory study that 17% of acquaintances

were able to guess the answers of the person in study in five tries or less.

(Nicholas Micallef, 2021)

Memorability

As it was mentioned before it is clear that memorability has a direct relationship with the security/difficulty of a question. The more secure a question and answer combination it is the less memorable it will be. In the graph below we can see that there is a direct relationship with memorability and time. The longer it has been since the enrollment of the account the less memorable the question is. However, this is even more true when the questions are more difficult. Questions that require you to remember of sequence of numbers such as a Library card number and a Frequent flyer number only have an average 15.6% success rate in account recovery. Yet the answers to those questions are not up to interpretation and don't change over time.

question	Overall success	After 1 month	After 3 months	After 6 months	After 12 months
<i>City of birth?</i>	80.2%	83.9%	79.9%	79.2%	79.5%
<i>Father's middle name?</i>	75.6%	85.9%	75.7%	74.4%	74.3%
<i>Childhood best friend?</i>	68.5%	82.9%	65.0%	64.6%	63.7%
<i>High school name?</i>	67.3%	78.8%	62.8%	62.6%	61.4%
<i>First phone number?</i>	55.2%	70.0%	55.4%	53.3%	50.1%
<i>Favorite food?</i>	48.0%	73.6%	52.8%	50.1%	46.6%
<i>First teacher's name?</i>	47.1%	71.7%	45.9%	43.2%	39.8%
<i>Library card number?</i>	22.5%	49.6%	24.3%	19.9%	17.7%
<i>Frequent flyer number?</i>	9.0%	32.1%	8.5%	6.4%	6.4%

A table demonstrating the success rate of answering questions by the number of months that have passed since the creation of account.

Another huge factor that affects memorability are inaccurate answers. People give inaccurate or false answers to security questions to make the more secure i.e., Giving a made up maiden name for "*What's your mother's maiden name*". This makes the answer a lot harder to remember as it no longer ties the question to the answer with any personal significance or meaning. This finding also applies to questions that are answered in shorthand or inaccurately. The question "*What was your first phone number*" may have a maximum field length of 13 characters, 3 for country code and 10 for actual phone number. However, a study carried

out on memorability shows that a fully accurate answer of the questions along with any spaces or punctuation has a 68% accuracy, which dips below to 28% if there is even one character left out. (Bonneau, 2015)

Geographical factors

Understanding cultural differences and geographical factors is key to designing security questions for fallback authentication. In figure 1 we can see that France has the highest recall rate for first phone number while having the lowest recall rate for fathers middle name. Other cultural differences affect the effectiveness of a shared secret, such as in some Spanish cultures the questions "Father's middle name?" is replaced with "Primer apellido del padre?" which means Fathers second surname as its quite common for people to have multiple surnames is Spanish speaking countries thus increasing memorability. (Bonneau, 2015)

Language	Country	<i>1 month since registration</i>	<i>3 months since registration</i>	<i>6 months since registration</i>	<i>12 months since registration</i>
<i>English</i>	US	70.0%	55.4%	53.8%	49.8%
<i>English</i>	UK	68.4%	52.1%	49.7%	44.2%
<i>German</i>	Germany	69.2%	44.6%	42.3%	37.7%
<i>Spanish</i>	US	70.0%	59.2%	59.1%	57.6%
<i>Spanish</i>	Spain	68.6%	47.5%	41.5%	37.9%
<i>French</i>	France	75.6%	59.2%	58.5%	57.0%

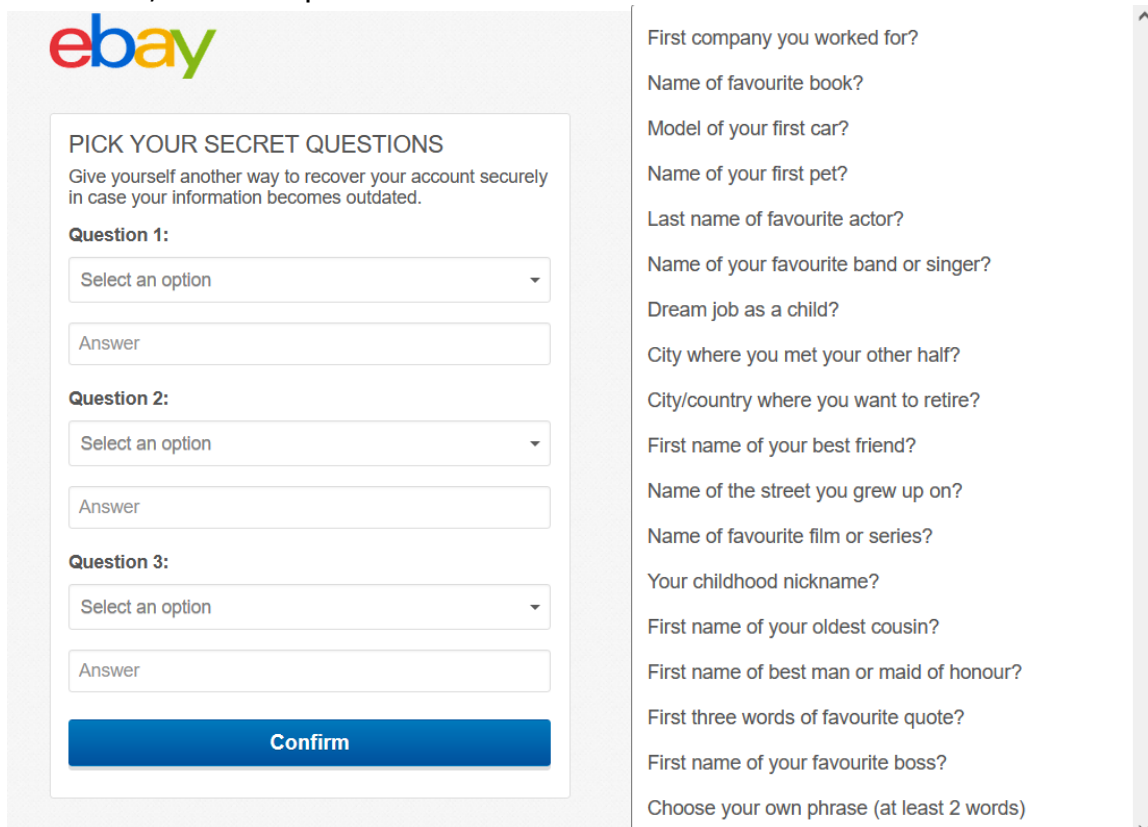
Figure 1: A table for memorability of the "First phone number?" In different countries

Language	Country	<i>1 month since registration</i>	<i>3 months since registration</i>	<i>6 months since registration</i>	<i>12 months since registration</i>
<i>English</i>	US	85.9%	75.7%	75.1%	74.4%
<i>English</i>	UK	81.2%	68.0%	64.6%	64.1%
<i>German</i>	Germany	81.9%	68.0%	64.4%	64.4%
<i>Spanish</i>	US	88.3%	81.3%	82.2%	80.8%
<i>Spanish</i>	Spain	85.3%	71.7%	70.2%	62.8%
<i>French</i>	France	56.8%	39.6%	37.6%	36.9%

Figure 2: A table for memorability of the "Father's middle name?" In different countries

Technologies

During my research for this project I discovered that still quite a few companies use security questions as a password recovery method. Some of these can be quiet surprising, especially as some of them store banking details and a lot of other private information that is related to their accounts, such as phone numbers and addresses.



ebay

PICK YOUR SECRET QUESTIONS
Give yourself another way to recover your account securely in case your information becomes outdated.

Question 1:
Select an option
Answer

Question 2:
Select an option
Answer

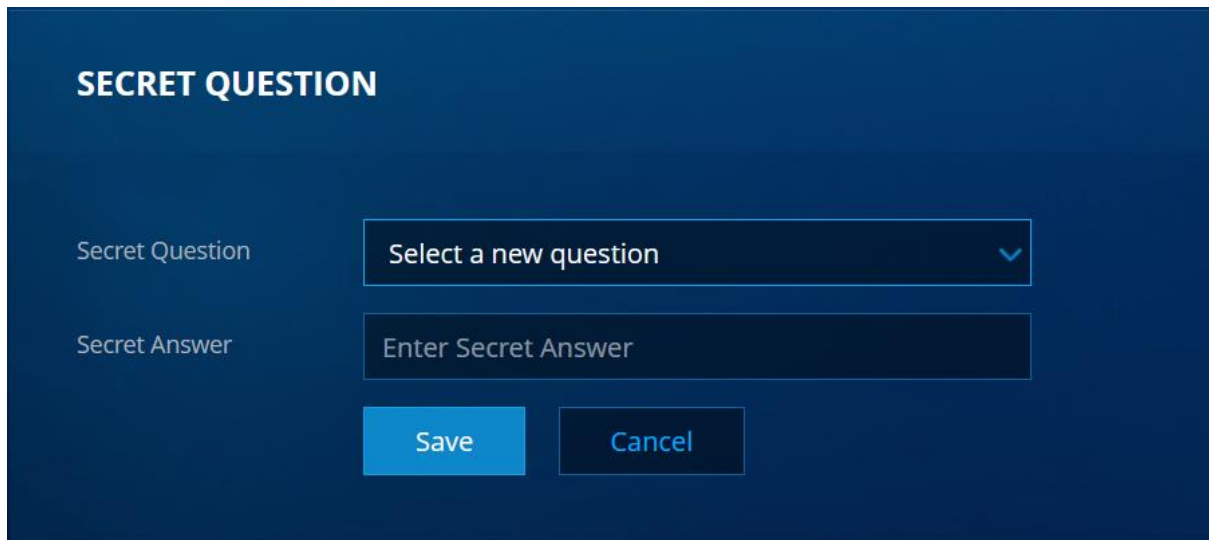
Question 3:
Select an option
Answer

Confirm

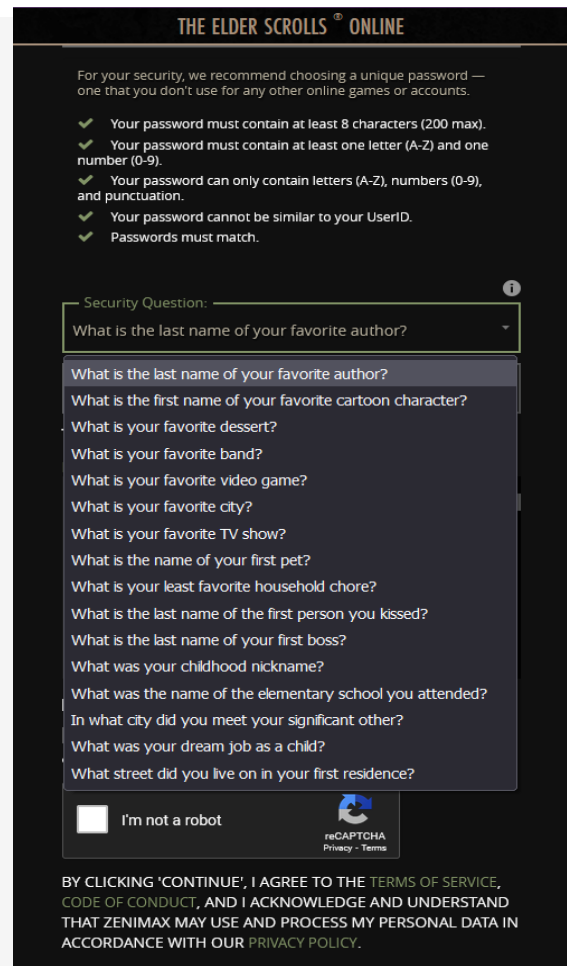
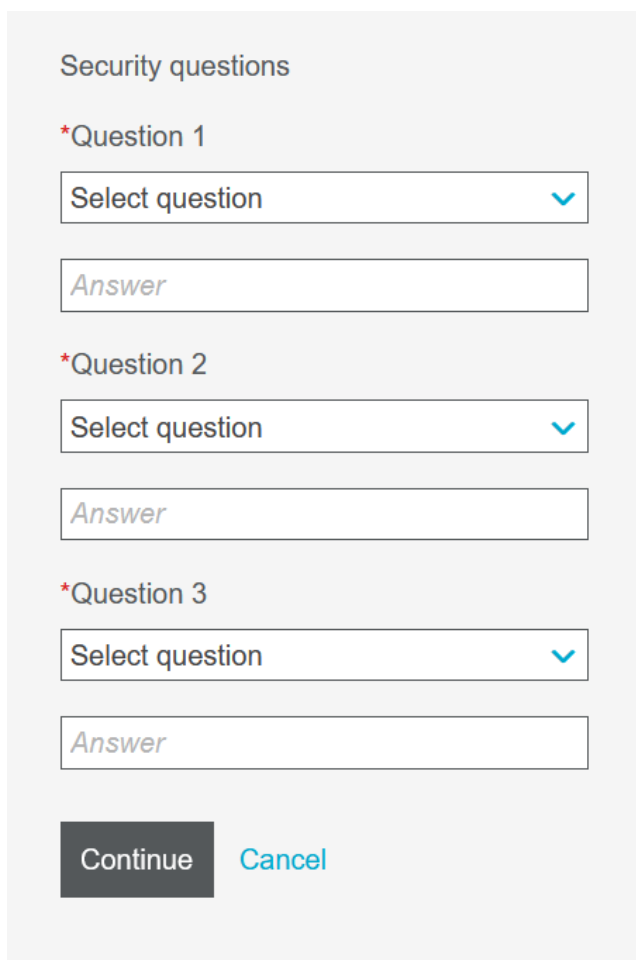
- First company you worked for?
- Name of favourite book?
- Model of your first car?
- Name of your first pet?
- Last name of favourite actor?
- Name of your favourite band or singer?
- Dream job as a child?
- City where you met your other half?
- City/country where you want to retire?
- First name of your best friend?
- Name of the street you grew up on?
- Name of favourite film or series?
- Your childhood nickname?
- First name of your oldest cousin?
- First name of best man or maid of honour?
- First three words of favourite quote?
- First name of your favourite boss?
- Choose your own phrase (at least 2 words)

(ebay, 2022)

These are screenshots of ebay's account recovery method as of March 2022. What is even more surprising is that this dialog option came up after account creation as a method to 'secure' your account and set up for recovery. Ebay are not the only company that are providing security questions after account creation to 'secure' your account.



(Battle.net, 2022)



(Primark, 2022), (The elder scrolls online, 2022)

Seeing technology like this is what built the motivation for this project and it's goal to secure accounts and protect user private information that is stored on websites like these.

Existing applications

While I haven't found any dedicated security question generating and storage solutions, I have discovered that this is a problem that other people have been looking to solve for a while. The most common method is to use password lockers. Password lockers are tools that store your credentials with your permission and the next time you need to log in to the same website they will fill out the username and password field for you, simple right? Well it doesn't work the same for security questions as password lockers and managers don't consider security questions as login credentials, therefore when recovering your account they can't be used to automatically fill out the answer field. What changes this is if a password locker or manager has a secure note section or tab where you can add a note that will be securely stored with your passwords. This allows users to enter their question and answer as a note and refer to it manually whenever they need it. (Notenboom, 2018) This solution offers a fix to only one of the main problems surrounding security questions: memorability. Guessability isn't affected by this at all and the fact that it is the more crucial of the two main issues to address just shows that there is a real need for a dedicated tool that solves both problems.

As I mentioned earlier future applications for security questions and shared secrets are being studied and they are moving away from personal information and rather towards shared memorable information between client and host. Despite this the more favorable approach to account recovery is two factor authentication via text message or email. My solution isn't so much as a permanent solution for security questions rather than a stop gap until they are phased out and not used at all anymore, at least not in their present form.

Proposed solution

My proposed solution to these issues is creating an application that will solve the two major issues to security questions: memorability and guessability. These two factors affect security questions the most in regards to account security so these will be the ones that I will focus on. For this project I intend to develop an android application as majority of people in the world use their mobile phone to create accounts for online shopping, social media and general web use. I will aim at android users as 71.7% of all phones used in the world run android. Another reason for targeting android users is that implementing applications in android is a lot more accessible in regards to functionality, local storage and

permissions all of which are necessary for this application. With less time constraints and greater development knowledge, it would be possible to develop an iOS application and may be a consideration for the future.

Development

The application will be developed in android studio using the java language as I am already familiar with this language and can spend more time getting familiar with application development rather than learning Kotlin. During development, I plan to use many different libraries however the most notable one will probably biometric prompt as it is key to the application. I will allow the user to authenticate with the application using biometrics such as facial recognition or fingerprint scanner. The application will have a choice for both because there are some devices that come with either one and not both. The plan is to have an application that will generate a secure answer to a security question for account creation and will then encrypt and store that answer on the device. A web plugin will also be created to assist in securing your account when creating on a PC browser.

Technology Stack

Android Studio

Android studio is the integrated development environment. This is a possible IDE that I will use for this application. It was created by google specifically for android development and has many integrated tools. Most notably, the android virtual device emulator which allows you to run android virtual machines to test you application. More information on Android Studio can be found here:

<https://developer.android.com/studio/releases>

MIT App Inventor

MIT app inventor is a visual programming tool that allows everyone to build fully functional app for android phones, iPhones and tablets. This is another possible tool that I may use for this application. The MIT app inventor uses a block based building system to assemble application on a web based interface. More information on the MIT app inventor can be found here: <https://appinventor.mit.edu/about-us>

Samsung A51

The Samsung A51 is the device I will be using to carry out the physical testing of my application. Since my application requires biometric authentication I need a physical device to test it on. I chose this device in particular because it has both a on screen fingerprint scanner and also facial recognition. Another reason I chose this mobile phone is because it's the one I own. More information can be found here:

<https://www.samsung.com/ie/smartphones/galaxy-a/galaxy-a51-black-128gb-sm-a515fzkveua/>

Emulation

Both Android Studio and MIT app inventor provide emulators for android devices to build and test your applications on. As mentioned previously I will be working on a Samsung A51 and so I would like for my emulator to match the real device. For this purpose Samsung provides size setting and skins to set up your emulators with as part of their development tools which can be downloaded free of charge for any device. Find out more about these here: <https://developer.samsung.com/galaxy-emulator-skin>

SQLite

SQLite is the most used database engine in the world and comes as a default for local storage in android, functionality for SQLite comes built in to all mobile phones in the world. I decided to use this database engine because while I was carrying out my research on this application I found that all local storage solutions used SQLite and it offered the most functionality and support. More information on SQLite here: <https://www.sqlite.org/index.html>

Chrome Browser

The solution that I plan to implement includes a browser extension that allows for QR code generation. I have never developed a browser extension but from my research it appears that browser extensions are developed in HTML and support languages such as JavaScript. There are minute differences from browser to browser in their extension development and Mozilla Firefox supports both Firefox and Chrome specific namespaces for ease of integration. (Mozilla , 21) I will focus on Google Chrome as the browser for developing my extension. The reason for this being is that the most of resources I found in my research were for Chrome. Chrome was originally developed and released in 2008 built with free components from the Apple WebKit and Mozilla Firefox. As of October 2021 it is estimated that chrome has a 68% market share worldwide for web browsers. Find out more about chrome here: https://www.google.com/intl/en_ie/chrome/browser-features/#

Languages

Java

Java is a fully supported language on android studio and I have studied java in varying application over the last 4 years. However I have never used java for android development. Using java will provide me with other functionality such as its cryptographic library that I am also familiar with and easy integration with SQLite.

Kotlin

Kotlin is Google's preferred android development language and it is designed to interoperate fully with Java. Kotlin's standard library depends on the Java class library and in terms of android development Kotlin compiles to java bytecode on android studio. The aim of Kotlin is to replace Java as its considered a stronger object-orientated language.

XML

XML or Extensible Markup Language is going to be used as the language to configure android views for each page of the application. This language defines data and how its organized. Specifically in android, XML is used to implement user interface related data. (GeeksforGeeks, 2021) It makes the UI lightweight and it categorises everything into tags making it easy to configure.

SQL

SQL or Structured Query Language will be used to pre construct statements for inputting data into the database and storing it. SQL is used to communicate with the database by writing queries. My use of SQL in the project should be minimal as I will only need a few statements but it's very necessary.

Other Possible Languages

HTML and JavaScript

HTML and JavaScript might be necessary for this application when building the web plugin as browser extensions are developed in HTML and from the research I have carried out, scripts that are required to add functionality to the extensions are written in JavaScript. For external operation such as generating QR codes and functions for retrieving information from webpages.

Conclusion

The next time a new account with a security question is created hopefully more thought and consideration is put into it, knowing the history and it's uses in the past. The consequences of choosing an answer that is too easy to guess or find out can be devastating to account security. Entering an answer that seems really secure at the time but in reality is something that is easy to forget will prevent most people from recovering their account. It's possible that in the near future, you could look at security questions on an account security page and it would look completely different. Instead of asking for personal information, it would challenge the user to memorise a phrase or an entire story. The next time a security question is encountered, an application made exactly to address all the issues talked about in this document may be available to aid the user in protecting their account.

Regardless of what the next encounter with security questions may be, they are without a doubt being phased out from our account security and any solutions proposed in this paper are more so a stop gap until they don't exist anymore rather than a permanent solution to making security questions 'secure'. Account security is important above all in this day and age and having peoples private information protected is not up for consideration but should be a must of every company, business and platform out there.

After carrying out in-depth research on the subject and exploring many options for security questions the only recommendation is to avoid them in account security. If it's necessary and cannot be avoided then use of the solutions proposed in this paper such as an application to generate and store secure answers should be used to ensure that user accounts that store personal information are as secure as they possibly can be.

Bibliography

- Armin Anvari, L. P. (2017). Generating security questions for better protection of user privacy. *International Journal of Computers and Applications*, 329-350.
- Battle.net. (2022, April 12). *Security*. Retrieved from Battle.net: <https://account.battle.net/security#recent-activity>
- Bonneau, J. (2015). Secrets, Lies, and Account Recovery. *World Wide Web Conference*, 18-22.
- ebay. (2022, 03 24). Retrieved from ebay.ie: https://reg.ebay.ie/reg/ChangeSecretQuestion?flow=SIGN_IN&ru=https%3A%2F%2Fwww.ebay.ie%2F
- GeeksforGeeks. (2021, July 23). *A Complete Guide to Learn XML For Android App Development*. Retrieved from GeeksforGeeks: <https://www.geeksforgeeks.org/a-complete-guide-to-learn-xml-for-android-app-development/>
- Haber, M. J. (2020, June 5). *Reused Security Questions can Pose a High Risk: Learn Tips & Tricks to Mitigate the Threat*. Retrieved from www.beyondtrust.com: <https://www.beyondtrust.com/blog/entry/reused-security-questions-can-pose-a-high-risk-learn-tips-tricks-to-mitigate-the-threat>
- Hayden, W. M. (1906). Systems in Savings. *The Banking Law Journal*, 909.
- Mozilla . (21, April 2022). *Chrome incompatibilities*. Retrieved from Mdn web docs: https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Chrome_incompatibilities
- Nicholas Micallef, N. A. (2021). Understanding users' perceptions to improve fallback authentication. *Personal and Ubiquitous Computing*, 893-910.
- Notenboom, L. A. (2018, November 7). *Ask Leo*. Retrieved from <https://askleo.com/how-can-i-use-a-password-manager-for-my-security-questions/>: <https://askleo.com/how-can-i-use-a-password-manager-for-my-security-questions/>
- Primark. (2022, 14 April). *Carrers Create Account*. Retrieved from brassring.com: <https://krb-sjobs.brassring.com/TGnewUI/Search/Home/Home?partnerid=30055&siteid=5837#CreateAccount>

Ruberg, B. (2017). What Is Your Mother's Maiden Name? *Feminist Media Histories*, 57-81.

The elder scrolls online. (2022, February 24). *Create an account*. Retrieved from accounts.elderscrollsonline.com:
<https://account.elderscrollsonline.com/register/account-information>

Vladimir Propp, *An introduction to Propp's 7 character types, spheres of action and narrative functions*. (n.d.). Retrieved from Media Studies: <https://media-studies.com/propp/>