DEGREE PROJECT IN COMPUTER ENGINEERING, FIRST CYCLE AND DEGREE PROJECT IN INFORMATION AND COMMUNICATION TECHNOLOGY, FIRST CYCLE
STOCKHOLM, SWEDEN 2017

Two-Factor Authentication

Selecting and implementing a two-factor authentication method for a digital assessment platform

NIKLAS TELLINI and FREDRIK VARGAS

KTH ROYAL INSTITUTE OF TECHNOLOGY
INFORMATION AND COMMUNICATION TECHNOLOGY

Two-Factor Authentication

Selecting and implementing a two-factor authentication method for a Digital Assessment Platform

Niklas Tellini and Fredrik Vargas

2017-05-31

Bachelor's Thesis

Examiner
Gerald Q. Maguire Jr.

Academic adviser
Anders Västberg

KTH Royal Institute of Technology
School of Information and Communication Technology (ICT)
Department of Communication Systems
SE-100 44 Stockholm, Sweden

Abstract

Two-Factor Authentication (2FA) is a two-step verification process that aims to provide an additional layer of security by requiring the user to authenticate himself/herself using a secondary means (ownership factor or inheritance factor). Without the use of 2FA, an attacker could gain access to a person's devices or accounts solely by knowing the victim's password, while with 2FA knowing only this password is insufficient to pass the authentication check. In this project, we analyze different methods in which 2FA could be implemented by a Digital Assessment Platform. These platforms allow test assessments to be built directly into digital content; therefore, an important requirement of these systems is secure authentication. Moreover, it is important to securely protect teachers' account in order to avoid unauthorized people gaining access to those accounts. We investigate how 2FA could be used to add an extra layer of security to teachers' accounts, focusing on cost, user experience, ease of use, and deployment of the solution. We arrived at the conclusion that 2FA through an ownership factor is a suitable method and we implemented a solution based upon One-Time Passwords. This thesis project will hopefully benefit Digital Assessment Platforms who wish to implement 2FA by providing broader knowledge regarding this subject. The project should also benefit society by increasing the general knowledge of 2FA, hence leading to more secure services.

Keywords

Two-Factor Authentication, Security, One-Time Passwords, Access control, Digital Assessment Platform

Sammanfattning

Tvåfaktorsautentisering (2FA) är en tvåstegs verifieringsprocess som syftar att ge en extra nivå av säkerhet, i och med att den kräver användaren att autentisera sig själv genom en sekundär faktor (något man äger eller har ärvt). Utan användning av 2FA, kan en förövare få åtkomst till en persons mobila enhet eller konto endast genom att kunna offrets lösenord. Att enbart kunna lösenordet är inte tillräckligt för att en autentiseringsprocess ska vara godkänd om 2FA är implementerad. I det här projektet analyseras olika 2FA som skulle kunna implementeras av en digital utvärderingsplattform. Sådana plattformar förvandlar tester och prov till digitalt innehåll och kräver därför en säker autentisering. Dessutom är det viktigt att säkra lärarnas konton för att undvika att icke auktoriserade personer loggar in på deras konton. Vi undersöker hur 2FA kan användas för att lägga till en extra nivå av säkerhet på lärarnas konton, med fokus på kostnad, användarupplevelse, lättanvändlighet och utplacering av lösningen. Vi kom fram till att 2FA via en faktor man äger är en passande metod och vi implementerade sedan en lösning grundad på engångslösenord. Detta projekt kan förhoppningsvis vara till förmån för digitala utvärderingsplattformar som vill implementera 2FA, genom att ge en bredare kunskap inom detta område. Projektet skulle kunna gynna allmänheten genom att bidra till ökad generell kunskap om 2FA, och därav leda till säkrare tjänster.

Nyckelord

Två-stegs autentisering, Säkerhet, Engångslösenord, Åtkomst kontroll, Digital bedömningsplattform

Acknowledgments

We would like to thank examiner Gerald Q. Maguire Jr. and supervisor Anders Västberg for their help. Also, we would like to thank Robin Andersson at DigiExam as well as the other employees for their support and help throughout this project.

Stockholm, May 2017
Niklas Tellini
Fredrik Vargas

Table of contents

List of Figures

List of Tables

List of acronyms and abbreviations

CAPEX	Capital Expenditure
CNSS	Committee on National Security Systems
DAP	Digital Assessment Platform
ECB	European Central Bank
EER	Equal Error Rate
ERP	Event Relevant Potential
ERR	Equal Error Rate
FAR	False Acceptance Rate
FER	Failure to Enroll Rate
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Rejection Rate
FTA	Failure to Acquire Rate
GSM	Global System for Mobile Communications
HCI	Human-Computer Interaction
HTTPS	Hypertext Transfer Protocol Secure
ICT	Information and Communication Technology
ITU-T	Telecommunication Standardization Sector
LMS	Learning Management System
LOA	Levels of Assurance
MNO	Mobile Network Operators
NIST	National Institution of Standards and Technology
OATH	Initiative for Open Authentication
OTP	One Time Password
PIN	Personal Identification Number
PKI	Public Key Infrastructure
QR	Quick Response
SFA	Single-Factor Authentication
SIM	Subscriber Identity Module
SME	Entrepreneurship and small and medium-sized Enterprises
SMS	Short Message Service
SSL	Secure Sockets Layer
SSO	Single Sign-On
TAR	True Acceptance Rate
TMR	True Match Rate
TNMR	True Non-Match Rate
TRR	True Rejection Rate
TOTP	Time-Based One-Time Password
2FA	Two-Factor Authentication
UX	User Experience
WWW	World Wide Web

1 Introduction

This chapter describes the specific problem that this thesis addresses, the context of the problem, the goals of this thesis project, and outlines the structure of the thesis.

Any system that implements access control must be able to verify that someone is who he or she claims to be. The ways in which someone can be authenticated are usually divided into three categories, known as authentication factors:

Knowledge factors something the user knows - this could be a password, a Personal Identification Number (PIN), or a shared secret

Possession factors something the user has – this could, for instance, be an ID card, a security token, or a smartphone

Inherence factors something the user is – these factors, also known as biometrics, are personal attributes such as fingerprints, face, and voice

It is important to state that some systems with more advanced requirements for security may use location and time as factors too.

The traditional authentication method is known as single-factor authentication (SFA). SFA is a security process in which only one factor (typically a password) is used to confirm a user's claimed identity. This method is considered insufficiently secure for many security-critical applications, such as online banking or logins to personal accounts. In fact, approximately 63% of confirmed data breaches involved a weak, default, or stolen password in 2016 [1]. The main problem with password-based authentication is that it requires knowledge and diligence to create and remember strong passwords. Further, passwords are also prey to threats such as brute-force or dictionary attacks. Hashing passwords before storing them, the introduction of password salting, and the use of non-dictionary passwords or passphrases are good means to strengthen the SFA method, but passwords are still insufficient for higher security applications.

Two-Factor authentication (2FA) strengthens access security by requiring two methods to verify a user's identity. This additional layer of security makes it harder for attackers to gain access to a person's devices or online accounts. Passwords are the most common form of SFA because of their low cost, familiarity, and ease of implementation. In contrast, the multiple challenge-responses cycles that come with 2FA can provide greater security and can protect against phishing, social engineering, and password brute-force attacks, hence securing logins from attackers exploiting weak or stolen credentials. When 2FA is enabled, a user has to provide both the first and the second factor for authentication. This means that the probability of the system being secure is calculated by the following simple formula

$$S = 1 – (F_1 * F_2),$$

where S is the probability of the authentication system being secure, F_1 is the probability of passing the first factor and F_2 is the probability of passing the second authentication factor. For instance, if the probability of guessing a password is 30% and the probability of passing the second authentication factor is 5%, then the probability of passing both the first and the second authentication factors is 1,5%.

This thesis project took place at DigiExam, a leading Swedish company in the digital assessment platform sector. The research was conducted in a way that hopefully would benefit this company. A digital assessment platform is a service that allows students to take exams digitally. This revolutionary way of taking exams has lots of benefits for teachers, students, organizations that use it, and the environment. Teachers utilizing DigiExam to create exams save up to 50% of their working time, can co-grade with their colleagues, have the possibility to create digital exams and assessments in a reliable, intuitive platform, and have everything they need in a single place. The schools and organizations using DigiExam save money and reduce their impact on the environment by reducing paper use, avoid investing in hardware that becomes outdated, and removing teacher bias by allowing anonymous grading. Students also benefit from DigiExam because one writes faster on a computer than by hand, it is easier to delete or edit written text, can forget about writer’s cramp, and can rely on a reliable platform that works both online and offline. Additionally, students can use their own familiar device.

Naturally, security is a major requirement for such a digital assessment platform service. It is of considerable and substantial importance that only authorized persons can access DigiExam’s service and that the authenticated user is given access to only those resources the user is approved for. This means that a student should never be able to access the teacher’s account because this could lead to undesirable consequences that could affect grading of the examinations or even the contents of the exams themselves. Today, DigiExam relies on SFA, i.e. each user can log in to their account providing the correct username and corresponding password. As stated before, SFA is considered weak security and insufficient security for a security-critical service such as DigiExam. The company would like to mitigate this problem and therefore wants to gain insight into different 2FA methods and evaluate which methods they could benefit the most from.

This thesis project analyzes different 2FA methods and attempts to determine which of them best suits the needs and requirements of a digital assessment platform, in terms of cost, deployment, security, and user experience (UX).

1.1 Background

Online security is crucial. The Internet has evolved tremendously over the last several years and computer networks are becoming bigger and bigger. A consequence is that network security has become one of the most important factors for companies to consider. Today we perform financial payments via the Internet using mobile money transfer services, we perform banking operations online, and we make use of online health services. As a result, our data is everywhere: on our phones, laptops, work PCs, servers, and retailers' computer networks. In addition, hackers are becoming more sophisticated and offensive hacking tools are numerous and cheap.

Lots of world-famous companies have been attacked, including Sony, Adobe, Evernote, and LinkedIn. The biggest data breach in history was revealed in December 2016 when Yahoo said 1 billion accounts were compromised in 2013 [2]. These cyber-attacks have enormous consequences in terms of cost for the involved businesses. A research that Juniper published in 2015 predicts that cybercrime will cost businesses over US$2 trillion by 2019 [3]. Attacks on large banks, retailers, and government agencies become worldwide news – but all business are actually at risk. According to Symantec, 52.4% of phishing attacks carried out in December 2015 were against Entrepreneurship and Small and medium-sized enterprises (SMEs) [4]. The earlier impact of these attacks has clearly shown that the consequences for SMEs who ignore security risks can be disastrous.

In addition to the loss of revenue, cyber-attacks may also cause damages to business reputation, breach of confidential information, and loss of customers. By increasing network security the risk of privacy spoofing and identity or information theft are decreased. Improving network security could be done in several ways. Creating a network security policy is certainly a good first. A clear and comprehensive network security policy outlines a user policy and is meant to govern data access, web-browsing habits, and use of passwords and encryptions. Keeping the network up-to-date by updating the operating system, antivirus software, firmware, and device drivers will strengthen the system against cybercriminals who launch their attacks by taking advantage of known security flaws in old versions of the software. Additional actions include installing a firewall, blocking users from installing software, and to adopting a strong password policy. All of these actions can improve network security.

Internet security is of vital importance because it is essential for protecting personal and business information. Cyber-attacks allow hackers to steal credit and debit card information, damage reputations, and cause financial losses for the victims. For businesses and government entities, the Internet provides a means to allow customers to carry out financial transactions without having to meet or talk on the phone. However, if this information is not secured in an appropriate way the consequences may be destructive.

There are lots of different methods an attacker could utilize in order to learn a victim's password. A dictionary attack is perhaps the most famous form of attack, as a dictionary attack compares the password file with pre-computed hashes and hence can quickly find one or more user's passwords. A brute force attack is similar to a dictionary attack, but in this attack, the hacker can find non-dictionary words by exhaustively trying all possible alpha-numeric combinations. Another famous attack is phishing, where the hacker sends email to the victim that directs them to a fake online banking, payment site, or another site to trick the victim into logging in via this site, hence providing the hacker with the username and corresponding password. Often an attacker will make use of social engineering in order to psychologically manipulate the victim into divulging confidential information. Other ways of performing a password attack include trying popular passwords against a large number of accounts or exploiting user's mistakes (writing down passwords, sharing passwords with friends, not changing a device or account's default password).

Creating a strong password is not difficult, but it requires knowledge and diligence to create such passwords, remember them, and keeping them safe. First of all, a unique password should be used for each account. Re-using passwords is risky because if someone figures out a password for one account, then it is obvious that this same password should be tried for the user's other accounts. Furthermore, a mix of letters, numbers, and symbols should be used when creating a password. An eight-character password that possibly contain numbers, symbols, and mixed-case letters is harder to guess because it has 30,000 times as many possible combinations as an eight-character password with only lowercase letters [5]. A password should be unrelated to personal information. Also, simple words or sequential patterns are strongly not recommended. Lastly, passwords should be kept secure, for example by using a password manager. As a consumer account security report by mobile security company TeleSign shows, the main problem with SFA is that users continue to ignore security advice in favor of convenience [6]. This report by TeleSign shows that despite 80% of consumers being worried about online security, 45% being very concerned about their accounts being hacked, and only 30% being confident that their passwords will protect the security of their online accounts, consumers rarely change their password (47% are using a password that has not been changed in 5 years). Moreover, 73% of accounts use duplicate passwords. In addition to that, people use simple passwords to protect their account. A report from Keeper [7] shows that over 50% of the 10 million passwords that were analyzed in 2016 are in the top 25 of the most common passwords and that nearly 17% of users are safeguarding their accounts with "123456".

The report from TeleSign also states that 68% of people say they want companies to provide an extra layer of security and that 86% of people who use 2FA feel their accounts are more secure. Therefore, it is time for companies who rely on SFA to change their password policies and authentication methods and to start utilizing 2FA to provide this extra layer of security. Currently, DigiExam does not yet support 2FA but wants to increase the security of their service by exploiting 2FA.

As noted earlier, the service offered by DigiExam should be properly secured. Without appropriate security, a student could perhaps access another student's account and take a test instead of the actual student; hence affecting the student's assessed performance. Worse, students could exploit and take advantage of the insufficient security that comes with SFA to gain access to a teacher's account and be able to edit grades or access exam questions *before* the exam takes place. It is easier to discover the first breach because proving that a specific person has not written a text can be done by looking at the syntax, grammar, and vocabulary (i.e. the words that have been used). On the other hand, the second type of breach is more difficult to detect. The student-teacher ratio within the EU ranges from 8.0 in Latvia, Malta, Lithuania, and Greece to 19.3 in Turkey [8]. While a teacher might suspect that a student's grades have been modified, it would be more difficult to detect an intrusion into the teacher's account that leads to a test being distributed to students in a class before the exam takes place.

However, the increasing number of students and flat or decreasing budgets negatively affect assessment logistics; hence automation and digitalization of assessment mechanisms seem to be a necessary and indispensable educational development today. The computerization of the assessment process facilitates procedures that exploit computer based design, delivery, analysis, and scoring of assessment instruments [9]. Teachers seem to be positive towards this type of assessment, as they save time in testing and marking tests [10]. In addition, teachers can take advantage of the easier scheduling and administration of assessments [10]. Students also seem to prefer digital assessment instead of written assessments, as they consider a digital assessment to be more objective, faster, and less stressful [11]. An increasing number of schools rely on digital assessment platforms; however, they also place a great importance on the security of this service. DigiExam reports that 555,000 assessments were handed in via their system during the year 2016 by over 154,000 students [12].

1.2 Problem

This thesis project aims to improve DigiExam's authentication security. As stated in the introduction and as it will be explained in greater detail in Chapter 2, DigiExam currently relies on SFA. That means that students, teachers, and employees at DigiExam only have to provide the system with the correct combination of username and password in order to gain access to any account. As noted earlier, SFA offers insufficient security for a security-critical service such as DigiExam's digital assessment platform.

Assessment is a key element in learning, as it entitles, directs, and drives students, while also furnishing criteria against which they can measure their advances and improvements. Due to assessments' educational impact, there is a growing need for certification of knowledge. Moreover, the consequences of a leaked examination could be serious. For instance, a student could crack a teacher's password and thus be able to login into the account for many years. Students' school results, grades, and achieved competencies could be falsified and hence be misleading. This is particularly serious in Sweden were admission to universities is based upon a grade calculated from a student's results during the 3 years of high school (gymnasium).

Given all of the above, it is natural that a platform such as that provided by DigiExam should allow teachers to authenticate themselves using 2FA. However, there are different methods by which 2FA could be implemented. These different methods vary in cost, UX, deployment, and security. Changing from SFA to a 2FA might require an overhaul of the company's legacy authentication system. For this reason, many companies prefer to wait to implement 2FA because although the risk of fraud that results from stolen password is large (as described earlier), it still represents a known cost while the risk of security incidents due to changing and updating the infrastructure are unknown. Also, two-factor authentication may result in customer dissatisfaction, especially if this change is forced and there could be excessive calls to the company's support service (leading to high costs). While SFA authentication is regarded as weak security, it appears to not be sufficiently weak enough to justify taking the risks that come with implementing 2FA. Therefore, it is important to offer a solution that would make migration less painful (for example, by allowing incremental changes). However, a number of questions remain, such as: How could DigiExam implement 2FA to allow teachers to strengthen their accounts' security? Which method best suits the needs and requirements of DigiExam?

1.3 Purpose

The purpose of this thesis is to offer the reader a deeper insight into 2FA solutions and to offer DigiExam the most appropriate solution for their needs and requirements. The problem should be solved because of the serious consequences that could arise due to the lack of sufficient security offered by the company's digital assessment platform services given the use of SFA for authentication at the time the user logs in.

The goals of this degree project are stated in the following subsection. If these goals are achieved, then this degree project would hopefully benefit several different parties. Firstly, it will benefit DigiExam, the company where this degree project is taking place. The solution that will be proposed for DigiExam will be the most appropriate solution with respect to their requirements and could hopefully be implemented in their service. Secondly, other digital assessment platforms could benefit from this work. Although the requirements and needs of other platforms are unlikely to be exactly the same as those of DigiExam, hence the solution may need to be adapted to each specific use case. The scientific community should also benefit from this work, as different 2FA methods will be analyzed, therefore it will be easier for other researchers to further analyze the methods that are explored in this thesis. Finally, society and the public will benefit from this thesis. Today, only 39% of consumers use 2FA and more than 60% of them only use it just because a site requires it [6]. Hopefully, this project will make consumers more aware of the risks of SFA and hence motivate them to take stronger measures to protect themselves in an appropriate way.

Ethical aspects of this project are related to computer and information security and privacy. Different ethical approaches to information technology exist. The morality of hacking, the morality of computer crime, and the moral importance of computer security are all ethical aspects of computer security. Moreover, privacy has a substantial moral importance and information technology has a significant impact on it. Privacy is also a social issue, and the benefits of maintaining privacy as well as the costs of failing to achieve or losing privacy need to be understood. As the concepts of privacy and information security are a part of Information and Communication Technologies (ICT), we also have to deal with the sustainability issues that come with the use of ICT. Technological advancements in the ICT domain are reaching new heights, but everything comes at a cost. Many businesses are driven by the motive of profit maximization and increasing competition motivates them to achieve greater market share, hence many companies neglect issues of sustainability. This tendency may cause damage to the society in the long term.

1.4 Goals

The main goal of this degree project is to provide DigiExam with a suitable 2FA method that allows teachers to securely access their accounts. A prototype implementation of the solution may be implemented. The main goal has been divided into the following sub-goals:

1. Analysis of existing 2FA methods,
2. analysis of what out-of-the-box solutions exist for 2FA, and
3. find out how account recovery works in the context of 2FA.

The deliverable of the project is an identification of a suitable 2FA method that meets the requirements of DigiExam. Hopefully, an implementation of a prototype of this method will also be delivered to the company. The results of the project will hopefully enable DigiExam to benefit from the proposed solution by implementing this method in their service (with or without modification). Alternatively, they might implement their own solution based on the insight offered by this thesis. In all cases, the aim is to increase the security of the authentication method adopted by DigiExam.

1.5 Research Methodology

This section introduces the procedures and techniques used to identify, select, and analyze the research problem. Details of the methodology and methods used will be given in Chapter 3.

The research method relies on is the empirical-analytical approach, in contrast to the interpretative approach that employs inductive reasoning. Deductive reasoning uses existing theories to formulate hypothesis. Moreover, the empirical-analytical approach is based on explanation. In contrast, the interpretative approach is focused on understanding phenomenon in a holistic/universal way. This research method focuses on subjective knowledge and requires the variables to be carefully interpreted. The interpretative method positions the meaning-making practices of humans at the center of the scientific explanation and it aims to show how those practices could generate observable outcomes. The empirical method is linked to quantitative research, while the interpretative method is linked to the qualitative method. The quantitative approach is therefore based on quantities that could be measured through quantifiable measurement processes, while the qualitative one is based on properties that cannot be measured to yield a numerical result, as it is more concerned about the nature of things. The sources of quantitative data are often surveys, observations, and secondary data. The main source of data for this thesis will be secondary data that has to be analyzed and interpreted before being used. This project will mainly follow the quantitative approach because it focuses on a larger view of the problem. This quantitative analysis will provide a possibility to summarize some characteristics or solutions. Also, a quantitative approach is more in line with the scope and goal of the project. As stated before, it is out of the scope of this thesis to analyze those factors that lead someone to access someone else's account. Qualitative findings are also harder to generalize to a more general solution and this generally requires more time than a quantitative approach. One limitation of the quantitative approach is that it relies on existing solutions, therefore making it hard to introduce a new solution. Also, collecting data is costly in term of time and money and therefore the actual study will be more theoretical than empirical and rely on secondary data, as stated above. This means that the project thesis will mainly be literature-based. This thesis will, therefore, give a theoretical analysis, in which theoretical material will be discussed and compared in terms of applicability. After the theoretical analysis, a prototype will be implemented and evaluated as well.

1.6 Delimitations

This thesis project will **not** investigate or implement a 2FA method **for students' accounts**. The reason is that DigiExam is currently uninterested in this feature because their service is mainly used during written exams taken by students that are physically present in a classroom. Therefore, 2FA is unneeded as there are human exam proctors in this classroom.

Moreover, the proposed 2FA solution may not be optimal from a user's point of view. The reason for this is that in order to optimize the solution for use by a teacher, the solution needs to be made available for teachers' accounts and then this could be followed by a survey. Based upon an analysis of the data from this survey the proposed solution could be adapted to produce a better practical solution. However, we believe that time available for this thesis project is insufficient for such an evaluation.

As the thesis is mostly literature-based and secondary data will be analyzed, the assumptions made depend on those data. For instance, some data may only refer to a particularly continent or region or there could be limitations in the ability to learn the exact scope of the desired population. Moreover, this thesis does not aim at analyzing the psychological reasons behind cheating prior to or during exams and tests. Also, this thesis only analyzes the problem of authentication and does not examine any other problems arising in digital assessment platforms. Finally, it might not be possible to analyze the current design and implementation of DigiExam's actual login processes due to the company's policies. This could result in a solution that is not truly adoptable and would need to be adapted before DigiExam could use it.

1.7 Structure of the thesis

Chapter 2 presents relevant background information about different authentication factors used to implement different authentication methods. Additionally, it gives a more detailed description of how DigiExam handles students' login. Also, Chapter 2 presents related work relevant to this project thesis. Chapter 3 presents the methodology and methods used to solve the problem, with a focus on research strategies, data collection, and data analysis. Chapter 4 presents different 2FA approaches and offers an analysis of them in terms of cost, deployment, UX, and security. Chapter 5 will present the design of the chosen method and its implementation. Chapter **Error! Reference source not found.** gives conclusions and suggests future work that could be done. Reflections will also be presented and limitations will be analyzed.

2 Background

This chapter provides relevant background information about different authentication factors and how they are used to implement different authentication methods. Additionally, this chapter describes how SFA is currently implemented by DigiExam. The chapter also describes related work and some future work is introduced in this chapter.

Authentication is the act of confirming the truth of an attribute of a single item of data or of information claimed to be true by an entity (with some stated identity). Authentication is relevant to multiple fields and it is very important in computer science. In computer science, authentication is the process through which software, computer, user, or system *verifies the identity* of another computer, software, or user that wishes to benefit from some service(s). Authentication is different from authorization (which means giving users the right to access specific resources related to information security and computer security, based on their identity).

2.1 Authentication Factors

Methods for authenticating people differ from those used to authenticate machines or programs, due to major differences in the capabilities of people and computers. Approaches for human authentication rely on the following factors:

Knowledge factors something the user knows (e.g. a password) - the most common form of authentication

Ownership factors something the user has (e.g. a smart card) - often a hardware or software token

Inherence factors something the user is (e.g. fingerprint) - authentication is based on something intrinsic to the principal being authenticated

The types of authentication available differ in their level of security. Moreover, by combining factors from the one or more of the three categories of factors the level of security can be further raised. The choice of the different methods for authentication depends on many factors, such as usability, the importance of the information that has to be protected, and the cost of the system. Sometimes a process called mutual authentication is used, where both parties authenticate each other.

2.1.1 Knowledge Factors

The idea behind the use of knowledge is that the party wishing to be authenticated know a secret (a password) that only they and the authenticating party know. Thus, knowledge of this secret distinguishes him/her/it from all others. The authentication system simply checks whether the person claiming to be X knows the secret associated with X.

Unfortunately, the use of a secret is not a panacea. As stated earlier, there are plenty of ways by which a hacker could learn a user’s password. The simplest attack is shoulder surfing: if the secret is entered at some sort of keyboard, an eavesdropper may observe the secret being entered. Furthermore, people tend to use passwords that are easy to remember, which often means that the password is easy to guess. In contrast, if a strong and non-easy to guess password is used, then it is likely that the password is difficult to remember and has to be written down (potentially making it easy for an attacker to find). Even if a password is non-trivial to guess, it could be the target of an offline search of the password space. In an offline attack, the attacker first gains access to the password hash (for example, by stealing the password file) and compares entries in the password file with a table of pre-computed hashes. In such an offline attack the attacker does not have to communicate with the target’s host until it has found a likely password. Finally, human intervention is required to change passwords and this means that compromised passwords could remain valid for longer than it is desirable. Moreover, in order to reset passwords, there must be some other mechanisms to identify the user – and these might be vulnerable to social engineering attacks. Such social engineering attacks are based on convincing a human with the authority to change or access information that it is necessary to do so. With this being said, there are three dimensions required for a password to be considered good. These are the following:

Length This is the easiest dimension for people to strengthen. An easy way of making passwords longer but still easy to remember is the use of passphrases, i.e., sentences that can be remembered. Such passphrases are easier to remember than non-dictionary words and are harder to break. For instance, it is 10 times more secure to use "this is fun" as a password than "J4fS<2” [13]

Character Set If a greater variety of characters are used in a password, then the number of possible combinations is greater and therefore the password space is larger. For instance, it takes 1 month to crack “jskerv” using brute-force attack while it takes 219 years to crack “J4fS<2” [13].

Randomness If a password is known to be in a language, then an attacker can leverage regularities in the language to reduce the password space.

The time periods given in the above examples are based upon testing 100 passwords per second. However, today this rate would be quite low as an offline attack could search the space in parallel and hence even 64 bits of password space can be searched in a short time.

When used to authenticate a user, the system has to check if the entered password is valid. Of course, storing a file with a list of usernames and password is a bad idea because if the confidentiality of the file was ever compromised then all security would be lost. Instead, a cryptographic hash of the password is used and stored. When a user enters a password, the system computes the hash of the password and compares it with the entry in the password file. Despite hashes being stored instead of passwords, the file's integrity still has to be protected. The problem with a file that is not confidential is that it could be stolen and used in an offline dictionary attack. An alternative to confidentiality to defend against such threats is the use of a salt. Password salting introduces randomness in the password hashing; hence the same password will be hashed differently. For each password entry, a salt is generated as a random value S. The password P is hashed together with the salt using the cryptographic hash function H as H(S+P). Then both S and H(S+P) are stored in the file. For verification of the salted password, the hash of the password and the salt is computed and the result matched against the password file. Without the use of salt, an attacker can pre-compute hashes of all of the words in a dictionary once for all password entries, while with salt an attacker has to compute hashes of all dictionary words for each password entry. For instance, with a 12-bit random salt, the same password could hash to 212 different hash values.

It is worth noting that an attacker who wishes to get someone's password does not have to perform any of the previously described attacks himself. As described in the previous chapter, lots of companies have been hacked and their password databases have been leaked. It is possible and very easy to check if an email address and password has been revealed during one of these attacks simply by doing a search on the Internet [14]. If the e-mail the attacker is searching for was leaked during one of the attacks, then it should be possible to find the database where this password was stored. Using tools on the Internet that decrypt hashed passwords means that the attacker may now have the target account's password.

2.1.2 Ownership factors

The ownership factor requires the user to be in possession of something. This factor is usually a security token. Typically, a token is a hardware device that has been issued to the user for the purpose of authentication. The fundamental requirements for a token are that it can be queried by the authentication system and that the token is uniquely associated with the user. The token is often authenticated using some form of cryptographic process. There are different forms of tokens:

- Paper tokens, for instance a list of "one-time passwords" that the user needs to enter in response to a challenge,
- hardware tokens, which are physical tokens that the user has to be in possession of, and

- soft tokens, which rely on a software component that is present on the user's device, such as a software token application.

A hardware token could be a small and simple LCD display, a token that connects to a USB port, or a smart card. These tokens can be divided into three categories

Disconnected tokens have no direct connection to the user's computer[*]

Connected tokens need to be physically connected to the computer

Contactless tokens utilize a wireless connection to the computer rather than a physical connection

A summary of the different ownership factors is presented in Figure 2-1. Software tokens are meant to mitigate some of the vulnerabilities that hardware tokens have. Loss and theft are the biggest risks with a token that has to be carried around by the user. Stolen tokens can be rendered useless by requiring a PIN be entered along with the information provided by the token. Procuring and replacing tokens has a cost and trade-offs between cost and usability have been demonstrated [15]. For instance, flaws in widely-used ATMs have been found [16]. Software tokens do not require expensive, single-use, hardware technology. Another advantage of software tokens are that no additional hardware is necessary since many people carry their mobile phones all the time. Lastly, software tokens can be sent electronically to geographically dispersed regions (avoiding waiting for shipping and paperwork).

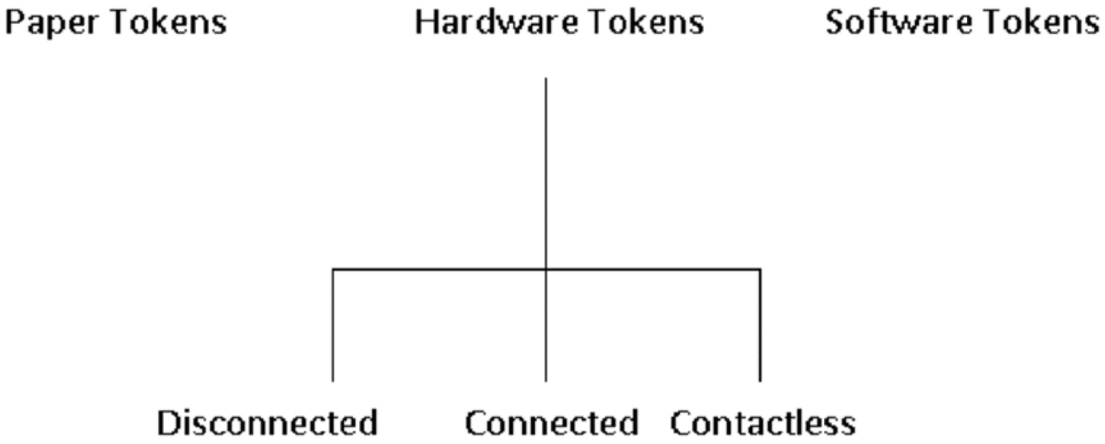


Figure 2-1: Summary of ownership factors

[*] These tokens generate information that the user manually uses in the authentication process. A common form of such tokens is device that displays on an LCD display the result of a cryptographic hash of an internal timestamp. This result is manually entered by the user.

Authentication through ownership factors rely on one-time passwords (OTPs) or a Public key infrastructure (PKI). OTPs uses a cryptographically generated PIN on each use and this PIN is only valid for that session. This process utilizes symmetric authentication, where an OTP is generated both on the server that performs authentication and on the token (software or hardware) that the user possesses. If the OTP generated by the token matches the one generated on the server, then the authentication is successful. These OTPs may be generated based on time-synchronization generation of OTPs by both the authentication server and the user, using a mathematical algorithm to generate new OTPs based on previous passwords, or using a mathematical algorithm where the new OTP is based on a challenge. Moreover, randomly generated numbers sent to the end user via a short message service (SMS) are in practice similar to OTP. Tokens that rely on time-synchronized OTPs may be realized by proprietary devices, mobile phones, or other devices running software that may be proprietary, freeware, or open-source. A time-based one-time password (TOTP) algorithm is an algorithm that generates time-synchronized OTPs. Authentication that uses PKI performs asymmetric authentication as it relies on a pair of keys: a private encryption key and a public encryption key. The private encryption key is not transferable when it is stored within a hardware token. When authenticating to a server, a challenge is issued by the server and the response is signed using the private encryption key. Authentication is successful and access is granted if the signed challenge validated using the public encryption key known by the server matches the plain-text of the challenge. The choice of using OTPs or PKI depends upon the requirements and needs of the organization that wants to exploit ownership factors for authentication. While OTP authentication may provide sufficient security for most enterprises, companies, and institutions that operate in fields that require greater security (such as e-health or e-banking) might use PKI authentication.

As described above, OTPs are sometimes sent via SMS, but this method is no longer considered highly reliable for services that need a high security such as e-banking and e-learning. Firstly, mobile network operators (MNOs) are unable to guarantee SMS text message delivery within an acceptable timeframe, because their mobile networks may be overloaded, e.g. during natural disasters or network coverage may be poor or non-existent, for instance underground. Moreover, researchers at Northeastern University and Technische Universität Berlin have come to the conclusion that OTPs via SMS are no longer a secure method of authentication for two reasons [17]:

- The Global system for mobile communication (GSM) technology is insecure due to several vulnerabilities such as lack of mutual authentication and weak encryption algorithms.
- Mobile phone malware, and especially Trojans, are designated to intercept SMS messages (and therefore OTPs) are a rising threat to material sent in an SMS message.

2.1.3 Inherence Factors

Authentication based on inherence factors relies on behavioral or physiological characteristics of the user. These characteristics have to be measured accurately. Inheritance factors might include retinal pattern, fingerprints, handprint, voice pattern, or keystroke timing. In order to implement a biometric authentication scheme, the representations of the characteristic of interest are stored. During the authentication process, the characteristic is measured and compared with the stored value. An exact match is not expectable due to error rates associated with biometrics sensors, which is at most 5% [18]. The percentage does not seem that high, but in a larger scale the number of errors would be significant. The assessment of biometric accuracy is usually expressed as a percentage, and there are four main measures of biometric accuracy:

- **True Acceptance Rate(TAR) / True Match Rate (TMR)** a measure that tells how well the system is able to correctly match the information from the same user.
- **False Acceptance Rate (FAR) / False Match Rate (FMR)** a measure of the likelihood that the system will falsely report biometric information of one person to match the information of another person. When this happens, an unauthorized user is granted access to an account.
- **True Rejection Rate (TRR) / True Non-Match Rate (TNMR)** a measure of the likelihood that the system will not match the biometric information against any of records in the database because, in fact, that person is not in the database.
- **False Rejection Rate (FRR) / False Non-Match Rate (FNMR)** a measure that represents the frequency of cases when biometric information from a user is not matched against any record in the database when it should have been because the person is in the database.

These above four measures are interdependently [19]. There is a mathematical relationship between the corresponding true and false rates so that if one rate is known the other one can be calculated. For instance, when working with percentages, the sum of the rates has to be 100%, so if one rate is 98% the other is 2%. Also, there is a trade-off in that a decrease in the frequency of false matches tends to also decrease the frequency of the true ones. Other measures of a biometric system's accuracy are:

- **Equal Error Rate (EER or ERR)** which indicates the point at which the False Acceptance Rate is equal to the False Rejection Rate.
- **Failure to Enroll Rate (FER)** the rate at which people are unable to enroll in a biometric system. This can happen when biometric characteristics are weak or missing.

- **Failure to Acquire Rate (FTA)** the rate at which biometric information could not be obtained even though the person was able to previously enroll. These fails can be caused by environmental conditions at the time of biometric system use.

Even if inherence factors may appear to be highly secure factors, there are drawbacks and problems related to biometric information. Firstly, physical characteristics are unstable and could vary in time. For instance, a sore or wound on a finger or on the hand could compromise the authentication process. Fingerprints have also been copied. In 2014, the German defense minister von der Leyen's fingerprint was copied by Chaos Computer Group [20]. Cost and availability are two important factors in the use of inherence factors for security. Moreover, people may be unwilling or unable to provide their biometric characteristics. A report from 2006 by the U. S. National Institute of Standards and Technology (NIST) shows that people find biometric systems less hygienic and more stressful than traditional PIN systems [21]. Moreover, their main drawback is the difficulty of revocation. Unlike knowledge factors, an inherence factor is not easily revocable.

2.2 Types of Authentication and Levels of Assurance

The types of authentication available differ in terms of the security they provide; hence combining factors from one or more of the three categories described above is commonly done to increase the level of security. Thus there are different levels of assurance (LoA) for entity authentication that describe the *degree of confidence* in the process of authentication.

2.2.1 Types of Authentication

There are three types of authentication methods. They differ in the level of security they provide and in the way they use the authentication factors. The following three types of authentication methods will be considered in this thesis:

Single Factor authentication

Single factor authentication is the weakest level of authentication, where only one of the previously described factors is used. Usually, single factor authentication relies on the use of a secret (i.e., a password). This authentication method has been described in the Section 2.1.1, with a focus on its weakness.

Two-factor authentication

When a 2FA method is used, two factors are required to be provided by the user in order for the authentication to be successful. Many 2FA methods rely on a password as one of the two factors [22].

Multi-factor authentication

In a multi-factor authentication process two or more factors are used during the authentication process

It is important to note that the term strong *authentication* has no unique meaning and it is defined differently by different institutions. The U. S. National Information Assurance by the Committee on National Security Systems (CNSS) defines strong authentication as *"The requirement to use multiple factors for authentication and advanced technology, such as dynamic passwords or digital certificates, to verify an entity's identity"* [23]. The European Central Bank (ECB) has defined strong authentication as *"a procedure based on two or more of the three authentication factors"* [24]. Additionally, according to the U.S. FFIEC[*], elements of at least two, and preferably all three types of authentication factors should be verified during an authentication process [25].

2.2.2 Levels of Assurance

The International Telecommunication Union's Telecommunication Standardization Sector (ITU-T) Entity Authentication Assurance Framework (EAAF) defines four LoAs for authentication [25]. **Error! Reference source not found.** summarizes the different levels of assurance that EAAF defines. LoA1 is the lowest level of assurance and LoA4 is the highest level of assurance. Determining which LoA is appropriate in a given situation depends on a variety of factors. The determination of the appropriate LoA is mainly based on risks: the consequences of an authentication error, the resultant harm, and their likelihood of occurrence. When the risks are higher, a higher LoA should be used.

Error! Reference source not found. **Levels of Assurance and corresponding risk**

Risk	Level of Assurance
Minimum	LoA1
Moderate	LoA2
Substantial	LoA3
Very High	LoA4

2.2.2.1 Level of assurance 1 (LoA1)

LoA1 is used when the risk associated with erroneous authentication is minimal. There is no a specific requirement for which of the authentication mechanisms to use, but it has to provide some minimal assurance. This level does not require the use of cryptographic methods.

[*] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, National Credit Union Administration, Office of the Comptroller of the Currency, and Office of Thrift Supervision

For example, LoA1 may be applicable for transactions involving websites that require registration in order to access materials or documentation.

2.2.2.2 Level of assurance 2 (LoA2)

LoA2 is used when the risk associated with an erroneous authentication is moderate. At this level, SFA is acceptable. A user that wants to authenticate has to prove control over the credential, through a secure authentication protocol. There should be controls that reduce the effectiveness of online guessing attacks and protect against attacks on stored credentials.

For example, an invoice company may operate a website which allows customers to change their address. The process by which a customer changes an address may be considered a LoA2 authentication process because it involves a moderate risk of inconvenience since notes regarding payment and account details are sent to the address. The company should therefore obtain some authentication assurance before the process of changing address is allowed.

2.2.2.3 Level of assurance 3 (LoA3)

At this level, there is a high confidence in the claimed identity of the entity. This level is used when the risk associated with an erroneous authentication is substantial. This LoA should provide multi-factor authentication. At this level, any secret information that is exchanged during the authentication process should be cryptography protected. Identity information should be verified and identity proofing procedures have to be dependent on this verification.

LoA3 transaction examples include, for example, transfer of money or funds from an online banking account.

Since, according to [25], single-factor authentication is acceptable for LoA2 and LoA3 should employ multi-factor authentication, any 2FA system provides a third level of assurance. Hence, the 2FA solution for DigiExam will provide a LoA3.

2.2.2.4 Level of assurance 4(LoA4)

At this level, there is a very high confidence in the claimed identity of the entity. This level of assurance is used when the risk associated with erroneous authentication is very high. LoA4 is the highest level of assurance defined by the recommendation. LoA4 is similar to LoA3 but it adds the use of tamper-resistant hardware devices for the storage of all secret keys. The sensitive data included in the protocols should be cryptographically protected.

For example, a pharmacist dispensation of a controlled medication may require a LoA4 protection.

2.3 Digi Exam's SFA implementation

Digi Exam has not (yet) implemented a 2FA method. Authentication of users relies only on SFA, i.e. passwords. Both students and teachers are only required to provide their username and their password to access their account.

Figure 2-2 shows the login form for DigiExam as seen by teachers and students wishing to log in to their account. As with the majority of such forms, there is a field for the username and a field for the password. Below the Log in button, there are two links, one for creating an account and another for a forgotten password. It is interesting to note that the form is implemented in such way that in the case of an erroneous login attempt, the same message (Login failed, e-mail or password is incorrect) is displayed independently of the error (whether it was an incorrect username or password). This is good because a possible attacker is unable to know if the username entered is valid, before trying to guess the password. However, it is not difficult to get a teacher's private or work mail address, thus the security of the username is weak. This means that the only real security comes from the security of the password.

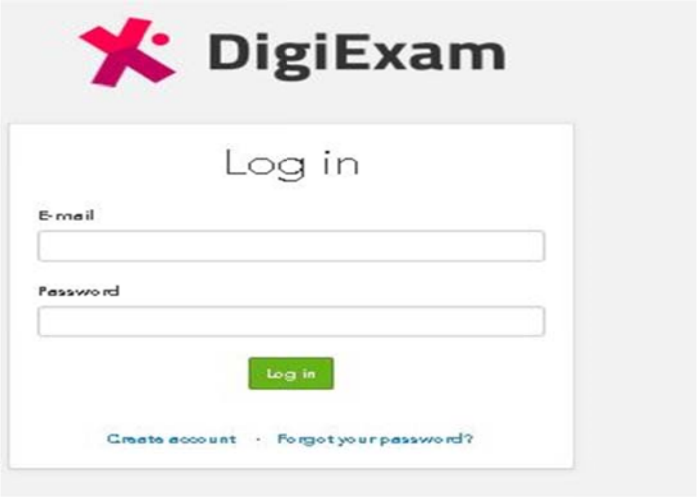


Figure 2-2: DigiExam Log In form

The back-end technology is implemented using Google's own programming language, Go. Go is a free and open source language created in 2007. It is a compiled and statically typed language, with garbage collection and memory safety features.

When the login is successful, a secure cookie is passed with the name of the person that logged in and other information that includes authorization information. The login function is throttled; hence there is a limitation on how many times the function can be called over time. DigiExam allows a user to attempt to log in 10 times over a period of 3 minutes. All login attempts are logged and monitored. The information that is logged at a login attempt are an e-mail address, IP address, and if the login attempt was successful or not.

The passwords are hashed using bcrypt, a hashing function first published in 1999. The function incorporates a salt to protect against rainbow tables, and the iteration count can be increased to make it slower[*]. Bcrypt is based on Blowfish cipher and it uses a key setup algorithm which becomes quite time expensive.

2.4 Related work

This section presents previous work related to this thesis project. Google implemented 2FA as early as September 2010 [26] and lots of studies have been done regarding 2FA and how to make authentication more secure. Also, over the last several years there has been a substantial increase in the use of eLearning- systems and digital assessment platforms due to the advantages that they bring.

The majority of the work in this section addresses the problem of securing eLearning systems or learning management systems, which are a broader concept than digital assessment platforms. However, the focus of this earlier has mainly been on how to best verify online student’s identity, rather than strengthening the account security of teachers.

2.4.1 Securing e-learning against impersonation

In 2016, Shauna Beaudin at the Nova Southeastern University [27] conducted an empirical study of different authentication methods to secure e-learning systems against impersonation fraud. The researcher refers to a previous work by Apampa, Wills, and Argles [28] to identify authentication control methods based upon their strength for e-learning activities that have a high potential for impersonation. The researcher performed quantitative research and the results of this research showed that e-learners perceive that the levels of authentication must vary in strength based upon the activity being considered. According to this researcher, summative e-assessments need a stronger authentication method than SFA, which should at least include authentication based on biometrics or a live proctor. The result of this study is not applicable to this project thesis due to the different focus of her research. Despite this, the work was useful as it identified previous work regarding multi-factor authentication.

2.4.2 Migration from SFA to 2FA

The problem of painless migration from SFA to 2FA has been addressed particularly well by Mao, Florencio, and Herley [29]. In their paper, they describe how they solved the problem by proposing an incremental change from SFA to 2FA. Their solution does not require any alteration of the existing system and it

[*] This iteration count is the number of rounds in which the keying algorithm is applied. This process can be made slow which helps avoid brute-force attacks.

makes it possible to have both the legacy password authentication and the 2FA systems coexist.

2.4.3 Preventing cheating during online exams

In order to prevent cheating during online exams, Bedford, Gregg, and Clinton [30] implemented a device called Remote Proctor that verifies a student’s identity through the use of biometric comparison. The authors assert that due to its functionality and low cost, the device will be more cost effective in the long run than the alternatives. Also, they show how faculty and students feel the device will help stop cheating. Despite addressing the problem of students cheating, they do not provide any higher security for the teachers’ account, which is still protected through SFA.

2.4.4 Usability of 2FA

A comparative usability study of 2FA by De Cristofaro et al. [31] was done to measure the usability of three popular 2FA solutions, namely codes generated by tokens, one-time PINs received via email or SMS, and dedicated smartphone apps. They performed a quantitative study on 219 users that showed that 2FA technologies are perceived as highly usable and that a user’s perception of 2FA is correlated with individual characteristics of gender, age, and background. The study provides a starting point for follow-up qualitative studies and is applicable to this thesis as UX is one of the aspects analyzed in this thesis to select the best 2FA option.

2.4.5 Authentication of students in online learning environments

Jortberg and Bailie [32] addresses the problem of authentication of students in online learning environments by comparing different authentication methods with their privacy impact. This work shows how a university and a corporation have partnered to find how to best verify the identity of online students and studies the use of biometrics, web video recording, and face-to-face proctored exams.

2.4.6 E-learning security issues

Adetoba et al. [33] have written a review of e-learning security issues and challenges. Authentication is listed as one of the five major security issues in this field. The authors describe current research in e-learning security, and as in previously discussed work, lots of efforts have been made regarding the use of biometric authentication. In 2011, Alotaibi and Argles [34] proposed a biometric system which requires human interaction and a fingerprint scan to utilize a service. Moreover, Song et al. [35] proposed performing authentication by combining both eye tracking and event relevant potential (ERP) of brain waves.

2.4.7 Using a graphical password

Vaithyasubramanian et al. [36] implemented a two-factor authentication method for secure login using a graphical password in addition to a password. The advantages are that the solution is easy to implement and that a graphical password is also easy to remember. However, the proposed solution is not a 2FA method because both a password and a graphical password are actually something that a user remembers (a knowledge factor); therefore, the work cannot be used in this thesis project.

2.4.8 Economics of authentication systems

Authentication systems are often seen from a technical point of view. Altinkemer and Wang [37] are the authors of the first paper that attempts to understand the decision of authentication systems from an economic point of view. The authors state that every authentication system can be seen as non-repairable or biometric. Non-repairable means that as time passes, there is a greater chance of the system failing. For instance, a password is non-repairable because it could be lost or stolen. Their study shows the expected costs and losses of different authentication methods and that managers who wish to implement 2FA need to take into account the implementation costs, the market share the company has, and the composition of customers. The paper is interesting but does not propose any technical solution. Because of cost is not one of the more important criteria for this thesis project, this work is only partially relevant.

A study carried out by ENCAP security in 2012 [38] compares the costs of authentication methods for enterprise for secure employees' access to the enterprise applications. The company analyzed the average cost of the six most prevalent 2FA approaches for an enterprise with 3,000 users over a three years period. The study reveals that a smart device-based software solution is 95 percent cheaper than a hardware OTP solution. According to Thomas Bostrøm Jørgensenhen, CEO of ENCAP, the time for hardware-based authentication has passed. The study is interesting and can be used as a reference in this thesis, though it is performed by a company that provides smart-device software solutions and therefore there is a risk that the study is not objective. Moreover, the study analyses the cost for deploying 2FA in order for the employees to securely access the enterprise's services and not the cost of deploying the solution to end users.

3 Methodologies and Methods

The purpose of this chapter is to provide information about the research methodologies and methods used in this thesis. These were briefly explained in Section 1.5, and this section is going to focus on research strategies, data collection, and quality assurance.

Section 3.1 describes the research process underlying the thesis project. The steps that have been followed in writing this thesis are presented in a list as well with a detailed explanation of them. Section 3.2 presents details about the methodologies while section 3.4 focuses on the methods adopted in order to select the most suitable 2FA method. Section 3.4 explains how data was searched and collected and presents some ethical aspects linked to it. Section 3.5 explains the techniques used to evaluate the reliability and validity of the data collected while section **Error! Reference source not found.** and section 3.7 describe the method adopted for the experimental design.

3.1 Research Process

Research involves a process which focus is on objectively gathering information that will be analyzed in order to come to a conclusion. This thesis project's research process includes multiple steps that are interlinked with each other. These steps are:

- selection of the topic area,
- formulation of the problem,
- formulation of the goals of the thesis,
- review of the literature,
- selection of methods of data collection,
- analysis of data, and
- reaching conclusions.

Selection of a topic area was the first step. This thesis project focuses on the weak security offered by passwords and on how to make authentication more secure. The topic area was selected by the authors of this thesis together with their supervisor at the company because it was of mutual interest. The authors have previously taken courses on internet security. Moreover, the topic of this thesis project was found to be interesting because it addresses an underrated but serious problem. The company finds the topic of special interest because of a new regulation (Regulation(EU) 2016/679) in the protection of personal data that will be applied from May 25th, 2018 [39]. The goal of this new set of rules is to simplify the regulatory environment for business and give citizens back control over their personal data.

Identifying a problem related to the topic area is the goal of the second step. The problem is the weak security of SFA (i.e., passwords) in the field of digital assessment platforms. After the problem has been identified, the aims and objectives of the thesis had to be stated in order to determine the scope, the depth, and the direction of the work. In order to avoid setting unrealistic goals and aims, the research objectives were formulated according to specific, measurable, achievable, realistic and timely (SMART) objectives [39]. Therefore, the goal of the project was set to identify and implement a suitable 2FA method for DigiExam for securing teachers' accounts. It is a goal that was regarded as achievable and realistic at the beginning of the project, and it is also considered to be timely. We believe the thesis project to be important for several different reasons. Firstly, it contributes to the elimination of a gap in the literature. Although 2FA is not a new topic and lots of solutions exists, many of them focus on using biometric as an additional authentication factor. Also, as described in Section 2.4 most of this prior work focused on securing students' accounts rather than teachers' accounts. Moreover, the practical goal of the thesis is to solve a tangible problem, i.e. the lack of sufficient security in the authentication system implemented by DigiExam.

Once the problem and the scope are identified, existing literature on the topic has to be searched and reviewed. This step allows the researcher to find previous work regarding the problem and data related to it. The first stage in the literature search was to identify search terms. The main search terms in this project were *2FA, authentication, data breaches, two-factor authentication with one time passwords, different 2FA methods, authentication factors, types of authentication, send OTP, UX with* 2FA, *cost of 2FA, cost of biometrics, cost of hardware tokens* and other. In the second stage, online literature was found using these search terms. Relevant literature was identified by visiting Questia[*], Google Scholar[†], and JURN[‡]. Data and statistic about data breaches, password security, and related information were collected from reports by companies (such as Juniper and Symantec). The collected data was then filtered according to the importance of the authors and the source. Secondly, the remaining literature was filtered according to the contribution of that text to this project thesis. Lastly, the literature was filtered by date of publication in order to focus on the latest developments in the area.

3.2 Methodologies

The different types of research methods can be divided into several categories depending on the nature, the scope, and the purpose of the research. The general classification of the different methods is quantitative and qualitative. The

[*] An online research and paper writing resource. https://www.questia.com/

[†] Bibliographic Database by Google. https://scholar.google.com/

[‡] Online search tool for searching academic articles and tools. http://www.jurn.org/#gsc.tab=0

quantitative research method is based on numbers and calculations while the qualitative one is based on perceptions and feelings and aims to provide depth of understanding. This thesis will mainly follow the quantitative approach due to the topic area that is analyzed. Moreover, a research method could be descriptive or analytical. A descriptive method involves surveys and studies which goal is to identify facts and describe a situation as it is at present. An analytical research is different because it is based on already available information that is analyzed in order to provide an evaluation. This thesis project follows the analytical method as it relies on existing data and information and not on surveys carried out by the researcher. Lastly, the purpose of the study will affect the research methods. This thesis project is an applied research because it primarily aims at finding a solution for a problem. Findings of an applied research could be applied to an issue while a fundamental research will simply aim at exploring issues. Also, applied and fundamental research differs in the context, as applied research objectives are often set by a company or a client in order to try to solve a problem they are facing. Fundamental research is on the other hand usually driven by the researchers primarily to expand their knowledge in that area.

This thesis project will be grounded on a deductive method since it relies on quantitative data, on an objective reasoning and it will try to answer a pre-specified question, therefore being outcome-oriented. Moreover, it is important to state how we will try to solve the problem. Due to the fact that an attempt to solve the problem will be done in this thesis and hopefully a final answer to the problem will be delivered, this project will adapt to a conclusive research design.

3.3 Methods

In section 3.1 the methods used to find information about different 2FA methods were described. This section will instead provide information about the methods used to decide which of those 2FA methods is the most suitable for DigiExam. The aspects we analyzed in order to carry out the result were cost, end UX, security, and deployment. These aspects were decided by the authors and the supervisor at the company as interesting and important ones. These aspects were analyzed both for possession factors and inherence factors. The possession tokens that were analyzed were hardware and software ones, with the exclusion of paper tokens. The main reason is that it is not in line with the idea of the company; the company’s idea is to make exams digital so using paper to perform authentication is contrasting. Furthermore, it was difficult to gather information about UX when using paper tokens. Finally, it is difficult to estimate the cost of such solution.

A suitable authentication method was found using an elimination method. First, we compared second authentication factors, namely inherence or ownership factor. Depending on which method were best regarding the chosen aspects, one of them was selected. At this point, the chosen method was analyzed and its difference solutions compared to find the most suitable one.

3.4 Data Collection

Data collection is the process by which the researcher collects information from all the relevant sources chosen in the previous steps of the research process. The collected data can be divided into two categories: secondary data collection and primary data collection. Primary data are collected by the researcher conducting the research through methods like surveys, interviews or questionnaires. Often, when the research problem is unique, there is no available related work, then the researcher has to collect the data him or herself. Unfortunately, collecting primary data can be costly and time-consuming.

In contrast, secondary data is data that have been already collected and available from other sources. The main sources of secondary data are reports, government censuses, magazines, journals, conferences, and the Internet. One of the main advantages of utilizing secondary data compared to primary data is that it is less costly. Also, it is more readily available and can provide a basis for comparison with data collected by the researcher. Furthermore, secondary data might be available when primary data cannot be obtained at all. Moreover, the authenticity and validity of secondary data should be questioned, especially when it comes from the Internet.

This thesis project is mainly literature-based, therefore the data used in this work is secondary data collected following the steps explained in Section 3.1 and those that are explained in Section **Error! Reference source not found.**.

3.4.1 Ethical considerations

Ethical issues are related to both primary data collection and secondary data collection. New technologies have made these issues more pressing, and data sharing and storage have become faster and easier. Moreover, there are concerns regarding security and confidentiality of the data used. The main issues with secondary data are related to potential harm to individuals as well as the issue of consent. Data could contain identifying information about participants, which means that the researcher re-using this data should describe how the privacy and confidentiality of the participant will be protected. If the data does not contain any identifying information, then the researcher should confirm that the data is anonymous and acknowledge the source of the data. Secondary data was not used to *answer* the problem presented in this thesis but rather was collected during the first two weeks of the project thesis in order to use it as background knowledge. This contained data and statistics about data breaches, passwords, and internet security.

3.5 Assessing reliability and validity of the data collected

Because of the disadvantages explained in the previous section, secondary data has to be evaluated. The data collected as background for this thesis project followed the requirements of availability, relevance, accuracy, and sufficiency. Lots of articles and papers have been written about the field of security and authentication, hence data was available. The data collected was also relevant as it was not outdated and it perfectly fits the problem of this thesis. Moreover, the data was assessed as being accurate. The data was detailed, exact, and comprehensive. The articles chosen presented a throughout discussion of the subject and all sides of the main issue were addressed.

3.6 Experimental design

This section will provide information about the decisions made when planning on implementing the prototype.

It is important to notice that the implementation of the prototype may differ in some details from the *theoretical* best solution as the authors may not have the skills, the time or the abilities to implement a fully functional prototype of that solution. Moreover, the solution may not be fully functional for the same reasons. This aspect will be discussed in a more detailed manner in Chapter 5.

3.7 Software used

With regard to the design and implementation of the prototype, the software used has to be decided. The way in which the software will be decided reflects the abilities and skills the authors have acquired during their studies. The main line that will be followed is to use well-known programs/software. Chapter 5 will provide information about the software used when implementing the solution.

4 Selection of a suitable 2FA method

In this chapter, we are going to select the most suitable 2FA method for DigiExam.

4.1 Ownership or inherence factor?

The first step was to select either the possession or the inherence factor as the second authentication method. We believe that the disadvantages of a biometric implementation for DigiExam outweigh the advantages.

First of all, DigiExam must deal with the regulations that regulate how personal and biometric information are stored. According to the Swedish Privacy Protection Law [40], personal information that is collected should be adequate and relevant in relation to the purpose, and no more information that necessary should be collected. This means that biometric profiles should technically be built in a way that eliminates needless information. Moreover, all the end users would need to consent to having their biometric characteristics stored by DigiExam. According to the EU regulations, the biometric data could not be centrally stored, but rather only stored as a hash. The consent should be voluntary which means that there has to be an alternative to a biometric based authentication. Another important factor to take into account is that if the solution is going to be extended to students' accounts too, then DigiExam has to deal with storing biometric characteristics of minors (who do not have the ability to give such consent, hence it would have to involve their legal guardian(s)).

With this being said, the main challenge for DigiExam would be the process by which the biometric characteristic of a teacher is captured in order to be mapped to an identity. Lack of accuracy in capturing the characteristic or partial capture can lead to system failure. The best way would be to have the teacher scan their biometric characteristic under the supervision of a system administrator who could then easily associate the information with the respective identity in a database. The second alternative would be to buy and send a biometric scanner to each teacher or oblige them to buy one and then have them scan their characteristic and then send the results to DigiExam. This is difficult because there has to be an entity that guarantees that it is the teacher that is scanning his/her own biometric information. None of the solutions make the enrollment phase easy. Moreover, we believe that none of these alternatives is effective in terms of cost, organization, and implementation, considering that today there are between 13,000 and 15,000 teachers' accounts at DigiExam (including both active and inactive accounts). Implementing a 2FA method is costly and time- consuming. The equipment has to be purchased (the cost of a fingerprint scanner is circa US$20 when a large order is made [41]), the software has to be installed (which probably means license costs), and the information must be stored securely.

An alternative to lower costs of implementation could be to use iris or facial recognition method, using the camera that is nowadays implemented in most laptops. Even assuming that all teachers are in possession of a laptop with a built-in camera, the challenges explained before regarding storing information and users who are unwilling to have their characteristics stored persist.

Additionally, the fact that physical characteristics are unstable and could vary with time could make authentication impossible. As explained in Section 2.1.3, a sore or wound on a finger could for instance compromise authentication process when using fingerprints authentication. Finally, it is difficult to deal with revocation as inherence characteristics are hardly revocable.

This being said, we believe that DigiExam should implement 2FA using ownership factor as the second mean of authentication. The deployment process of rolling out a biometric authentication for teachers is too challenging; moreover, as explained before, DigiExam could be interested in implementing 2FA for students' accounts too, which would make the challenge even more demanding.

4.2 Hardware or soft ownership token?

In the previous section, we came to the conclusion that DigiExam should rely on ownership factors when implementing 2FA for their teacher accounts. In this section, we are going to select the most suitable ownership factor, namely a hardware-based or a software-based token. To summarize, a hardware token could be a small and simple LCD display that generates and displays an OTP, a token that connects to a USB port, or a smart card. On the other hand, a soft token is a software-based token that simplifies distribution and lowers costs compared to a hardware token.

4.2.1 Deployment

Software-based authentication mechanisms are easier to deploy on a large scale compared to hardware ones. They are more advantageous regarding both time and effort for deployment. The main reason is that software tokens do not have to be provided physically to the end users, thus eliminating shipping time and cost. Hardware tokens cannot be distributed immediately and hence require logistical planning. Another advantage of using a software-based solution is that users would realize very quickly if they lost their mobile device and therefore report the loss quicker than they would have done in the case of a hardware token, thus further limiting the risk of an unauthorized user being able to log in.

Moreover, even if it was easy to set up and start using a hardware token, DigiExam has to support that hardware token. This means that if DigiExam makes an agreement with a hardware token supplier, then this type of token has to be integrated into DigiExam's service, a process that could be painful or require time.

Another aspect that makes a soft solution better than a hardware-based one is that the 2FA solution has to be rolled out to the end customer, rather than just to the company's employees. Today there are between 13,000 and 15,000 teachers' accounts and nearly 200,000 students' account, which means that the solution has to be deployed on a very large scale. A hardware-based solution would be more suitable where the goal to secure employees' login into a company's services or networks, as the number of users is likely to be smaller than the number of users of the company's services. Moreover, companies with very large numbers of users (such as Google, Facebook, LinkedIn, and Yahoo) implements 2FA using a software-based token[*].

Furthermore, hardware tokens are not eco-friendly as they would lead to more electronic waste, hence adopting them would not be environmentally responsible.

4.2.2 Cost

Software-based alternatives eliminate the need of a hardware token, thereby providing savings both initially and over time. Relying on a software-based solution, an enterprise saves shipping costs, both for initial deployment and over time as new hardware may be needed to replace stolen or broken hardware. Generally, software solutions also present advantages from a licensing standpoint. If the company does not develop the software itself, it only has to pay for the software license, while hardware solutions may involve both a software license and a fixed subscription cost. Also, hardware token manufacturer may force enterprises to replace tokens after a fixed amount of time which could lead to additional costs. This means that the capital expenditure (CAPEX) cost of a hardware-based solution would be substantially higher than for a software-based solution. Moreover, in the case of SMS OTP the operating costs are spread over time and in case of TOTP the operating cost is nearly zero as the software is license free.

Additionally, avoiding hardware tokens mean no extra devices or "wallet-fillers" for the customers, as a software token is just add-on to the device the user already is in possession of.

4.2.3 Security

The security of hardware and soft tokens depends on how they are implemented and on the resources available to a potential attacker who attempts to defeat the selected 2FA method.

When it comes to security, hardware tokens might be considered more secure than software ones. The main problem with soft tokens is that they are completely dependent upon the device they reside on, which could be a computer, a mobile phone, or a tablet. A virus infecting the device or an attacker gaining physical access to the device can completely compromise the token. A software PKI solution

[*] Googles uses Google Authenticator, Facebook uses Login Approval, while Yahoo and LinkedIn use a solution based on the same idea: provide a personal code sent to a mobile phone when trying to log in from a new device.

is vulnerable to being hacked during the key generation process and to keylogger attacks. A software OTP solution is also vulnerable to attacks. The OTP application's data could be extracted, which means that an attacker could access the victim's OTP; moreover, if the OTP is sent via SMS, the weakness of the GSM network could be exploited to intercept the OTP, as was described in Section 2.1.2. Also, malware applications and Trojans could be used to intercept the SMS messages.

4.2.4 User Experience (UX)

To start, it has to be said that increased security comes with a loss of convenience and ease of use. Therefore, designing a user-friendly 2FA is a hard task due to the fact that authentication poses many contradictions to the principles of Human-Computer Interaction (HCI), as secure authentication often compromises the user's experience[31, 42]. Moreover, the study in [31], concluded that the usability of different 2FA methods varied depending on the users' demographic background rather than the actual technology or the context the 2FA method was utilized in.

An inconvenience with software tokens is that phones can unexpectedly run out of battery, while the lifetime of a hardware token's battery is expected to be many years: Citi Private Bank (CITI) states that the battery life of their hardware tokens is at least 7 years [43], Duo's hardware tokens' battery life is also expected to be 7 years [44], while the hardware token provided by Stanford University has a battery lifetime of four years [45]. This means that when using a hardware token a user does not have to worry very much about having enough battery power for access. On the other hand, a hardware token is inconvenient to carry and can be lost. Software solutions come in handy when implemented or relying on a mobile phone.

4.2.5 Summary of hardware-based and soft-based solution

Deployment, cost, security and UX have been examined for hardware-based and soft-based solutions. summarizes the different aspects of a 2FA solution that would most suit DigiExam's needs.

We believe that a soft-based solution is the most suitable for DigiExam, as it is easier to deploy than a hardware solution, it is less costly, and it should offer a better UX. The security of a soft-based solution is lower than that of a hardware solution, but a soft-based solution will provide enough security for most of DigiExam's customers.

Table 4-1: Comparison of Hardware-based and soft-based solutions

Examined aspects	Hardware-based solution	Soft-based solution
Deployment	Not optimal for DigiExam	Optimal for DigiExam
Cost	$$$	$$
Security	More secure	Less secure
UX	*	**

4.2.6 YubiKey, an out-of-the-box hardware solution

As stated in the previous sections, we believe that a soft-based 2FA solution is the most suitable solution for DigiExam. However, there is an interesting hardware-based solution that is worth analyzing. The Swedish company Yubico was founded in 2007 and their hardware 2FA solution named YubiKey aims to change the historical trade-off between security, low cost, and high usability [*]. The YubiKey is a touch-sensitive USB and NFC security key offering multiple functions for protecting users' access to services and does not require any software installation or battery. A YubiKey offers many different options, such as OTP, OATH TOTP and HOTP, and open GPG.

The most common pattern is to use Yubico OTP in combination with a username and password to strengthen the security of users' logins. The main advantages offered by the OTP solution provided by Yubico are that there is no need for client software further easing its implementation. In order to verify the OTPs, there are two options:

- Use Yubico's own web server called YubiCloud, which is available after getting an free API key or
- host a verification server. Implementations of such a server written in Go, Python, and PHP are available on Yubico's website[†].

Regarding implementation, Yubico offers two solutions: integration plugins (if supported software is used) or libraries for programming languages such as Go, Python, and PHP, which help when connecting to the YubiCloud for Yubico OTP.

It is worth saying that today Yubikey is used in education, as 1,000 schools including 450 higher education institutions rely on Yubico for securing schools' applications or computers.

[*] All the advantages of Yubikey can be found at https://www.yubico.com/products/why-yubikey-wins/

[†] Different validation servers implementations are available in PHP, Python and Go https://developers.yubico.com/Software_Projects/Yubico_OTP/YubiCloud_Validation_Servers/

The solution offered by Yubico is good, but we believe that some issues would limit DigiExam's ability to use it. The main challenge is how DigiExam could deploy the solution to all teachers. Buying a YubiKey and having it delivered to all teachers is definitely too costly for DigiExam. An option could be to support Yubico's hardware and leave it up to the user to buy a YubiKey in order add an extra layer of security to their account. This raises the question of which percentage of the users would be willing to buy a YubiKey and implement 2FA. An investigation or survey should be done by DigiExam in order to understand what percentage of the users are actually willing to adopt YubiKey and whether supporting YubiKey would pay off or not for both the users and DigiExam. Moreover, the fact that 2FA may not be adopted by all user raises a possible ethical and moral question, namely who is liable to ensure that an account is properly secure. Is it DigiExam's responsibility or the end user's responsibility? Is it enough to simply make it easier for the user to use 2FA or is it actually necessary to force all users to use 2FA when they log in?

4.3 OTP or PKI

In the previous sections, we came to the conclusion that DigiExam should rely on a soft-based 2FA solution in order to properly secure all teachers' accounts. As seen in Section 2.1.2, there are different software-based solutions, most prominently OTPs and PKI. In this section, the best software-based solution is going to be selected.

While a PKI solution is generally considered more secure than an OTP-based one, for most enterprises an OTP-based solution provides sufficient security. However, a PKI-based solution should be used when a higher level of assurance is needed, for example in e-health and e-banking services. Regarding cost, an OTP-based solution is generally more economic as well as easier and quicker to deploy as a PKI-based solution requires setting up a PKI infrastructure, which could be a complex and costly process[*]. It is feasible to buy one certificate from a Certificate Authority and then use it to sign certificates for each user, but there are general challenges for a safe and reliable implementation of a PKI. These challenges include the human and technical administration challenge of running and maintaining a reliable Registration Authority (RA) and ensuring that PKI-systems are usable, intuitive and trusted by the end users. It is worth noting that there are open source solutions for PKI Certificate Authority software, such as EJBCA[†]. EJBCA is scalable, flexible and suitable to build a complete PKI infrastructure for a large organization or enterprise. Moreover, an enterprise could run a PKI infrastructure itself, which eliminates the need of buying certificates from a certificate authority. For example, a school could sign certificates for each teacher.

[*] If relying on not free third party services

[†] https://www.ejbca.org/

However, we believe that for DigiExam an OTP solution would be suitable. Moreover, we make the assumption that for an enterprise such as DigiExam the added security offered by a PKI solution is not vital, as a PKI-based solution is more suitable for e-health and e-banking services. Therefore, we will focus on an OTP-based solution for this thesis.

4.3.1 OTP, HOTP, and TOTP

An OTP value can be generated through two standards governed by the Initiative for Open Authentication (OATH): HOTP or TOTP. In this section details about these OTP schemes are given and the two algorithms are presented.

4.3.1.1 OTP

RFC 2289 [46] describes a One-Time Password System. OTPs were mainly to counter a "replay attack", in which information is eavesdropped and captured on a network connection in order to be used later to access a system. The system described in the RFC relies on a secret pass-phrase that is used to generate a series of OTPs. The strength of the system is that the user's pass-phrase never has to traverse the network, therefore it is invulnerable to replay attacks; moreover, added security is given by the property that no secret information has to be stored on any system. However, OTP offers no protection against social engineering or active attacks.

There are mainly two requirements for an OTP-system to work. First, the generator must generate the OTP from the user's secret pass-phrase and from the information provided by the server. The server, on the other hand, must send a challenge to the generator, must verify the received OTP, must store the last valid OTP it received, and must store the OTP sequence number. The OTP generator passes the pass-phrase, along with a seed received from the server through a secure hash function to produce a one-time password. Details of the system are present in RFC 2289.

4.3.1.2 HOTP

RFC 4226[47] describes an algorithm to generate one-time passwords based on the Hashed Message Authentication Code (HMAC). The algorithm relies on two factors: a shared secret key and a moving factor. The symbols introduced by the RFC are:

- an 8-byte counter value C that is the moving factor. This counter has to be synchronized between the client (generator) and the server (validator),
- a shared secret K between client and server. Each generator has a unique and different K,
- t is a throttling parameter that indicates the number of unsuccessful authentication attempts before the server refuses connection, and

- s is a resynchronization parameter.

HOTP is based on using the HMAC-SHA-1 algorithm described in RFC 2104[48]. Since the output of the algorithm is 160 bits, the value is truncated in order to easily be read by the user. The function *Truncate* converts the output into an HOTP value:

$$\text{HOTP}(K,C) = \text{Truncate}(\text{HMAC-SHA-1}(K,C))$$

The RFC describes in detail how the HOTP is generated. The implementation must extract at least a 6-digit code and possibly a 7 or 8-digit code, depending on the security requirements.

With regard to security RFC 4226 states that even if an adversary is able to observe numerous message exchanges of successful authentication attempts and knows how build a function in order to generate HOTP values, then the adversary will **not** have a significant advantage over a random guess.

A protocol implementing HOTP as an authentication method between a prover and a verifier has to meet the following requirements. It:

- must support two-factor authentication,
- should **not** be vulnerable to brute force attacks, and
- should be implemented over a secure channel in order to protect users' privacy.

Regarding validation of the HOTP value, the client increments its counter and then calculates the next HOTP value. If the value that is received by the authentication server matches the value calculated by the client, then the HOTP is validated and the server increments its value by one. If there is not a match, the server initiates the resynch protocol before it requests another pass. If the maximum number of authorized attempts is reached, the server should lock the account and inform the user.

4.3.1.3 TOTP

RFC 6238[49] presents an extension of the HOTP algorithm, where the moving factor is based on a time value. This time-based OTP algorithm provides short-lived OTPs which strengthen security. The basic different between the TOTP algorithm and the HOTP algorithm is that a value T derived from a time reference replaces the counter C. Moreover, TOTP may use HMAC-SHA-256 or HMAC-SHA-512 functions instead of the HMAC-SHA-1 function used in the HOTP implementation. The requirements of a TOTP algorithm implementation are:

- The prover and the verifier must know the current UNIX time[*],
- The prover and the verifier must share the same secret,

[*] The number of seconds elapsed since midnight UTC of January 1, 1970

- The algorithm must use HOTP as a key building block,
- The prover and the verifier must use the same time-step value X,
- There must be a unique secret key for each prover,
- The keys should be randomly generated, and
- The keys must be protected against unauthorized access.

The value X represents the time step in seconds, and its default value is 30 seconds. The number of time steps is computed relative to Unix time, with a default value of 0.

RFC 6238 defines TOTP as TOTP = HOTP (K, T). The implementation of the algorithm has to support a time value T larger than a 32-bit integer when beyond the year 2038.

Regarding validation of the TOTP, the potential delay between the client and the server has to be taken into account. If a TOTP value is generated at the end of a time-step window, then the receiving time will probably fall into the next time-step value, hence two different TOTPs are generated. Therefore, a validation system should not only compare the OTPs with the receiving timestamps but also with the past timestamps taking into consideration the one-way transmission delay. RFC 6238 recommends that at least one-time step is allowed for network delay. A time-step of 30 seconds is recommended by the RFC for an optimal balance between security and usability. Moreover, the next OTP must be generated in the next time-step window, hence a user has to wait until the clock moves to the next time-step before they will get a different OTP. This means that a too-large window will be unsuitable for a typical login authentication system. Details about validation and resynchronization are described in RFC 6238.

4.3.2 HOTP or TOTP

The main difference between HOTP and TOTP is that the OTPs generated through the TOTP algorithm are short-lived (generally with a lifetime of less than 30 seconds), while the OTP generated through the HOTP algorithm is potentially longer-lived. This means that the security provided by a TOTP implementation is better than an HOTP implementation for two reasons. First, if an HOTP password is compromised, it can be valid for a “long time”, while a TOTP password is only valid for a number of seconds (if implemented following the recommendations). Also, the attempt to attack a key in the TOTP protocol may be invalidated because the target keeps moving. Therefore, a TOTP implementation is preferred.

4.4 TOTP application or SMS delivery

As described in the previous section, a TOTP implementation of the OTP is preferable due to its enhanced security. If the OTP scheme is based on a software solution, then the OTP is generated by an application on the phone. Another way of using OTPs it to send a (randomly generated) code (often called a *nonce*) to the end user via SMS. In this section, we analyze which method is best in terms of cost, deployment, security, and maintenance.

The main advantages of receiving a OTP via SMS is that the user does not have to own a smartphone, since all mobile phones are able to receive SMSs. Although nearly 80% of the people in Sweden owned a smartphone in 2015 [50], implementing a TOTP application solution means that not everyone is able to use it. Moreover, SMSs are convenient since there is no need to download an application or perform any setup. In terms of deployment, an SMS-based solution is simpler than an application-based solution since it does not require the company to write an application to generate TOTPs or to modify its servers in order to meet the requirements of the TOTP algorithm. Also, the client device generating the TOTP does not need to be connected to the Internet (as the protocol also works in offline mode).

On the other hand, there are some disadvantages in using SMS delivery of OTPs. The first disadvantage is due to limited network coverage as poor network coverage would imply the impossibility for the user to authenticate using 2FA. Moreover, as seen in Section 2.1.2 sending OTPs though SMS is no longer considered secure due to the vulnerabilities of the GSM network and due to malware & Trojans designated to intercept SMSs. An SIM exchange will not compromise the TOTP codes since the codes are independent of the SIM card. Moreover, an application generating TOTP is not vulnerable to man-in-the middle or replay attacks. This means that from a security point of view, a TOTP-based implementation is more secure than a scheme that does not consider time.

The main disadvantage of a TOTP based approach is that an application-based solution may be less user-friendly - as the user has to perform an initial setup and an app has to be started every time the user needs to authenticate himself/herself using 2FA.

In terms of cost, the TOTP solution is more convenient since it is nearly free (once it has been implemented). The TOTP protocols are open source, which means that the only cost will be based upon the amount of time necessary for DigiExam to implement the protocol in its servers and to develop an app that generates TOTPs. Implementing an SMS 2FA will have an on-going cost to DigiExam since the company will have to utilize a communication platform, such as Twilio[*], GatewayApi[†], or SmsApi[‡] in order to have the OTPs sent. This means that by using a nearly free solution, such as a TOTP application, the 2FA solution could be offered to the end users at every login, while relying on an SMS–based solution means that some challenges have to be faced in order to make the solution

[*] https://www.twilio.com/

[†] https://gatewayapi.com/

[‡] https://www.smsapi.com

affordable. Considering that the average cost for sending an SMS using existing services is roughly $0.040, which means that it would be infeasible for 10,000 teachers to use 2FA every time they login to DigiExam's webpage. This can be seen in detail by assuming three logins per day, which would cost DigiExam circa $40,000 per month for every teacher to login using 2FA. In order to make such a solution more economical, a 2FA solution relying on SMS has to be offered to the end user *only* on particular occasions; for example, when trying to log in from a new unrecognized device or after clearance of the cache. This approach of only performing the full process occasionally would significantly lower the costs for DigiExam but still strengthen the account security of the users. However, at the same time it would make it possible to perform some forms of attack during the time when entries in the cache are valid and/or between times when the user logins via a new device.

A summary of the two different solutions regarded as most suitable for DigiExam: (1) an application running on the end user's phone generating an OTP and (2) an OTP randomly generated and sent via SMS to the end user. Table 4-2 illustrates these two solutions in terms of cost, deployment, UX, and security.

Table 4-2: Summary of OTP application and SMS OTP delivery

Examined aspects	OTP application	SMS OTP delivery
Cost	Nearly free	Costly over time
Deployment	May be more problematic to deploy	Easy deployable
UX	Good	Very good
Security	Highly secure	Average security

Both of these software-based solutions are good for an enterprise, such as DigiExam. While the SMS solution may offer a better UX and is more easily deployed than an OTP solution, the latter offers greater security and is less costly. We believe that DigiExam should utilize an OTP-based solution for 2FA as the difference in cost and security are greater than the difference between the deployment cost and UX between the two solutions. Compromising the user's experience a little bit and allowing for a nearly free deployment that provides a highly secure solution provides greater advantages for DigiExam.

4.5 Account Recovery

This section describes and analyzes several different potential account recovery methods for the two solutions presented above.

Every authentication system has to deal with account recovery, which is the process of regaining control after losing access to an account. The recovery process differs depending on how the authentication system is implemented and could be more challenging if 2FA is used. The more secure the authentication process is the harder it is to regain control over the account, as more information has to be proven. This means that while 2FA strengthens the security of the account at login, it also makes it more difficult and troublesome for the account possessor to regain control over an account. This difficulty in account recovery should be kept in mind, as if not implemented thoughtfully it could enable an intruder to steal an account. An unavailable mobile device implies the inability of the user to use the second authentication factor. Depending on which 2FA method is used and whether the device is broken or lost there are different ways of dealing with account recovery. Table 4-3 summarizes the different ways in which account recovery could be handled if 2FA is enabled.

Table 4-3: Summary of how account recovery could be handled for different 2FA implementations

Cases	SMS 2FA	OTP application 2FA
Forgotten password	E-mail with reset link (SMS if number stored)	E-mail with reset link (SMS if number stored)
Broken device	SMS with code	Backup code or *peer recovery*. (SMS if number stored).
Lost device	Backup code or *peer recovery*. (SMS if number stored)	Backup code or *peer recovery*. (SMS if number stored)
Inability to provide any 2FA method.	Account recovery handled manually	Account recovery handled manually

4.5.1 Forgotten password without 2FA enabled

A forgotten password is one of the most common reasons for being unable to access an account. The most common method used in order to recover the account in this scenario is to send a recovery link or temporary password to the user’s email address, which was registered during account creation. Sometimes the password or the link is sent to an alternative mail address, still provided by the user during account creation. This method is the easiest and probably the most logical one to deal with forgotten passwords, but it actually might increase the chances for a possible intruder to access the account, since the intruder can gain access through the email.

4.5.2 Forgotten password with 2FA enabled

A forgotten password in case of a 2FA enabled account presents more challenges when recovering the account. For security reasons, accessing the account based upon account recovery when 2FA is enabled should **not** be easier than accessing it the normal way (2FA login). If enabled, 2FA should not be bypassed during account recovery.

Regardless of the 2FA method used, the account recovery mechanism should work in the same way as described in the previous section; hence 2FA would not be bypassed. Even if a potential intruder manages to reset the password, the second authentication factor would still be required in order to gain access and therefore the account is still secured with 2FA.

If the user's phone number was stored by the company implementing 2FA, then another solution for dealing with a forgotten password would be to send a code to the user's SMS number - thus giving the user the possibility to reset their password.

4.5.3 Inability to provide the second factor of authentication

The inability to provide the second factor of authentication could be due to either loss of the user's mobile device or it being broken. By the term broken we mean a mobile device that is in possession of the user but is unusable. A lost mobile device means that the user has no physical access to the device.

Storing the user's phone number during the registration process makes it easier to deal with account recovery in case of a broken mobile device, regardless of the 2FA method chosen for login. This is true since the SIM card is still in possession of the user, hence the user could move this SIM card to another devices and an SMS could be sent with either a code to login or a restore code to reset the password.

On the other hand, the process of recovering the account becomes complicated if 2FA relies on OTP and no phone number is linked to the account. This case is handled in the same way for both a lost and broken mobile device. The problem that arises, in this case, is that authentication is dependent upon the application.

One solution that still would provide 2FA security is to have a backup code in form of a paper token that the user has to store safely. This token is utilized during the account recovery process in order to gain access to the account. Another solution is to rely on *peer account recovery*. In this solution, two users have the opportunity to identify each other as trusted peers, which gives them both the possibility to reset each other's account through their personal page. For instance, if user A has no access to his (or her) mobile device, then user B could log into his (or her) personal account and disable 2FA for user's B account. This solution could actually be a suitable solution for DigiExam as their customers are teachers who work with each other. This could make it easier for a user to choose one person to be *peer*, in contrast to a service rolled out to users that might not have a relation to each other.

4.5.4 Inability of providing any 2FA factor

If the user is unable to provide the right password and neither the paper token solution or e trusted peer solution are available, we believe that in order to guarantee the real user’s authenticity account recovery should be handled manually by an administrator. The more information stored during registration or user activity on the page, the more information that can be proven by the user to the administrator. For instance, in case of a betting site – a common requirement would be the last digits of a credit card or a list of recent placed bets.

In the case of DigiExam, information that could prove someone’s identity includes time and date for assessments and correction. However, this information could be also known or guessed by students, which means that they are not truly reliable for proving a teacher’s identity. Therefore, we propose that a user should provide a telephone number of a trusted person during registration. In this way, the customer support at DigiExam could contact the person and verify that the user is unable to login.

5 Implementation of the prototype

This chapter presents details of the implemented prototype of the solution. Section 5.1 presents the idea underlying the implementation and describes different use cases, thus giving a theoretical view of the solution. It also describes the design choices that were made and why they were made. Section 5.2 provides more detailed information about the implementation, describing how the solution was implemented and which tools were used. Section 5.3 describes the limitations that were determined during the implementation of the prototype of the solution.

5.1 Design of the implementation

This section presents the idea behind the implementation and how the solution was designed. It also includes an activity diagram that explains how the system implemented is meant to be done. We implemented the solution as a web page. The web page was hosted on a localhost running with Apache[*] and using Uniserver[†].

5.1.1 Registration of the user

We first dealt with registration of the user. DigiExam stores the full name of the user, the mail address and the gender of the user. For the purpose of the implementation, we chose to not take into account the user's full name or gender. In order to implement the SMS 2FA option, our implementation requires the user to provide a valid phone number during registration. The email address, password, and phone number are then stored in a database. The password is hashed with the *sha256* hashing algorithm before being inserted into the database in order to better simulate the functionality of the system. However, it is not salted. The reason is that the focus of this implementation is on the second authentication factor and not on the security of the first one. Making the first authentication factor secure is out of the project scope. A random 16-digit base32 encoded code is generated for every user during registration and stored in the database. This 16 digit code is the secret key the user's device and the server has to know in order to calculate the OTP for 2FA login through an application. The code is presented to the user as a QR-code. The following use case describes how a user registers an account:

1. Basic flow:

   This flow starts when an actor wishes to register an account:

   a. The system requires that the user provides an email address, a password, and a phone number,

[*] Apache is a free web server developed by Apache Software Foundation. https://www.apache.org/

[†] The Uniform Server is a lightweight WAMP solution for running a web server under the Windows OS. http://www.uniformserver.com/

b. The user enters the information needed, and

c. The system validates the entered email address, password, and phone number and registers the user into the database.

2. Alternative flow

Invalid mail address/password/phone number
If in the Basic Flow the user provides an invalid e-mail address, password, or phone number, an error message is displayed and the user can decide whether to try again or go back to the initial page. If the information provided by the user is correct, then the user is registered and a new page is displayed and this page contains a link to the login page.

5.1.2 Login of the user

After defining the registration process for a user, we designed the login system for our implementation. Figure 5-1 shows a use case diagram that describes how a user logs into the system. At login, the user has to provide their e-mail address and the password provided during registration. The first login after creating an account is an SFA login where a correct combination of username and password are sufficient to gain access to the account. We chose to make the 2FA option user enabled because we believe that DigiExam's responsibility is to *provide* the user with a 2FA *option*, but it is up to the user to choose whether to use this options or not. Therefore, users have the ability to enable or disable 2FA for their account.

When 2FA is disabled, then our login implementation starts in the same way as DigiExam's current implementation, i.e. the entered username and password are checked in the database and if there is a match, then the user is granted access. After login, a welcome page is displayed where information about 2FA is displayed as well as the QR-code for OTP login. Via this page the user can choose to continue with SFA (the default option) or enable 2FA login. For the sake of simplicity, the user cannot do anything else on the welcome page other than log out. Our system handles failed login attempts in the same was as DigiExam does, i.e. the user can fail to login ten times during a 3 minutes long period before the account is blocked for 3 minutes. If 2FA is enabled, ´then the user has to provide the correct combination of username and password and is then redirected to a page via which the code received by SMS or the code generated on their mobile device can be entered and if validated access is granted.

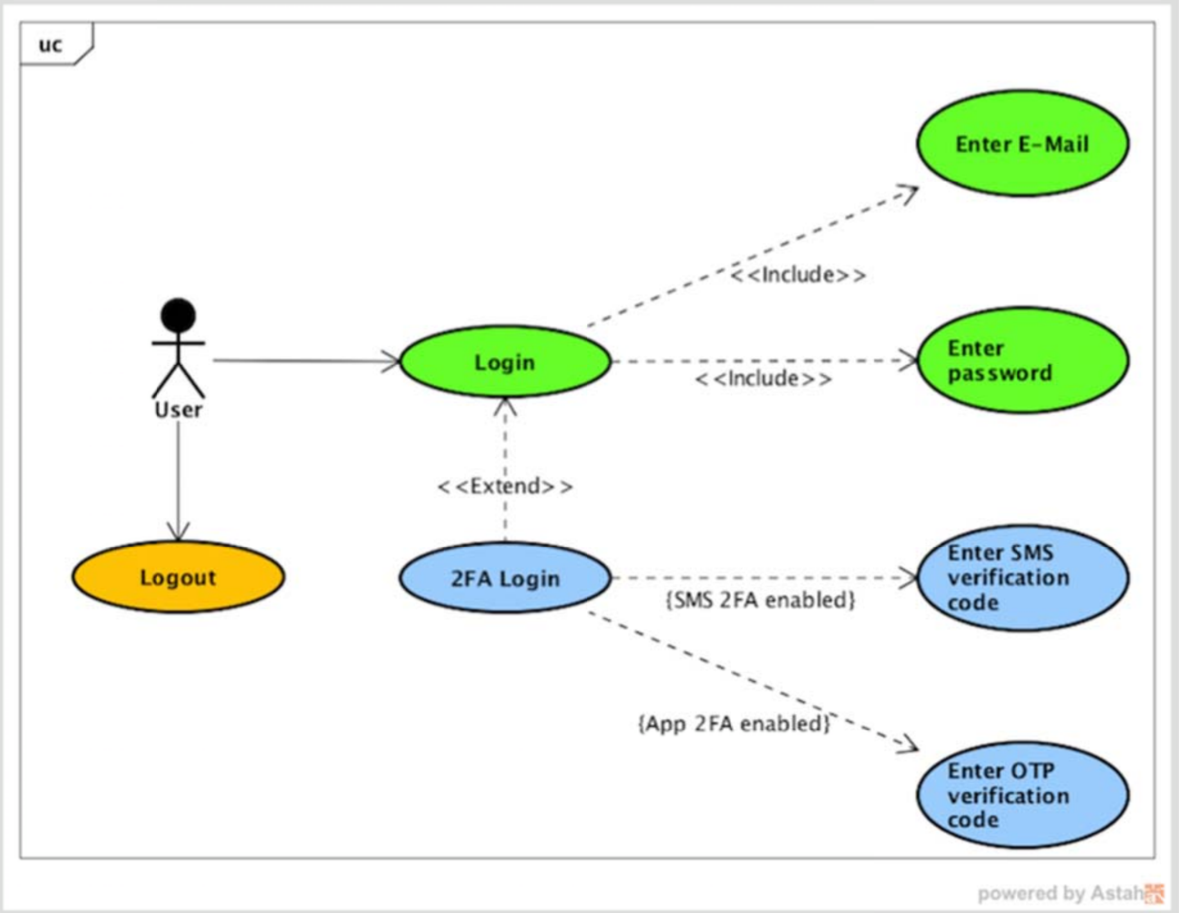


Figure 5-1: Use case diagram showing how the user logs into the system

5.1.2.1 SMS 2FA

If the user enables 2FA through SMS, a randomly generated 6-digit code will be sent to their phone number after the correct combination of username and password has been entered. We chose to use a 6-digit code since many existing services use a 6-digit code and because this length of code offers good security s(as the chance of an intruder guessing the correct code is low). If the user enters an wrong code twice in a 40 second period, then before they are redirected to the initial login page they are blocked for 3 minutes. If the entered code is valid, then the user will be redirected to the welcome page. A user could request a new code to be sent if the first one was not received; however, for security reasons only the last sent code is valid. When deciding upon the time between two incorrect attempts and the number of failed attempts, we came to the conclusion that two attempts in 40 seconds was reasonable as this limits the time for an intruder to guess the code and still gives time for the user to enter an incorrect code once. The chances of guessing a 6-digit randomly generated code within two attempts is *1 / 500,000*, which corresponds to 0.002‰.

Further note that for our implementation, we decided to rely on Twilio to send SMS codes to the user who is logging in. The reason behind this decision is that

Twilio is a well-established company who offers a reliable service and competitive prices for their services. Moreover, Twilio is used by companies, such as WhatsApp, Airbnb, eBay, and Uber.

5.1.2.2 OTP 2FA

If the user enables 2FA through OTP, the process is not as straightforward as for the SMS implementation. The user has to download an application that can generate TOTPs from a secret key. At the user's welcome page, a QR-code representing the secret key is presented for the user to scan with the application in order to start generating TOTPs. After providing the right combination of username and password, the user is redirected to a page in which the generated TOTP has to be inserted. If the inserted TOTP is the same as the TOTP generated by the validating server, then the user is granted access.

As with the SMS authentication, an incorrect code can only be inserted twice in a period of 40 seconds before the account will be blocked. The OTP will be regenerated after 30 seconds, but it is valid for two minutes in order to deal with bad synchronization between client and server clocks. This makes the choice of 40 seconds for the interval acceptable as even if the client and server are out of cycle by one cycle one of the two entries should be correct.

5.1.2.2.1 Choice of TOTP generating application

Since the Google Authenticator is currently proprietary software, we chose to utilize another application. FreeOTP* is a free 2FA application for systems utilizing OTP protocols. FreeOTP offers the possibility to add tokens by scanning a QR-code and it implements both the HOTP and TOTP standards which means that no proprietary server-side component is necessary. FreeOTP is available for both Android and iOS. However, the solution implemented in this thesis project also works using Google Authenticator.

5.1.3 Esthetical design of the implementation

Regarding the aesthetical design of the implemented solution, we decided to use the same design as DigiExam's actual login and registration page. The reason behind this choice was that it gives an idea of how DigiExam's login pages would look if the proposed solution was actually implemented by the company.

5.2 Software implementation

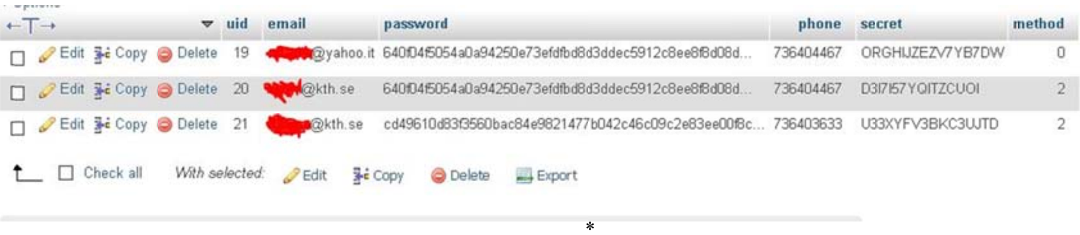
This section describes how the design was implemented and which tools were used during the development of this implementation. Some important parts of the code are shown as well as the database scheme.

* https://freeotp.github.io/

The first choice to me made was which programming language to use for implementing the server side of the solution, the client side, and the database. The web page was written in HTML and CSS — as these are the standard languages for creating web pages. JavaScript would have streamlined some processes but the authors preferred to use languages the selected as they had more experience with them and felt more comfortable with them. The server side was implemented using PHP for the same reasons. The database used was MySQL as it was already known and had already been used by the authors.

5.2.1 Registration process

The registration process was handled using PHP and MySQL. Figure 5-2 shows the table used to handle both the registration and login processes. As described earlier, the user inserts an email address, a password, and a phone number into the form presented on the registration page shown in Figure 5-3. When the user submits the HTML form a PHP page will process the data and according to the validity of it proceed with the registration of the user. The column *secret* contains the 16-digit base32 encoded code that is the secret key used by the TOTP algorithm to generate TOTPs. The base32 alphabet is described in RFC 4648[51], and it uses an alphabet of A-Z followed by 2-7. The digits "0" and the "1" are not used since they are similar to the letters "o" and "I". In the implementation, the secret key was generated using the PHP function *mt_rand*, as shown in Figure 5-4. The column *method* contains a single integer and provides information about whether 2FA is enabled or not and indicates which of the 2FA solution is enabled. This column has a default value of 0, meaning that the user only has to provide a username and password at the first login after registration.



←T→				uid	email	password	phone	secret	method
□	Edit	Copy	Delete	19	[redacted]@yahoo.it	640f04f5054a0a94250e73efdfbd8d3ddec5912c8ee8f8d08d...	736404467	ORGHIJZEZV7YB7DW	0
□	Edit	Copy	Delete	20	[redacted]@kth.se	640f04f5054a0a94250e73efdfbd8d3ddec5912c8ee8f8d08d...	736404467	D3I7I57YQITZCUOI	2
□	Edit	Copy	Delete	21	[redacted]@kth.se	cd49610d83f3560bac84e9821477b042c46c09c2e83ee00f8c...	736403633	U33XYFV3BKC3UJTD	2

*

Figure 5-2: Table used by the registration and the login processes. The value in *secret* is 16-digit 32based encoded.

* As seen in the figure, the password entry is not salted as explained in section 5.1.1

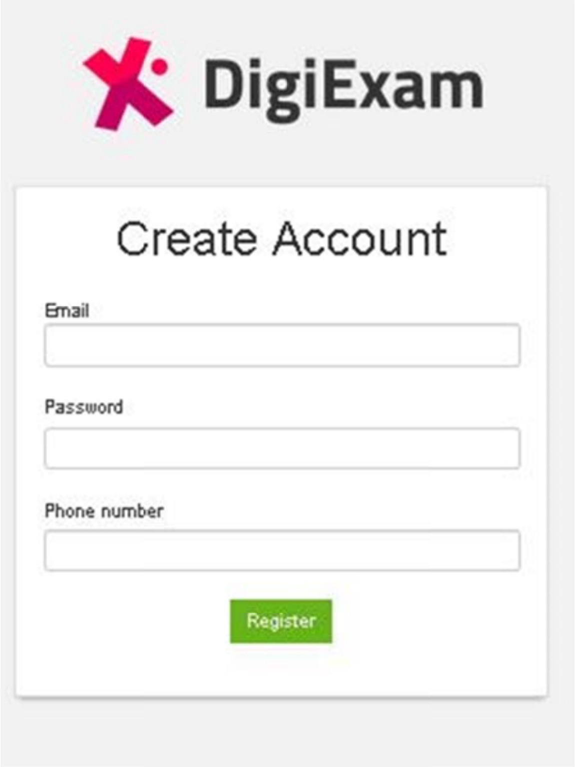
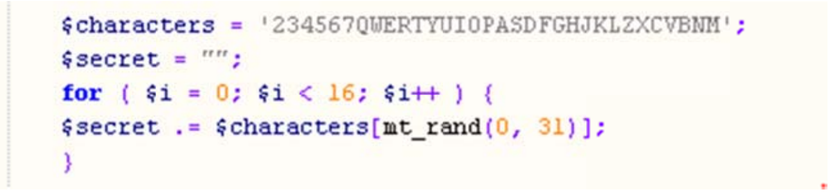


Figure 5-3: Registration page where the user inserts a valid e-mail, a password, and a phone number.



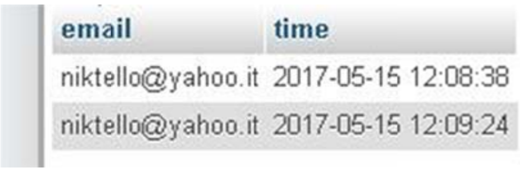
```
$characters = '234567QWERTYUIOPASDFGHJKLZXCVBNM';
$secret = "";
for ( $i = 0; $i < 16; $i++ ) {
$secret .= $characters[mt_rand(0, 31)];
}
```

Figure 5-4: PHP code snippet showing the creation of the secret key used by the application and the server to generate TOTPs.

5.2.2 Login process

The table used for the login process is the same as used for the registration process. When the user enters their e-mail address and the password, a PHP page handles the data and depending on whether 2FA is enabled or not the user is redirected to the welcome page or to a new page where the code sent via SMS or the OTP code has to be entered. A new session is created to keep track of the user proceeding with login.

In the case of a failed login attempt, the information about the user trying to log in is stored in a new table shown in Figure 5-5. The PHP function *Now()* is used to return a timestamp recording at which time the login attempt took place.



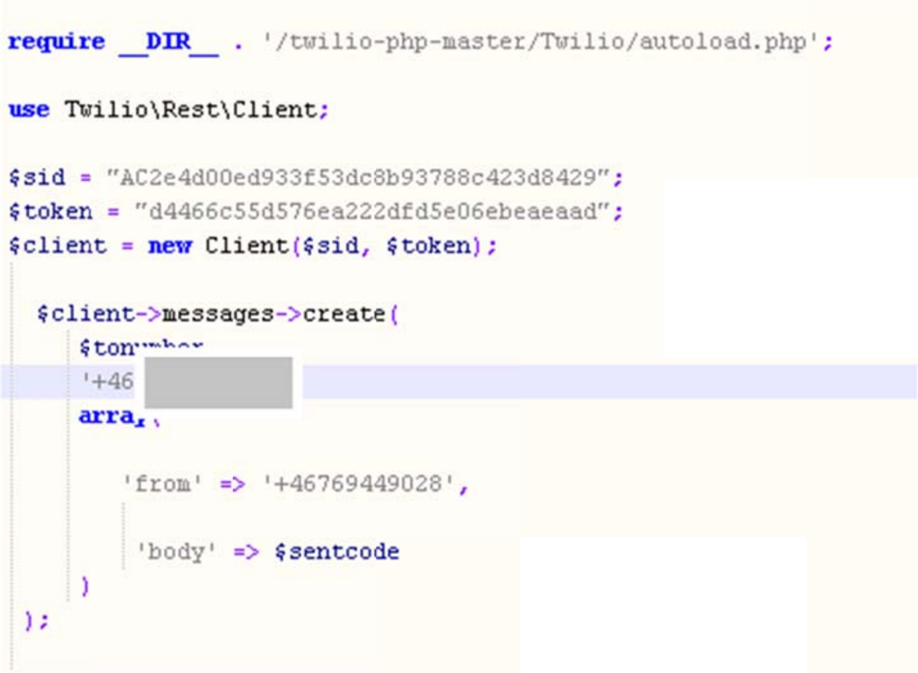
email	time
niktello@yahoo.it	2017-05-15 12:08:38
niktello@yahoo.it	2017-05-15 12:09:24

Figure 5-5: The table in which failed login attempts are stored

If ten attempts are made in a period of three minutes, then the user is blocked from trying to log in again for three minutes. A table similar to the previous table is used to stored failed 2FA login attempts, in order to deal with the limited number of maximum number of attempts a user has to provide the correct SMS code or TOTP within the bounded number of attempts.

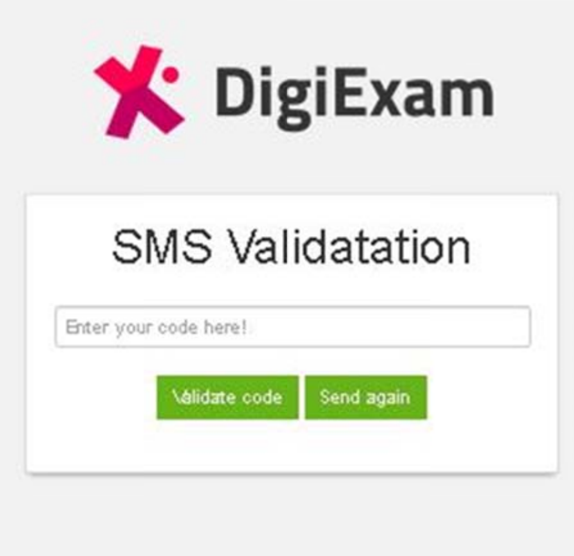
5.2.2.1 SMS authentication

If the user proceeding with login has SMS 2FA enabled, three additional PHP files are used to handle the login. First, a 6-digits random code is sent using the same function as was shown in Figure 5-4, then the code is sent using Twilio's Rest API as shown in Figure 5-6. The user is then redirected to a page that displays a form into which the SMS code has to be entered as shown in Figure 5-7. When the user enters a code and clicks the validate button, a PHP validation file will check whether the code is correct and if so, then the user is granted access. In this case, a new session will be created for the logged in user. After two failed attempts to provide the correct code, the user will be blocked and the session destroyed.



```
require __DIR__ . '/twilio-php-master/Twilio/autoload.php';

use Twilio\Rest\Client;

$sid = "AC2e4d00ed933f53dc8b93788c423d8429";
$token = "d4466c55d576ea222dfd5e06ebeaead";
$client = new Client($sid, $token);

  $client->messages->create(
    $ton
    '+46 [redacted]
    arra

        'from' => '+46769449028',

         'body' => $sentcode
    )
 );
```

Figure 5-6: PHP code that sends an SMS using Twilio



DigiExam

SMS Validatation

Enter your code here!

Figure 5-7: The page showed to the user when having SMS 2FA enabled and after having provided the right combination of username and password

5.2.2.1.1 How to use Twilio

In order to use Twilio to send SMSs, the first step is to sign up at their website*. After signing up, it is possible to buy a Twilio phone number which will be used for sending SMSs. We implemented Twilio's service *without* using Composer, therefore we manually downloaded and used the PHP SDK. The PHP SDK is available from GitHub†. In order to use the SDK, the file has to be unzipped in the same directory as where the code it is going to be used and then the autoload has to be required. The code in Figure 5-6 shows how we implemented this.

5.2.2.2 OTP application

To log in using TOTP, the same code has to be generated on the server and on the user's device. This is made possible by a secret key that is shared between the user and the server. As described before, our implementation relies on the FreeOTP application available for free both for iOS and Android. To set up the application and start generating TOTPs, the user has to start the application and scan the QR-code containing the secret key. Figure 5-8 shows what the application looks like after scanning the QR-code. Since the secret key is now known by the server and the application and both the server, now the client is able to derive the current Unix time and generate the appropriate TOTP. Is the client and server are synchronized, then the response will be authenticated as valid.

* https://www.twilio.com/

† https://github.com/twilio/twilio-php/archive/master.zip

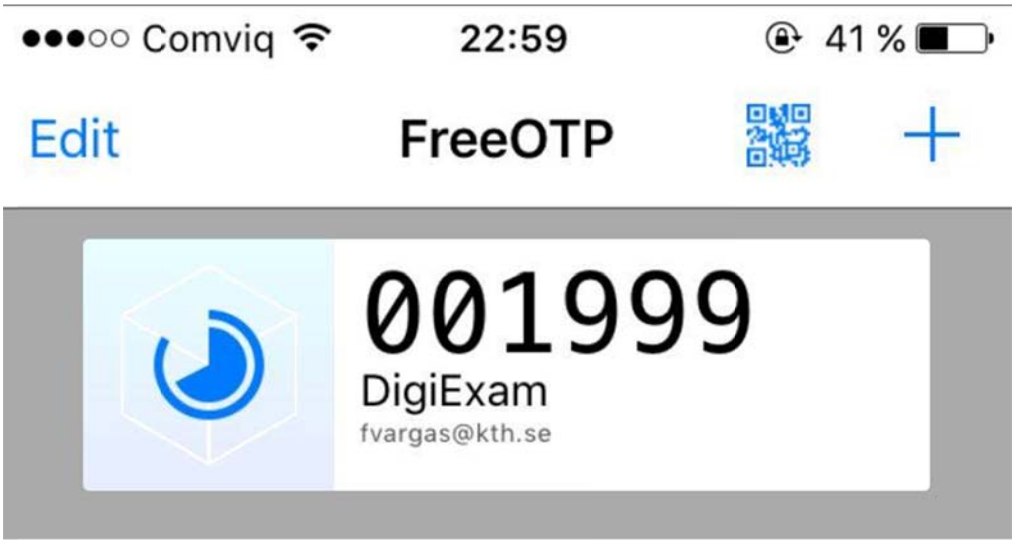
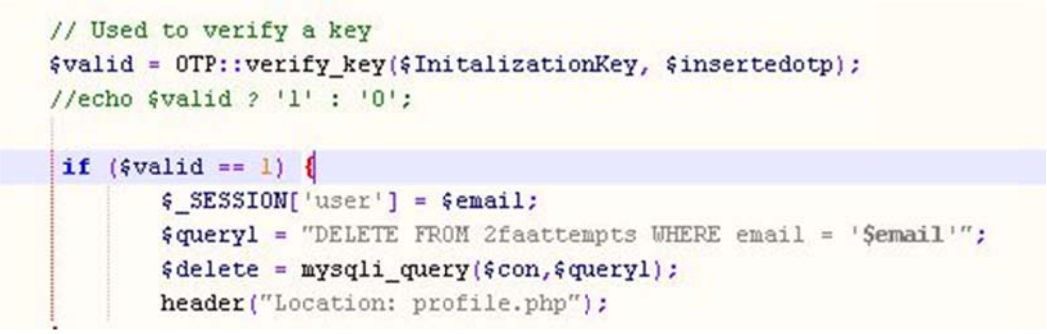


Figure 5-8: FreeOTP applications showing the TOTP after scanning the QR-code.

The page that is shown to the user in which the code has to be inserted is similar to the page shown in Figure 5-7. After enter the code, the code is sent to a PHP validation page on the server that generates a TOTP according to the user's secret key and the Unix time. A class with five PHP function is created in order to generate and validate the code. First, the function has to acquire the time. After that, the secret key is decoded in base32. The third function takes the secret key and the timestamp and returns the one time password, accordingly to RFC 6238. The last two functions truncate the string created by the SHA1 hash in order to extract the TOTP and the last is used to verify the code. Figure 5-9 shows how the code entered by the user is verified by the server. The function *verify_key* is used to check the validity of the entered code and returns *true* or *false*. A *true* value is converted to a string with value *1*, while a *false* value is converted to an empty string. If the function returns *true*, then a user session is created and the user is redirected to the profile page.



```
// Used to verify a key
$valid = OTP::verify_key($InitalizationKey, $insertedotp);
//echo $valid ? '1' : '0';

if ($valid == 1) {
        $_SESSION['user'] = $email;
        $query1 = "DELETE FROM 2faattempts WHERE email = '$email'";
        $delete = mysqli_query($con,$query1);
        header("Location: profile.php");
```

Figure 5-9: Validation of the TOTP by the server

5.2.2.2.1 Notes about implementation of TOTP server

The client's clock and the server's clock need to be synchronized in order for both the client and server to generate the same code. In order to ensure that this requirement is met, either the clocks have to be in perfect synchronization or the token inserted is checked against a delayed version of the current server's time. For security reasons, this is not a good choice since it allows for a greater chance of guessing the token, but it is a simple solution that can be modified to allow a minor difference in time of day. The secret key has been chosen to be 16 digits long in order to prevent an attacker from guessing the value. If the key is too small and the attacker is able to intercept few tokens, then the seed value may be discovered by brute-force.

5.3 Delimitations in the implementation

This section outline the delimitations of the implemented prototype. The aspects that we decided not to focus on during the implementation are UX, esthetical view, account recovery, and security. These aspects were not part of the scope of the prototype's implementation since we decided to focus on demonstrating the essentials of 2FA. Moreover, we chose to not prioritise these aspects due to the limited duration of this thesis project.

Regarding the user's experience, we built the application in order to demonstrate 2FA without putting too much focus on how a user might feel using the prototype. For instance, when scanning the QR-code, the application does not recognize when the user has scanned the code. Instead, the application remains on the same page which could confuse a user as feedback is not given and the 2FA OTP setup process might not seem to be complete. Moreover, there are other small details that would have to be fixed before a customer ready solution could be rolled out.

With regard to the esthetic view, we used DigiExam's existing layout and thus some esthetical aspects were left out or could be improved. For instance, JavaScript was not used during validation. Using JavaScript might have made validation messages more effective and esthetically pleasing.

Account recovery was discussed in Section 4.5. However, we did not implement any of the suggested methods for handling account recovery, because this was deemed to be out of the scope of the implementation of a prototype.

When thinking about the security of the implementation, some aspects were taken into account, while other were not. Security aspects such as hashing passwords, preventing SQL injection, blocking a user after a number of failed login attempts and limiting the validity time of TOTPs were considered and implemented in the prototype of the application. However, aspects such as securing the connection with SSL/HTTPS were not taken into account during the implementation due to limited time. Fortunately, these limitations do not directly affect the 2FA implementation.

Furthermore, our proposed solution is a functional implementation of a prototype which aims at presenting how a possible 2FA method could work. This means that no focus has been put on testing in order to measure scalability of the product.

6 Conclusions and Future work

In this chapter, we present the conclusions of our thesis project and suggest some future work. Section 6.3 states some of the limitations of this project, while Section 6.4 suggests some future work. Finally, the thesis ends with some of our reflections about the work we accomplished.

6.1 Conclusions

To summarize, the main goal of this thesis project was to provide DigiExam with the most appropriate 2FA solution for securing teachers' accounts. The problem underlying this thesis project was the lack of security regarding authentication due to single factor authentication, i.e., SFA, methods offering insufficient security for security-critical services. Implementing 2FA in an authentication system could provide greater security protection against phishing, social engineering, and stolen credentials.

Currently, DigiExam does not implement 2FA in their authentication system, hence the users of the system could be subject to the risks mentioned above. Changing from SFA to 2FA should be a painless migration in order to avoid user dissatisfaction and in order to make the overhaul of the company's legacy authentication system easier.

There are different methods by which 2FA could be implemented. However, the conclusion of this thesis project is that a software OTP solution through an application best suits the needs of DigiExam. Implementing an OTP solution avoids a painful migration to 2FA as the proposed prototype allows an incremental roll-out, while providing a desirable solution in terms of low cost and increased security. Moreover, as it is a software solution, it does not require the user to have an extra device as it can run on the user's existing mobile device. Despite having many advantages, a 2FA solution also has drawbacks. The main drawback is the challenges remaining with regard to account recovery. Losing the second authentication factor could imply difficulties for the user regaining access to their account. An alternative and still valid solution that might mitigate the challenges related to account recovery is to implement 2FA through SMS. However, this solution could be costly over time and is *not* as secure as the OTP solution.

The results of our thesis and the implementation of our solution were presented to DigiExam's supervisor and employees. The solution was believed to be feasible and of value by DigiExam. Due to the fact that the company wants to roll out a 2FA solution for its end users in the near future, the supervisor was content to know that there is a painless and cheap manner in which 2FA could be implemented.

For the reasons presented and explained in the background section, we believe that every company that uses an authentication system should take into serious consideration the risks arising with SFA and therefore should implement an appropriate 2FA method in order to offer to end users the possibility to enhance the security of their accounts.

6.2 Achieved goals

The two main goals of the thesis project were met. A prototype 2FA solution was provided to DigiExam. Moreover, we analyzed different 2FA methods in terms of cost, UX, deployment, and security. To analyze out-of-the-box solutions and how account recovery could be handled when implementing 2FA were goals as well. Account recovery was analyzed with regard to SMS and OTP 2FA. However, the goal of analyzing different out-of-the-box solutions was not entirely met. The thesis only presents and describes the YubiKey solution. In order to achieve all of the goals, additional out-of-the-box solutions should be analyzed and compared with it.

The goal we set at the beginning was to *possibly* implement *one* 2FA solution. However, a prototype was implemented for both the SMS solution and for the OTP solution; hence we managed to develop **two** working 2FA solutions. Despite not being an optimal solution, we first implemented the SMS solution because we thought it would have required more time than it was available to develop the OTP solution. However, after having developed the SMS-based solution, we decided to implement the OTP one because we realized we had sufficient time to implement it as well.

6.3 Limitations

During the thesis project, some limitations were encountered. The main limitation concerns the analysis of existing out-of-the-box solutions. Due to limited time and insufficient background knowledge, it was challenging to analyze other out-of-the-box solutions than the one described. Another limitation was the lack of an evaluation of the UX when applying either of the two different 2FA methods. While it was sometimes possible to rely on existing research to determine the expected UX, some assumptions based on the factors affecting UX had to be made in other cases. It would have been possible to avoid making these assumptions by conducting a tailor made study, but the limited time made this infeasible. This means that the proposed solution may not be optimal in terms of UX.

6.4 Future work

This section suggests some future work that could build on this thesis. The major future work is related to the implementation of the solution in DigiExam's production software. In our prototype once the 2FA option is selected, the user is always required to provide the additional means of authentication, regardless of when the last login occurred or from which device. A useful and interesting improvement would be to optimize the implementation so that the user only has to provide a TOTP when logging in from a new and unrecognized device or after having cleared their browser's cache or deleted their history. The same thing applies for SMS 2FA.

Moreover, although account recovery was theoretically analyzed account recovery was **not** implemented in the prototype. This means that account recovery functions should be added to the solution in order to make it fully functional.

The next obvious thing to be done is implement the proposed solution in DigiExam's existing service. While our solution is functional, it is still a prototype and needs further improvements and modifications in order to suit DigiExam's programming environment. Furthermore, modifications to the company's server and database have to be made before the solution could be rolled out by DigiExam.

Moreover, the scalability of the solution has not been tested. It could be interesting to test it in order to see how it works when lots of requests are sent simultaneously.

6.5 Reflections

This section presents the authors' reflections regarding economic, social, environmental, and ethical aspects related to the thesis project.

The rapid technological advancements in the ICT domain have led to an overall digitalization of services and processes. The growth of ICT has also embraced the educational area, in particularly leading to the establishment of digital assessment platforms. Schools and universities now use these digital assessment platforms to digitalize processes that were previously handled manually and thus time-consuming. For instance, exams can now be taken on-line and corrected automatically which leads, among other things, to a savings in time and costs. Digitalization of assessments saves costs by reducing paper use, thus having a positive impact on the environment.

On the other hand, the growth of ICT is linked to a growth in the production and consumption of electronic devices, which could negatively affect the environment if these devices are not disposed of in a proper manner. Moreover, more energy is consumed which depending upon the source of this energy may lead to other issues such as climate changes.

There are also ethical impacts of ICT as personal information is stored in the cloud and communicated via the Internet, potentially negatively affecting the privacy of the users. Privacy has a substantial moral importance and information technology has a significant impact on it. Privacy is also a social issue, and the benefits of maintaining privacy as well as the costs of failing to achieve or losing privacy need to be understood. This is particularly relevant for companies, such as DigiExam, as they handle sensitive information and hence they have to ensure the security of their users' accounts.

References

[1] '2016 Data Breach Investigations Report', Verizon, Executive Summary, Apr. 2016 [Online]. Available: http://www.verizonenterprise.com/resources/reports/rp_dbir-2016-executive-summary_xg_en.pdf. [Accessed: 23-Mar-2017]

[2] V. Goel and N. Perlroth, 'Yahoo Says 1 Billion User Accounts Were Hacked', *The New York Times*, 14-Dec-2016 [Online]. Available: https://www.nytimes.com/2016/12/14/technology/yahoo-hack.html. [Accessed: 22-Mar-2017]

[3] J. Moar, 'The Future of Cybercrime & Security: Financial and Corporate Threats & Mitigation', Juniper, Dec. 2015 [Online]. Available: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion. [Accessed: 21-Mar-2017]

[4] 'Internet Security Threat Report 2015', Symantec, Dec. 2015 [Online]. Available: https://www.symantec.com/security_response/publications/monthlythreatreport.jsp?id=2015-12. [Accessed: 24-Apr-2017]

[5] 'Creating a strong password - Accounts Help'. [Online]. Available: https://support.google.com/accounts/answer/32040?hl=en

[6] 'TeleSign-Consumer-Account-Security-Report-2015-FINAL.pdf', TeleSign, Jun. 2015 [Online]. Available: https://www.telesign.com/wp-content/uploads/2015/06/TeleSign-Consumer-Account-Security-Report-2015-FINAL.pdf. [Accessed: 21-Mar-2017]

[7] 'Most-Common-Passwords-of-2016-Keeper-Security-Study.pdf', Keeper, 2017 [Online]. Available: https://keepersecurity.com/public/Most-Common-Passwords-of-2016-Keeper-Security-Study.pdf. [Accessed: 21-Mar-2017]

[8] 'Secondary education statistics', Eurostat, Dec. 2015 [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Secondary_education_statistics#Teachers_and_pupil.E2.80.93teacher_ratios. [Accessed: 22-Mar-2017]

[9] G. Sim, P. Holifield, and M. Brown, 'Implementation of computer assisted assessment: Lessons from the literature', *ALT-J Assoc. Learn. Technol. J.*, vol. 12, no. 3, Sep. 2004 [Online]. DOI: 10.1080/0968776042000259546

[10] M. Thelwall, 'Computer-based assessment: A versatile educational tool', *Comput. Educ.*, Mar. 2013 [Online]. DOI: 10.1016/S0360-1315(99)00037-8

[11] A. C. Croft, M. Danson, B. R. Dawson, and J. P. Ward, 'Experiences of using computer assisted assessment in engineering mathematics', *Comput. Educ.*, vol. 37, no. 1, pp. 53–66, Feb. 2001. DOI: 10.1016/S0360-1315(01)00034-3

[12] 'Going Digital, Top Business Schools Transition to Digital Assessments', *DigiExam*, 16-Feb-2017. [Online]. Available: https://www.digiexam.com/company/blog/going-digital-top-business-schools-transition-digital-assessments/. [Accessed: 22-Mar-2017]

[13] T. Baekdal, 'The Usability of Passwords', *Baekdal*, 08-Nov-2007. [Online]. Available: https://www.baekdal.com/insights/password-security-usability/

[14] T. Hunt, 'Have I been pwned?', n.d. [Online]. Available: https://haveibeenpwned.com/

[15] D. Wang, D. He, P. Wang, and C. H. Chu, 'Anonymous Two-Factor Authentication in Distributed Systems: Certain Goals Are Beyond Attainment', *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 4, pp. 428–442, Jul. 2015. DOI: 10.1109/TDSC.2014.2355850

[16] M. Bond, O. Choudary, S. J. Murdoch, S. Skorobogatov, and R. Anderson, 'Chip and Skim: cloning EMV cards with the pre-play attack', Paper, University of Cambridge, UK [Online]. Available:

http://sec.cs.ucl.ac.uk/users/smurdoch/papers/oakland14chipandskim.pdf. [Accessed: 28-Mar-2017]

[17] C. Mulliner, B. Ravishankar, P. Stewin, and J.-P. Seifet, ‘SMS-Based One-Time Passwords: Attacks and Defense’, in *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, Springer, Berlin, Heidelberg, 2013, vol. 7967, pp. 150–159 [Online]. DOI: 10.1007/978-3-642-39235-1_9

[18] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, ‘Performance Evaluation of Fingerprint Verification Systems’, *IEEE Trans. PATTERN Anal. Mach. Intell.*, vol. 28, no. 1, Jan. 2006 [Online]. Available: http://biometrics.cse.msu.edu/Publications/Fingerprint/FVC/Cappellietal_FpPerformanceEvaluation_PAMI06.pdf. [Accessed: 28-Mar-2017]

[19] A. S. Patrick, ‘Fingerprint Concerns: Performance, Usability, and Acceptance of Fingerprint Biometric Systems’, Essay, National Research Council of Canada, Ottawa, ON Canada, 2008 [Online]. Available: https://www.andrewpatrick.ca/essays/fingerprint-concerns-performance-usability-and-acceptance-of-fingerprint-biometric-systems/. [Accessed: 28-Mar-2017]

[20] ‘German Defense Minister von der Leyen’s fingerprint copied by Chaos Computer Club’, *DW*, 28-Dec-2014 [Online]. Available: http://www.dw.com/en/german-defense-minister-von-der-leyens-fingerprint-copied-by-chaos-computer-club/a-18154832. [Accessed: 28-Mar-2017]

[21] M. Theofanos, ‘Health and Safety Perceptions of Biometric Devices’, *NIST*, Nov. 2006 [Online]. Available: http://zing.ncsl.nist.gov/biousa/docs/Health_Safety.pdf. [Accessed: 28-Mar-2017]

[22] ‘What is 2FA?’, *securenvoy.com*. [Online]. Available: https://www.securenvoy.com/two-factor-authentication/what-is-2fa.shtm. [Accessed: 28-Mar-2017]

[23] ‘National Information Assurance (IA) Glossary’. CNSS, 26-Apr-2010 [Online]. Available: https://www.ncsc.gov/nittf/docs/CNSSI-4009_National_Information_Assurance.pdf. [Accessed: 28-Mar-2017]

[24] ‘Recommendations for the security of internet payments’, European Central Bank, Frankfurt, Germany, Jan. 2013 [Online]. Available: https://www.ecb.europa.eu/pub/pdf/other/recommendationssecurityinternetpaymentsoutcomeofpcfinalversionafterpc201301en.pdf. [Accessed: 28-Mar-2017]

[25] ‘Text for ITU - T Recommendation X. 1254 | ISO/IEC DIS 2911 5 -- Information technology – Security techniques – Entity authentication assurance framework’, *ISOIEC JTC 1SC 27*, Nov. 2011 [Online]. Available: https://www.oasis-open.org/committees/download.php/44751/285-17Attach1.pdf. [Accessed: 28-Mar-2017]

[26] E. Feigenbaum, ‘A more secure cloud for millions of Google Apps users’, *Google Cloud Official Blog*. 20-Sep-2010 [Online]. Available: https://cloud.googleblog.com/2010/09/a-more-secure-cloud-for-millions-of.html. [Accessed: 29-Mar-2017]

[27] S. Beaudin, ‘An Empirical Study of Authentication Methods to Secure E-learning System Activities Against Impersonation Fraud’. Nova Southeastern University, 2016 [Online]. Available: http://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1957&context=gscis_etd. [Accessed: 30-Mar-2017]

[28] K. M. Apampa, G. Wills, and A. David, ‘User Security Issues in Summative E-Assessment Security’, *Int. J. Digit. Soc.*, vol. 1, no. 2, Jun. 2010 [Online]. DOI: 10.20533/ijds.2040.2570.2010.0018

[29] Z. Mao, D. Florencio, and C. Herley, ‘Painless Migration from Passwords to Two Factor Authentication’, in *Proceedings of the 2011 IEEE International Workshop on*

Information Forensics and Security, Washington, DC, USA, 2011, pp. 1–6 [Online]. DOI: 10.1109/WIFS.2011.6123150

[30] W. Bedford, J. Gregg, and S. Clinton, 'Implementing Technology to Prevent Online Cheating: A Case Study at a Small Southern Regional University (SSRU)', *J. Online Learn. Teach.*, vol. 5, no. 2, Jun. 2009 [Online]. Available: http://jolt.merlot.org/vol5no2/gregg_0609.pdf. [Accessed: 30-Mar-2017]

[31] E. De Cristofaro, D. Honglu, J. Freudiger, and G. Norcie, 'A Comparative Usability Study of Two-Factor Authentication', in *8th NDSS Workshop on Usable Security*, San Diego, CA, USA, 2014, pp. 1–10 [Online]. Available: https://pdfs.semanticscholar.org/028a/70fc1836e113fd18f12b99e08fb024f6bb04.pdf. [Accessed: 30-Mar-2017]

[32] J. L. Bailie and M. A. Jortberg, 'Online Learner Authentication: Verifying the Identity of Online Users', *J. Online Learn. Teach.*, vol. 5, no. 2, Jun. 2009 [Online]. Available: http://jolt.merlot.org/vol5no2/bailie_0609.pdf. [Accessed: 30-Mar-2017]

[33] B. T. Adetoba, O. Awodele, and S. O. Kuyoro, 'E-learning security issues and challenges: A review', *J. Sci. Res. Stud.*, vol. 3, no. 5, pp. 96–100, May 2016.

[34] S. Alotaibi and D. Argles, 'Abstract Paper: FingerID, A New Security Model Based on Fingerprint Recognition for Personal Learning Environments (PLEs)', *IEEE Eng. Educ. 2011*, pp. 142–151, Apr. 2011.

[35] K.-S. Song, S. M. Lee, and S. Nam, 'Combined Biometrics for e -Learning Security', in *The 7th International Conference on Information Security and Assurance*, Korea National University of Education, Korea, 2013, vol. 21, pp. 247–251 [Online]. Available: http://onlinepresent.org/proceedings/vol21_2013/63.pdf. [Accessed: 30-Mar-2017]

[36] A. Vaithyasubramanian, A. Christy, and D. Saravanan, 'TWO FACTOR AUTHENTICATIONS FOR SECURED LOGIN IN SUPPORT OF EFFECTIVE INFORMATION PRESERVATION AND NETWORK SECURITY', *J. Eng. Appl. Sci.*, vol. 10, no. 5, Mar. 2015 [Online]. Available: http://www.arpnjournals.com/jeas/research_papers/rp_2015/jeas_0315_1713.pdf. [Accessed: 30-Mar-2017]

[37] K. Altinkemer and T. Wang, 'Cost and benefit analysis of authentication systems', *Decis. Support Syst.*, vol. 51, no. 3, pp. 394–404, Jun. 2011. DOI: 10.1016/j.dss.2011.01.005

[38] 'Study: Companies can halve authentication costs by ditching hardware tokens', *Encap Security*. 04-Jul-2012 [Online]. Available: https://www.encapsecurity.com/study-companies-can-halve-authentication-costs-by-ditching-hardware-tokens/. [Accessed: 26-Apr-2017]

[39] 'Businesses have different aims and objectives that can change over time.', *BBC.com*. [Online]. Available: http://www.bbc.co.uk/schools/gcsebitesize/business/aims/partnershiprev2.shtml. [Accessed: 05-Apr-2017]

[40] Riksdagsförvaltningen, *Personuppgiftslag*, vol. 1998:204. 1998 [Online]. Available: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/personuppgiftslag-1998204_sfs-1998-204. [Accessed: 28-Apr-2017]

[41] J. Anil, R. Arun, and P. Salil, 'An Introduction to Biometric Recognition', *IEEE Trans. CIRCUITS Syst. VIDEO Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004. DOI: 10.1109/TCSVT.2003.818349

[42] L. Coventry, G. I. Johnson, T. McEwan, and C. Riley, 'Biometrics in Practice: What Does HCI Have to Say?', in *Human-Computer Interaction – INTERACT 2009*, 2009, pp. 920–921 [Online]. DOI: 10.1007/978-3-642-03658-3_111

[43] ‘Security Token | Citi Private Bank’, *privatebank.citibank.com*. [Online]. Available: https://www.privatebank.citibank.com/our_services/online/digital_key.htm#q7. [Accessed: 26-Apr-2017]

[44] ‘Security Tokens’, *Duo Security*. [Online]. Available: https://duo.com/product/trusted-users/two-factor-authentication/authentication-methods/security-tokens. [Accessed: 26-Apr-2017]

[45] ‘How to Use a Hardware Token for Two-Step Authentication | University IT’. [Online]. Available: https://uit.stanford.edu/service/webauth/twostep/token. [Accessed: 26-Apr-2017]

[46] N. Haller, C. Metz, P. Nesser, and M. Straw, ‘A One-Time Password System’, 1998, vol. RFC 2289 [Online]. Available: http://www.rfc-editor.org/rfc/rfc2289.txt. [Accessed: 22-May-2017]

[47] D. M’Raihi, M. Bellare, F. Hoornaert, D. Naccache, and O. Ranen, ‘HOTP: An HMAC-Based One-Time Password Algorithm’. Request for Comments, Dec-2005 [Online]. Available: http://www.rfc-editor.org/rfc/rfc4226.txt. [Accessed: 22-May-2017]

[48] H. Krawczyk, M. Bellare, and R. Canetti, ‘HMAC: Keyed-Hashing for Message Authentication’. Request for Comments, Feb-1997 [Online]. Available: http://www.rfc-editor.org/rfc/rfc2104.txt. [Accessed: 22-May-2017]

[49] D. M’Raihi, J. Rydell, M. Pei, and S. Machani, ‘TOTP: Time-based One-time Password Algorithm’. IEEE, 2011 [Online]. Available: https://tools.ietf.org/html/rfc6238. [Accessed: 02-May-2017]

[50] O. Findahl and P. Davidsson, ‘Svenskarna och Internet’, Internetstiftelsen i Sverige, 2105 [Online]. Available: http://www.soi2015.se/sammanfattning/. [Accessed: 02-May-2017]

[51] S. Josefsson, ‘The Base16, Base32, and Base64 Data Encodings’. Network Working Group, Oct-2016 [Online]. Available: https://tools.ietf.org/html/rfc4648. [Accessed: 22-May-2017]

TRITA-ICT-EX-2017:46

www.kth.se