Secure and Usable User Authentication

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Peter Mayer, M.Sc.
geb. in Gelnhausen

Tag der mündlichen Prüfung:	15.11.2019
Hauptreferent:	Prof. Dr. Melanie Volkamer
Korreferent:	Prof. Dr. Karen Renaud

Abstract

Authentication is a ubiquitous task in users' daily lives. The dominant form of user authentication are text passwords. They protect private accounts like online banking, gaming, and email, but also assets in organisations. Yet, many issues are associated with text passwords, leading to challenges faced by both, users and organisations. This thesis contributes to the body of research enabling secure and usable user authentication, benefiting both, users and organisations. To that end, it addresses three distinct challenges.

The first challenge addressed in this thesis is the creation of correct, complete, understandable, and effective *password security awareness materials*. To this end, a systematic process for the creation of awareness materials was developed and applied to create a password security awareness material. This process comprises four steps. First, relevant content for an initial version is aggregated (i.e. descriptions of attacks on passwords and user accounts, descriptions of defences to these attacks, and common misconceptions about password and user account security). Then, feedback from information security experts is gathered to ensure the correctness and completeness of the awareness material. Thereafter, feedback from lay-users is gathered to ensure the understandability of the awareness material. Finally, a formal evaluation of the awareness material is conducted to ensure its effectiveness (i.e. whether the material improves participant's ability to assess the security of passwords as well as password-related behaviour and decreases the prevalence of common misconceptions about password and user account security). The results of the evaluation show the effectiveness of the awareness material: it significantly improved the participants' ability to assess the security of password-related behaviour as well as passwords and significantly decreased the prevalence of misconceptions about password and user account security.

The second challenge addressed in this thesis is shoulder-surfing resistant text password entry with gamepads (as an example of very constrained input devices) in shared spaces. To this end, the *very first investigation of text password entry with gamepads* is conducted. First, the requirements of authentication in the gamepad context are described. Then, these requirements are applied to assess schemes already deployed in the gamepad context and shoulder-surfing resistant authentication schemes from the literature proposed for non-gamepad contexts. The results of this assessment show that none of the currently deployed and only four of the proposals in the literature fulfil all requirements. Furthermore, the results of the assessment also indicate a need for an empirical evaluation in order to exactly gauge the shoulder-surfing threat in the gamepad context and compare alternatives to the incumbent on-screen keyboard. Based on these results, two user studies (one online study and one lab study) are conducted to investigate the shoulder-surfing resistance and usability of three authentication schemes in the gamepad context: the *on-screen keyboard* (as de-facto standard in this context), the *grid-based scheme* (an existing proposal from the literature identified as the most viable candidate adaptable to the gamepad context during the assessment), and *Colorwheels* (a novel shoulder-surfing resistant authentication scheme specifically designed for the gamepad context). The results of these two user studies show that on-screen keyboards are highly susceptible to opportunistic shoulder-surfing, but also show the most favourable usability properties among the three schemes. *Colorwheels* offers the most robust shoulder-surfing resistance and scores highest with respect to participants' intention to use it in the future, while showing more favourable usability results than the grid-based scheme.

The third challenge addressed in this thesis is secure and efficient storage of passwords in portfolio authentication schemes. Portfolio authentication is used to counter capture attacks such as shoulder-surfing or eavesdropping on network traffic. While usability studies of portfolio authentication schemes showed promising results, a verification scheme which allows secure and efficient storage of the portfolio authentication secret had been missing until now. To remedy this problem, the (t, n)*-threshold verification scheme* is proposed. It is based on secret sharing and key derivation functions. The security as well as the efficiency properties of two variants of the scheme (one based on Blakley secret sharing and one based on Shamir secret sharing) are evaluated against each other and against a naive approach. These evaluations show that the two (t, n)-threshold verification scheme variants always exhibit more favourable properties than the naive approach and that when deciding between the two variants, the exact application scenario must be considered. Three use cases illustrate as exemplary application scenarios the versatility of the proposed (t, n)-threshold verification scheme.

By addressing the aforementioned three distinct challenges, this thesis demonstrates the breadth of the field of usable and secure user authentication ranging from awareness materials, to the assessment and evaluation of authentication schemes, to applying cryptography to craft secure password storage solutions. The research processes, results, and insights described in this thesis represent important and meaningful contributions to the state of the art in the research on usable and secure user authentication, offering benefits for users, organisations, and researchers alike.

Acknowledgements

First and foremost, I want to thank my supervisor, Melanie Volkamer, for the unwavering support and guidance as well as the time taken for the many discussions about my research. Her insightful mentoring has been an invaluable help in completing this dissertation. Thank you for believing in me, my research, and this whole endeavour.

Furthermore, I want to express my deepest gratitude to Karen Renaud. Thank you for the invaluable feedback on my research projects and this dissertation in particular.

Also I want to thank Andreas Oberweis and Alexander Mädche for agreeing to be part of my dissertation committee and for the interesting discussions during the defence.

Moreover, I would like to give many thanks to my colleagues in the SECUSO research group for their valuable feedback on research ideas, paper drafts, and practice presentations. Their advice and insights helped me to improve my experiments and analyses considerably. In particular, I want to thank all my colleagues for the very fruitful discussions leading to our joint publications.

I also want to thank the project partner usd AG for helping making the KMU AWARE project a success, for making my research possible within the project, and last but not least for the recruiting of the participants for the research underlying the first part of this dissertation.

My thanks also go to Dennis Hofheinz, Willi Geiselmann, Johannes Braun, and Matthias Schempp for their valuable feedback and insights on the third part of this dissertation. I highly appreciate the time taken to listen to my questions and ideas and discuss my research.

I want to thank my family and friends for their steadfast support at all times. I could never express in mere words how much their understanding, enthusiasm, and moral support have helped me throughout the years. I am looking forward to having more time to spend with you now that this dissertation is completed.

Finally, I want to thank all the participants of the studies I conducted and the people helping in the recruiting. Much of my research would not have been possible without them.

Contents

Introduction

1 Introduction

Authenticating as legitimate user is a ubiquitous task in everybody's daily digital life. In the private life, user authentication protects accounts for e.g. online banking, email, and social networks. In the professional life, it protects all assets of organisations. In both instances, text passwords are the dominant authentication scheme [98]. Every user has on average 6 to 7 text passwords for Internet accounts alone and enters one of these passwords on average 8 times each day [71]. Due to this ubiquity of text passwords, many investigations regarding their properties exist and research has identified serious issues associated with them, leading to challenges faced by both, users as well as organisations.

On the user side, the plethora of challenges caused by text passwords has been well documented over the years. The sheer number of text passwords users are expected to handle has been shown to overwhelm them [194]. While users have on average 25 to 80 [71, 92, 196] accounts, these are only protected on average by 6 to 7 unique text passwords [71, 92, 196]. However, the number of actual unique passwords users appear to be able to memorise is with around 4 to 5 even lower [34, 81]. Additionally, user authentication – and security as a whole – is usually not the user's primary task, but rather overhead which interferes with them getting their actual (primary) tasks done [33]. Therefore, while users do not have an inherent lack of willingness to behave securely [2, 177], their willingness to spend time and effort is understandably low [97]. However, even users willing to spend additional effort and time to inform themselves about secure practices and clear up their misconceptions might end up following advice that was created unsystematically and contradicts the state of the art in terms of password security [149, 236]. Consequentially, many users face problems when choosing, handling, or remembering their text passwords [107, 194, 195], ultimately leading them to develop potentially insecure coping strategies [194,195] or even authentication fatigue [80,178]. This situation is further aggravated by the increasing need to use text password-protected accounts on devices with constrained input devices relying on on-screen keyboards [217], such as touchscreens on smartphones or gamepads on game consoles. When such devices are used in public spaces (smartphones) or shared spaces (gamepads on game consoles) attacks such as shoulder-surfing become a feasible threat [169]. While traditionally, all the challenges faced by users have caused users being referred to as "the weakest link" in the information security chain [181], this thesis is part of the research focusing on addressing these challenges instead of blaming the user [163, 177, 240].

On the organisation side, the aforementioned challenges on the user side are equally faced by employees of the organisations and therefore translate directly to these organisations. Targeting passwords is the most prevalent tactic for attacks on organisations: in 2016 [213] 63% of breaches could be attributed to leveraging weak, default, or stolen passwords. In 2017 [214] this number was found to have increased to 81%[1]. Likewise, a 2019 report [18] focusing on the threat landscape in Germany indicates that attacks on passwords are the ones most frequently faced by organisations. In addition to the challenges arising for employees, organisations face issues arising from the operation of text password authentication infrastructure. Two of the most important best practices when operating text password authentication infrastructure is offering

[1] As of July 2019, the "2017 Data Breach Investigations Report" seems to be the last one including this metric. Therefore, no newer values for this metric can be reported.

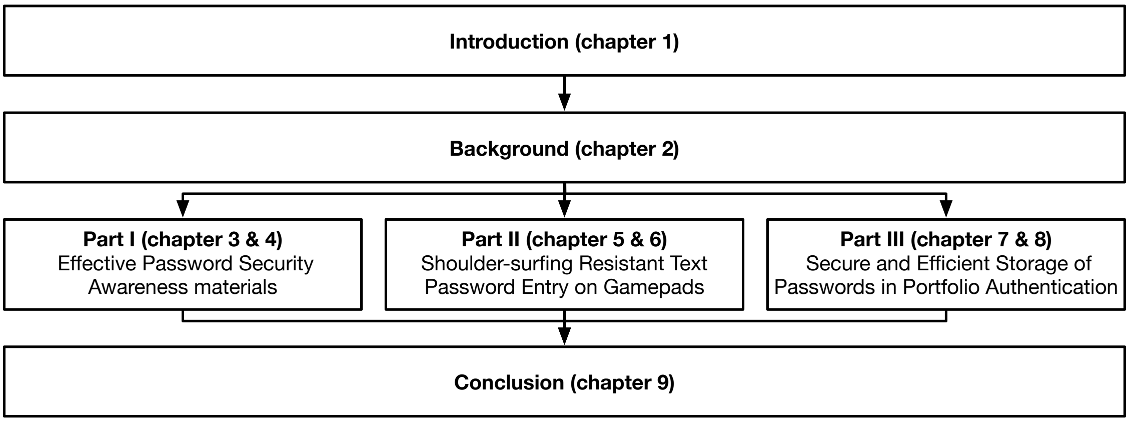

Introduction (chapter 1)
Background (chapter 2)
Part I (chapter 3 & 4)
Effective Password Security Awareness materials
Part II (chapter 5 & 6)
Shoulder-surfing Resistant Text Password Entry on Gamepads
Part III (chapter 7 & 8)
Secure and Efficient Storage of Passwords in Portfolio Authentication
Conclusion (chapter 9)

Figure 1.1: Overview of the structure of this thesis.

secure transmission of the password from the client to the server (i.e. using encrypted transmission [28]) and secure password storage on the server side (i.e. using salted hashes [73]). Unfortunately, organisations struggle with these long standing prerequisites of password security on the server side [28]. While *Let's Encrypt*[2] has enabled broad availability of securely encrypted transmission of passwords since its launch in 2016 [4], the continued struggle with secure password storage is evidenced by recent lapses regarding this best practice by large Internet companies such as Twitter [5], Google [76], Facebook [43], and Github [225] as well as the long list of other websites storing password in plain text[3].

In summary, both, on the user side as well as the organisation side, a plethora of challenges arises pertaining to text passwords.

1.1 Scope and Structure of this Work

User authentication involves three entities: the *user* who wants to authenticate, the *verifier* (e.g. device or remote server) the user authenticates to, and the *authentication scheme* (including its interface for input/output from/to the user and communication with the verifier) providing the actual authentication procedure. The work presented in this thesis contributes to the body of research on secure and usable user authentication, by addressing three distinct challenges – one with respect to each of the aforementioned entities user, authentication scheme, and verifier: (I: *User*) the availability of effective password security awareness materials, allowing interested users who are willing to spend time and effort to learn about secure password-related behaviour and clear up common misconceptions about password security, (II: *Authentication Scheme*) the availability of shoulder-surfing resistant text password entry on constrained input devices (gamepads) in shared spaces, and (III: *Verifier*) the availability of secure (i.e. using salted hashes) and efficient (i.e. low storage requirement) password storage in portfolio authentication schemes. Each of these three challenges is presented in one distinct part of this thesis. Thereby, the three parts – together with a preceding chapter introducing the necessary background (chapter 2) and a concluding chapter summarising the overall findings and contributions (chapter 9) – form the structure of the remainder of this thesis (cf. figure 1.1). In the following, the scope of each of the three main parts is described.

[2] `https://letsencrypt.org/` (accessed: 2019-08-05)

[3] `https://github.com/plaintextoffenders/plaintextoffenders/blob/master/offenders.csv` (accessed: 2019-08-05)

1.1.1 Part I: Effective Password Security Awareness Materials

The first challenge addressed in this thesis is the availability of effective awareness materials, allowing interested users who are willing to spend time and effort to learn about secure password-related behaviour and clear up their misconceptions about password security. In general, password security awareness materials are the prime way to (a) making users aware of the different attacks on passwords and user accounts as well as defences against these attacks and (b) clear up the misconceptions about password security the users might have. However, most advice available to end-users was not created in a systematic way, does not reflect the state of the art in password research and best practice, and is seldom formally evaluated. Therefore, the first part of this thesis (chapter 3 & 4) presents a procedure to systematically create awareness materials and describes the application of the procedure to create a password security awareness material, and subsequently evaluates this material with employees in three German small and medium-sized enterprises (SME) in order to validate its effectiveness.

Chapter 3 presents the first step in the creation procedure: a systematic literature review identifying the misconceptions about password security prevalent in users. Identifying and clearing up these misconceptions is vital in order to decrease insecure coping strategies and promote secure coping strategies. The literature review revealed that misconceptions exist in basically all aspects of password security and indicates a strong need to create interventions clearing up these misconceptions. In combination with the attacks and respective defences described in section 2.3 of the background chapter, the identified misconceptions form the basis for the remaining steps in the procedure as presented in chapter 4.

Chapter 4 then presents the development of the password security awareness material. The systematic development process employed as part of the creation procedure is comprised of four iterative steps: (1) the initial version of the material aggregates the relevant content, i.e. descriptions of relevant attacks and defences and newly created interventions to clear up the identified misconceptions, (2) structured feedback of independent experts ensures the material's correctness and completeness, (3) feedback of lay-users ensures the material's visual appeal and understandability, and (4) an evaluation in the field with employees of three organisations ensures the material's effectiveness and provides additional feedback for improvements to create the final version of the material. The results of the evaluation show that the password security awareness material was not only received very positively by the employees, but also validate its effectiveness by showing that it (a) significantly improves the employees capacity to behave securely with respect to their passwords and (b) significantly decreases the prevalence of the identified misconceptions among the employees.

The research presented in the first part of this thesis directly benefits both, users and organisations. On the one had, users who are willing to spend time and effort to improve their password-related behaviour and clear up their misconceptions about password security find in the awareness material a correct, complete, and understandable source of information about password security. On the other hand, organisations can distribute the password security awareness material to their employees and rest assured that the time and effort of their employees is well spent to effectively increase the password-related security. This is especially important for organisations which are obligated to distribute awareness materials for compliance reasons, since they can reap additional security benefits beyond pure compliance certifications. Additionally, apart from the direct benefits for users and organisations, the systematic process described in this part can guide researchers and practitioners in the creation and evaluation of awareness materials on topics other than password security.

1.1.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

The second challenge addressed in this thesis is the development of shoulder-surfing resistant text password entry, when using constrained input devices such as gamepads in shared spaces. Using gamepad-driven devices like games consoles is an activity frequently shared with others and thus takes place in shared spaces. Consequently, shoulder-surfing becomes a threat in this setting. The second part of this thesis (chapter 5 & 6) presents the first investigation of text password entry on gamepads regarding the shoulder-surfing resistance and the usability of existing and novel authentication schemes in this context.

Chapter 5 first outlines the requirements of text password entry using gamepads and then presents an assessment along these requirements of schemes currently deployed on gamepad-driven devices as well as shoulder-surfing resistant authentication schemes proposed in non-gamepad contexts. The results of this assessment show that none of the currently deployed and only four of the proposals in the literature fulfil all requirements of authentication in the gamepad context. Yet, since the four authentication schemes fulfilling all requirements were developed for non-gamepad context, an empirical evaluation is needed to gauge the performance in terms of shoulder-surfing resistance and usability of the incumbent schemes and the potential replacements.

Chapter 6 presents an empirical evaluation in the form of two user studies - one online study and one lab study - to comparatively evaluate the shoulder-surfing resistance and the usability of three authentication schemes in the gamepad context: the on-screen keyboard (the incumbent and de facto standard in this context), the grid-based scheme (an existing shoulder-surfing resistant scheme identified as viably adaptable during the assessment in chapter 5), and “Colorwheels” (a novel shoulder-surfing resistant authentication scheme specifically geared towards the gamepad authentication context). The results of this evaluation show that on-screen keyboards are highly susceptible to shoulder-surfing attacks, but fare best in terms of usability when compared to the other two schemes. In contrast, the “Colorwheels” scheme, which is specifically designed for usage with gamepads, seems to offer the most robust shoulder-surfing resistance and is rated highest by the participants in terms of intention to use the scheme in the future, while still exhibiting more favourable usability results than the adapted grid-based scheme.

The research presented in this part of the thesis informs the future exploration of text password entry in the gamepad context. The results can inform the design of authentication using gamepads. Furthermore, the evaluation of authentication schemes in this context can benefit organisations that want to enable users to enter text passwords in a socially acceptable way even in shared spaces (i.e. not asking family or friends to look away during password entry) and therefore ultimately users as well.

1.1.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

The third challenge addressed in this thesis is the secure and efficient storage of passwords in portfolio authentication. Portfolio authentication is a technique to counter-shoulder surfing attacks, originally intended for usage with graphical passwords. In portfolio authentication the password is regarded as a set of elements and every authorised subset of password elements is sufficient to authenticate a user. However, while results in usability studies of portfolio authentication schemes are promising, the lack of a verification scheme which allows secure and efficient storage of passwords in portfolio authentication schemes is an open question. Therefore, the third part of this thesis (chapter 7 & 8) presents the novel (t, n)*-threshold verification scheme* offering secure and efficient storage of passwords in portfolio authentication schemes. Its usage is illustrated based on three use cases.

Chapter 7 presents the novel (t, n)-threshold verification scheme, which allows to derive the same secret from all authorised subsets of the password elements by utilising cryptographic secret sharing and key derivation functions. Two variants of the scheme are described: one based on Blakley secret sharing and one based on Shamir secret sharing. Security as well as efficiency properties of these two (t, n)-threshold verification variants are evaluated against a naive implementation and against each other. The results of these evaluations show that the (t, n)-threshold verification scheme offers in all instances more favourable properties than the naive approach and that choice between the two variants depends on the specific application scenario.

Chapter 8 then presents three use cases of the (t, n)-threshold verification scheme, namely its application to three knowledge-based authentication schemes: graphical recognition-based passwords, partial passwords, and ZeTA (an authentication scheme designed to harness the human capability to build up semantic networks of related concepts). In addition, the application of the proposed (t, n)-threshold verification scheme beyond the knowledge-based authentication use cases is discussed.

The research presented in this part directly benefits all organisations which want to employ portfolio authentication schemes to counter capture attacks such as shoulder-surfing or surveillance of network traffic. This in turn allows users to benefit from the deployment of such authentication schemes. The three use cases described in chapter 8 can serve as blueprints for implementations. In addition, the use cases can guide researchers towards secure and efficient storage solutions in the design of novel portfolio authentication schemes.

1.2 Main Contributions

This research contributes to the field of usable and secure user authentication. It addresses three challenges in three distinct parts, as outlined in the last section. The main contributions of each of the three parts are summarised below.

1.2.1 Part I: Effective Password Security Awareness Materials

- Overview of misconceptions about password and user account security reported in the literature. The systematic literature review identifies 23 different misconceptions, which can be grouped into four categories: composition, handling, attacks and miscellaneous. This shows that there exist misconceptions with respect to a wide range of aspects of password security.
- Systematic process for the development of correct, complete, understandable, and effective awareness materials. The systematic iterative development process for information security awareness materials presented in this part of the thesis combines the state of the art described in the current scientific literature, the expertise of independent experts from academia and industry, feedback from lay-users, and an evaluation of the awareness material. Thereby, it enables the creation of correct, complete, understandable, and effective information security awareness materials.
- Evaluated password security awareness material. The password security awareness material created by application of the aforementioned systematic process significantly improves users' ability to assess

password-related behaviour, improves their ability to assess the security of passwords, and clears up the identified misconceptions decreasing their overall prevalence.

1.2.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

- Requirements of authentication in the gamepad context. Overall six requirements which all authentication schemes in the gamepad context should fulfil are identified. These requirements can be grouped into three categories: security, technical, and usability. The requirements can inform the design of new authentication schemes in the gamepad context.
- Assessment of authentication schemes along the identified requirements. The authentication schemes currently deployed in the gamepad context as well as a representative set of shoulder-surfing resistant authentication schemes proposed in the literature are assessed along the identified requirements. The results of this assessment show that none of the currently deployed and only four of the proposals in the literature fulfil all requirements.
- Baseline in terms of usability and shoulder-surfing resistance for the incumbent on-screen keyboard. By conducting two user studies – an online study and a lab study using similar methodologies – a baseline for the on-screen keyboard (as incumbent and de-facto standard of authentication in the gamepad context) is established.
- Comparative assessment of two alternatives to the on-screen keyboard. By evaluating them based on the same methodology as in the baseline assessment, the viability in terms of shoulder-surfing resistance and usability of two alternatives of the on-screen keyboard is assessed:

  (a) The grid-based scheme, which is a shoulder-surfing resistant authentication scheme proposed in the literature and identified as the most viable candidate adaptable to the gamepad context.

  (b) The novel Colorwheels scheme which was specifically designed for the gamepad context.

  Overall, the Colorwheels scheme seems to exhibit the most robust shoulder-surfing resistance among all three evaluated schemes. In terms of usability, Colorwheels scores better than the grid-based scheme, but the on-screen keyboard scores best in this regard indicating a trade-off between security and usability. However, since Colorwheels is rated highest by the participants in terms of intention to use the scheme in the future, this seems to be a trade-off users are willing to make.

1.2.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

- Requirements of verification for portfolio authentication schemes. A compilation of the general operations and properties required of verification schemes in the domain of portfolio authentication gives an overview of the requirements any verification scheme should fulfil.
- The (t, n)-threshold verification scheme. The (t, n)-threshold verification scheme directly addresses the challenge of secure and efficient password storage in portfolio authentication schemes. Two variants of the verification scheme are presented using two different secret sharing schemes, i.e. Blakley secret sharing and Shamir secret sharing.

- Comparison of the two proposed (t, n)-threshold verification variants and a naive approach regarding security and efficiency. In a comparative evaluation, the security and efficiency properties of the two proposed (t, n)-threshold verification variants are assessed against a naive approach and against each other. The (t, n)-threshold verification variants generally perform better.
- Three use cases for the (t, n)-threshold verification scheme. Application of the (t, n)-threshold verification scheme is described for three exemplary use cases: graphical recognition-based passwords, partial passwords, and the ZeTA authentication scheme.
- Further areas of application. While the three use cases of the (t, n)-threshold verification scheme described in this thesis all belong to the field of knowledge-based authentication, the area of application for the scheme covers far more general contexts including different types of authentication (i.e. knowledge-based, token-based, biometric) in order to provide fault tolerance or authentication schemes which adapt to the types of authentication available in the particular environment where a user acts at any given time (e.g. specific biometric sensors).

1.3 Related Publications

The research presented in this thesis is based on and extends previous publications. Furthermore, several publications have indirectly contributed to this thesis. In the following these publications are presented for each of the three main parts.

1.3.1 Part I: Effective Password Security Awareness Materials

This part of the thesis is based on the following publications:

- **P. Mayer** and M. Volkamer, “Addressing Misconceptions About Password Security Effectively”, Workshop on Socio-Technical Aspects in Security and Trust (STAST, co-located with ACSAC), 2017.
- **P. Mayer**, C. Schwartz, and M. Volkamer, “On The Systematic Development and Evaluation Of Password Security Awareness-Raising Materials”, Annual Computer Security Applications Conference (ACSAC), 2018, pp. 733–748.

The following publications contributed indirectly to this part of the thesis:

- **P. Mayer**, A. Kunz, and M. Volkamer, “Analysis of the Security and Memorability of the Password Card”, Annual Computer Security Applications Conference Posters (ACSAC Posters), 2017.

  The password card is intended to increase password memorability. The investigation of the properties of this solution allowed deciding whether it should be recommended in the password security awareness material or excluded from it.
- **P. Mayer**, J. Kirchner, and M. Volkamer, “A Second Look at Password Composition Policies in the Wild: Comparing Samples from 2010 and 2016”, Symposium on Usable Privacy and Security (SOUPS), 2017, pp. 13–28.

  The overview of password composition policies given in this paper allowed gauging the requirements put on passwords in the wild and therefore consider these requirements in the material.

- S. Stockhardt, B. Reinheimer, M. Volkamer, **P. Mayer**, A. Kunz, P. Rack, and D. Lehmann, “Teaching Phishing-Security: Which Way is Best?”, IFIP International Conference on ICT Systems Security and Privacy Protection (SEC 2016), vol. 471, 2016.

  The lessons learned from the evaluation of the materials in this publication informed the design of the evaluation of the newly developed password security awareness material in this part of the thesis.

1.3.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

This part of the thesis is based on the following publication:

- **P. Mayer**, N. Gerber, B. Reinheimer, P. Rack, K. Braun, and M. Volkamer, “I (don’t) see what you typed there! Shoulder-surfing resistant password entry on gamepads”, Conference on Human Factors in Computing Systems (CHI), 2019.

The following publication contributed indirectly to this part of the thesis:

- **P. Mayer**, S. Neumann, D. Storck, and M. Volkamer, “Supporting Decision Makers in Choosing Suitable Authentication Schemes,” International Symposium of Human Aspects of Information Security Assurance (HAISA), 2016, pp. 67–77.

  The work on authentication scheme features presented in this publication informed the work on requirements presented in this part.

1.3.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

This part of the thesis is based on the following publications:

- **P. Mayer** and M. Volkamer, “Secure and Efficient Key Derivation in Portfolio Authentication Schemes Using Blakley Secret Sharing”, Annual Computer Security Applications Conference (ACSAC), 2015, pp. 431–440.
- **P. Mayer** and M. Volkamer, “Poster: Secure Storage of Masked Passwords”, European Symposium on Security and Privacy Posters (Euro S&P Posters), 2017.
- A. Gutmann, K. Renaud, J. Maguire, **P. Mayer**, M. Volkamer, K. Matsuura, and J. Muller-Quade, “ZeTA-Zero-Trust Authentication: Relying on Innate Human Ability, Not Technology”, European Symposium on Security and Privacy (Euro S&P), 2016, pp. 357–371.

No other publications contributed to this part of the thesis.

2 Background

The goal of this chapter is to provide a broad overview of the required background knowledge and terminology surrounding user authentication, as needed for the understanding of the subsequent chapters. First, the different types of user authentication are described (section 2.1) and the phases involved in a general user authentication procedure are presented (section 2.2). Then, an overview of attacks on user authentication is provided (section 2.3, needed in particular for part I). Thereafter, the background on awareness materials is described, outlining the status quo and design considerations (section 2.4, needed in particular for part I). Last but not least, the portfolio authentication technique is described (section 2.5, needed in particular for part III).

2.1 Types of User Authentication

User authentication serves the purpose of controlling access to protected resources, i.e. grant access to the resources to authorised users and prevent access to the resources from unauthorised users [182]. There are three general types of user authentication [168], each based on a different characteristic of the user [64]:

- Something the user knows (knowledge-based authentication or memometric authentication)
- Something the user possesses (token-based authentication)
- Something the user is (biometric authentication)

In the following, each of the basic working principles and properties of these three types of user authentication will be described[1]. Also, any of these three types can be combined to harness the properties of multiple types.

2.1.1 Knowledge-based Authentication

In knowledge-based authentication, the user must prove knowledge of an authentication secret, which is generally referred to as a password [22]. Knowledge-based authentication schemes can be categorised along two dimensions: the type of data the password is comprised of and the way the user's memory is strained during the authentication procedure.

The most prevalent type of data used for passwords are *textual characters*. While text passwords were originally conceived only as a measure to track computation time on mainframe systems, it has evolved into the dominant authentication scheme [98]. However, other types of data have been used for passwords as well. The most promising alternative to textual characters are *graphical elements* [20]. The motivation behind using graphical elements in passwords is the possibility to exploit the vast human capacity to store and

[1] The explanation of *knowledge-based authentication* will be more detailed than those of token-based authentication and biometric authentication, since it represents the focus of this thesis.

process visual information, the so-called pictorial superiority effect [151]. Paivio's dual-coding theory [156] explains that visual information is processed and stored differently in the human brain compared to abstract information such as text. This difference in encoding leads to a higher probability of memory imprinting and therefore to a higher memorability of visual information. The idea to use graphical elements for passwords was first documented in 1992 [6], but a plethora of schemes aiming to exploit the superior memorability have been proposed in the literature since, displaying different usability and security properties [20, 141, 198].

With respect to the the strain put on the user's memory when authenticating, three categories of authentication schemes are described in the literature [20]: (a) free recall-based schemes, (b) cued recall-based schemes, and (c) recognition-based schemes. *Free recall-based schemes* are the traditional implementation of knowledge-based authentication: the user has to freely recall the complete secret needed for authentication and enter it in a predefined blank field (e.g. empty text field for text passwords, a blank canvas for graphical authentication schemes, etc.). Such free recall is a cognitively demanding task. Therefore, memorability of free recall-based schemes is often impaired [148, 232]. The ubiquitous text password is the canonical example of an authentication scheme falling into this category. *Cued recall-based schemes* try to overcome this deficiency in memorability by providing additional information as cues to help users recall their password during authentication. Analogously to the actual password, cues can be of different data types, such as textual characters (e.g. a hint towards the password) or graphical elements (e.g. a rebus). Examples of cued recall-based authentication schemes are the rebus passwords by King [124] and the cued click points scheme by Chiasson et al. [45]. The best option in terms of memorability, however, are *recognition-based authentication schemes*. Recognition is a cognitively much easier task than free or cued recall [148, 206]. Recognition-based authentication schemes rely on a different mechanism than free recall-based schemes and cued recall-based schemes: instead of recalling the elements of the password, users have to decide whether presented information is part of the password. Usually the password is randomly assigned by the system for recognition-based schemes and presented alongside distractor information in a challenge-response fashion. A simple example of a recognition-based scheme would be to assign a random string as traditional text password to the user during the enrolment phase. During the authentication phase this password could be displayed alongside other random textual strings. The task for the user would then be to point out her/his password among the choices. This scheme is obviously insecure beyond practicality and serves only illustration purposes. However, multiple different recognition-based schemes have been proposed and studied with promising results using both, graphical passwords (e.g. [60, 100]) and (with slightly worse results in terms of memorability) text passwords (e.g. [230]). In particular, combining the memorability advantages of recognition-based authentication and graphical authentication schemes results in highly memorable passwords. There are numerous studies which found that graphical recognition-based schemes offer higher success rates and lower reset rates when compared to other knowledge-based authentication schemes (e.g. [32, 60, 141]). The most prominent example of graphical recognition-based authentication (which will also be revisited in chapter 8) is the Passfaces scheme presented in [167] and its derivatives found in the literature. Hlywa et al. [100] evaluated three different Passfaces-style schemes. The password in these schemes is a set of images and usually randomly assigned to the user during enrolment. Each of the images in the password is part of a larger group. During authentication, multiple grids of images are presented to the user, one after another, where each grid displays all images in one of the groups. The user has to point out the one image that is part of their password among the other images (so-called distractors). Figure 2.1 shows a grid resembling one of their schemes.

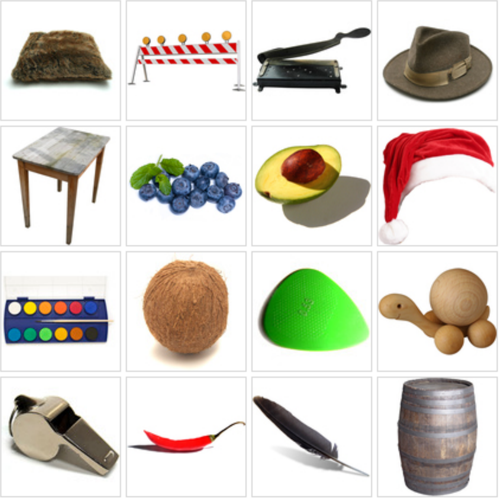


Figure 2.1: The interface of a recognition-based graphical authentication scheme resembling the interface[2] as used in [100]. The scheme shows multiple such grids one after the other. The task for users is to point out the one image that is part of their password.

2.1.2 Token-based Authentication

In token-based authentication, the user must prove possession of a certain object. Any object used for authentication purposes is typically referred to as *token* [168]. Tokens are usually categorised with respect to three characteristics [164]: whether they need to be plugged in, whether a specialised reader is required, and whether they are active devices.

One of the typical examples of tokens which have to be plugged in during the authentication procedure are smart cards. Smart cards are among the most widely deployed authentication tokens used, e.g. in banking [118], public transport [162], and physical access [231]. Other tokens which need to be plugged in are USB tokens (such as the Yubikey [52]).

In contrast, tokens that do not need to be plugged in are for example those where users have to transcribe a code from the token to the device. Such tokens come in electronic form (e.g. the RSA SecurID [65]) and in paper form (e.g. TAN lists for online banking [26]).

The second characteristic, namely whether a specialised reader is required, only applies to tokens which need to be plugged in [164]. Requiring a specialised reader which is not widely available in commodity consumer hardware (e.g. for smart cards) might hinder adoption as opposed to tokens without that requirement (e.g. USB tokens).

The third characteristic, namely whether the token is an active device, is determined by the type of storage and cryptographic functionality provided by the token [164]. Thereby, tokens relying on passive storage such as smart cards using magnetic strips are classified as passive. In contrast, tokens offering the possibility to perform computations during the authentication procedure (such as cryptographic functions) are classified as active.

[2] The interface was recreated using images from the same source (www.freeimages.com). All images © Getty Images.

2.1.3 Biometric Authentication

In biometric authentication, the user must prove that they exhibit a certain biometric characteristic. Biometric authentication schemes form two distinct categories: physiological biometrics and behavioural biometrics [51, 168].

Physiological biometrics represent static biometric information which is stable over time [168]. In physiological biometric authentication schemes, the static biometric information is captured using a sensor and matched to a capture of the same biometric stored during enrolment. The most widely applied example of physiological biometrics are fingerprints. In the last years they have become especially popular on mobile devices, such as TouchID [9] on the iPhone and iPad. However, a multitude of other physiological biometrics exist, such as iris [185] or retina [99] scans, palm vein recognition [238], face recognition [8], and many more [51, 168].

Behavioural biometrics represent the user's usage and activity patterns, e.g. mouse movements [114], keyboard typing characteristics [145], or gait [30]. Consequently, they can be captured only while the user is performing certain actions and exhibiting the necessary usage and activity patterns, e.g. using the mouse to capture the mouse usage patterns. Therefore, if the user is not performing the required actions anyways, the authentication procedure might take longer in order to collect enough data to judge whether the usage or activity pattern the user exhibits can be matched to the stored pattern for verification.

A problem inherent to all types of biometric authentication is that the sensors used to capture the biometric characteristic of the user have variances that might cause the characteristic of a legitimate user to be misclassified as illegitimate (a so-called false negative) or the characteristic of an illegitimate user might be classified as legitimate (a so-called false positive). For example Apple's FaceID system has gained attention after its deployment due to false positives [61].

2.2 The User Access Control Procedure

Authentication is part of the procedure to grant users access to protected resources. Thereby, the same overall procedure applies to all knowledge-based authentication schemes. This procedure comprises multiple phases, each fulfilling a distinct function (one of which is the actual authentication step). In the following, the function and scope of all phases are explained. Figure 2.2 gives an overview of all phases.

2.2.1 Enrolment

Before a user can authenticate using a certain authentication scheme, they have to be enrolled. The operation associated with this phase is the *creation of the verification information.* This phase ideally has to be

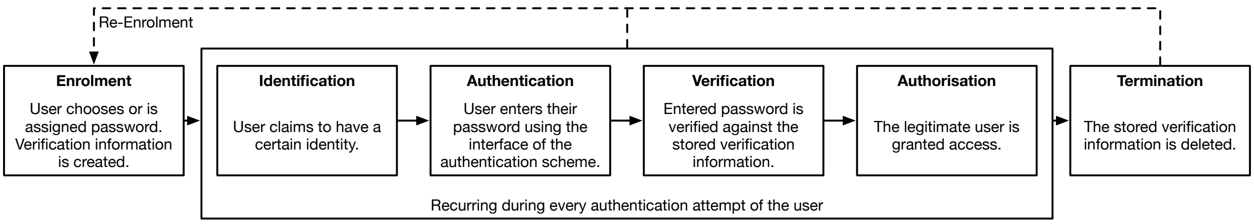

Re-Enrolment
Enrolment
User chooses or is assigned password. Verification information is created.
Identification
User claims to have a certain identity.
Authentication
User enters their password using the interface of the authentication scheme.
Verification
Entered password is verified against the stored verification information.
Authorisation
The legitimate user is granted access.
Termination
The stored verification information is deleted.
Recurring during every authentication attempt of the user

Figure 2.2: Overview of all phases comprised in a typical knowledge-based user access control procedure.

completed only once, but re-enrolment may be necessary in case a password is lost or stolen. In the traditional text password setting, this phase refers to the selection of a password by the user (in case the password is not assigned by the system), hashing the password, and storing the salted hash together with the user name for verification.

2.2.2 Identification

The identification phase is the first phase recurring whenever the user wants to access a protected resource. In this phase the user *claims to have a certain identity.* Traditionally the action required from the user in this phase is to supply the system with the user name. In this phase the password is not entered yet.

2.2.3 Authentication

In the authentication phase the user *enters their password* using the authentication scheme. In the traditional text password setting the respective user action is entering the password in a text field and submitting it to the system.

2.2.4 Verification

In the verification phase the authentication secret entered by the user during the authentication phase is *verified* against the verification information which was stored during the enrolment phase. In the traditional text password setting this refers to hashing the password supplied by the user during the authentication phase and comparing the calculated hash to the hash stored during the enrolment phase.

2.2.5 Authorisation

The fourth recurring phase is the authorisation phase. This phase is only reached when the verification was successful and determines what the authorised user is allowed to do. In this phase, *the legitimate user is being granted access* to the protected resource.

2.2.6 Termination

The termination phase corresponds to *deleting the verification information stored during enrolment.* The user will not be able to authenticate afterwards. This phase can usually occur only once, unless the user re-enrols.

2.3 Attacks on Passwords

Attacks on passwords[3] can be either capture attacks (i.e. the goal is to capture the password by recording or observing it in the clear) or guessing attacks (i.e. the goal is to guess the password utilising some form of

[3] Due to this thesis' focus on knowledge-based authentication this section uses the term *password* instead of *authentication secret*.

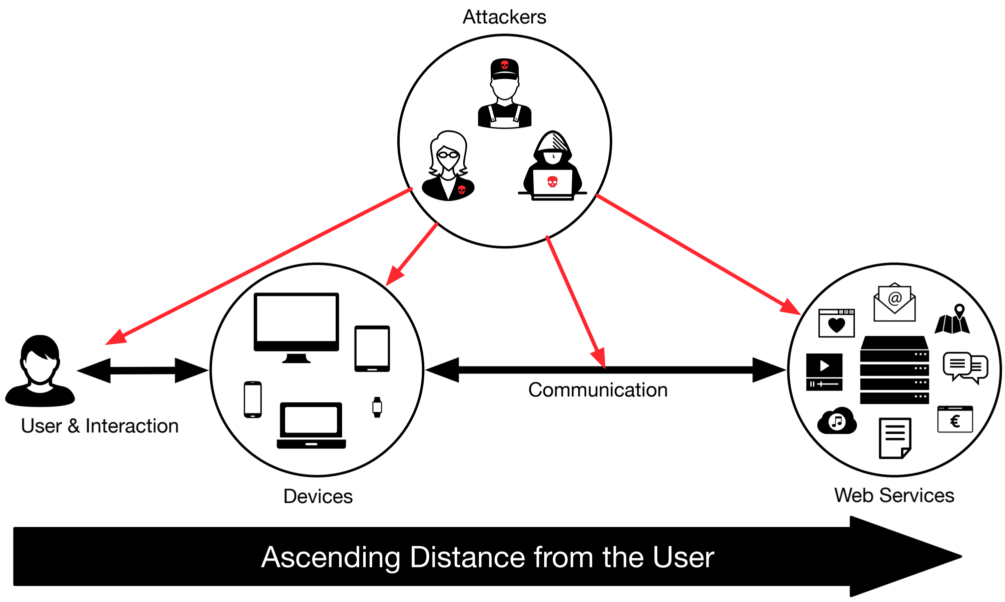


Figure 2.3: The attacks are ordered in ascending distance from the user.

oracle). The strength of passwords (i.e. their character composition and length) is thereby relevant only for guessing attacks, since capture attacks will always target the passwords in the clear and therefore capture all passwords regardless of character composition and length. Attacks are also classified regarding the number of accounts they impact [74] as full compromise (compromising potentially all accounts of the user), group compromise (compromising a group of accounts protected by the same password)[4], or single compromise (compromising a single account without acquiring the password, e.g. through session hijacking).

In the remainder of this section, an overview of the possible attacks on passwords and respective defences is given. The selection of attacks is based on those represented by the 11 different security benefits of authentication schemes as defined in the framework for comparative evaluations of authentication schemes[5] by Bonneau et al. [26]. Since the framework of Bonneau et al. [26] does not specify a systematic order for the attacks, such an order is introduced here: the attacks are ordered in “ascending distance from the user”, i.e. first attacks on the users themselves, then on the users’ devices, then the communication between the devices and remote services, etc. (see figure 2.3). Table 2.1 gives a birds-eye overview of the attacks on user accounts with the respective classification as full, group, or single attack and the corresponding benefits from [26]. The explanation of the *shoulder-surfing* attack will be more detailed than those of the other attacks, since it represents the basis of parts II and III of this thesis.

2.3.1 Attacks targeted directly at the user and the interaction with their devices

The attacks “closest to the user” are those targeted the user themselves (e.g. through her/his behaviour) and those targeted at the interaction of the user with her/his devices (e.g. observation of the user).

[4] Note that it is of course possible to achieve a *full* compromise in case all passwords can incrementally be obtained through *group* compromises. Also the group can – in the worst case for the attacker – be of size 1.

[5] Relying on this selection implies that firstly, only full and group attacks will be considered, i.e. those attacks whose goal it is to acquire the actual password. This excludes in particular the technique of exploiting a weak reset mechanism. While this is sometimes regarded as an attack, it is in essence the application of any attack presented in this section to the reset mechanism (which is just another authentication scheme), e.g. *targeted guessing* in the case of “personal security questions” [25]. Secondly, only specific attacks (e.g. phishing) and not categories of attacks (e.g. social engineering) are described.

Table 2.1: Overview of the attacks on user accounts with the respective classification as full, group, or single attack and the corresponding counteracting benefits from [26].

Attack	**Class**	**Counteracting Benefits**
Attacks targeted directly at the user and the interaction with their devices		
Phishing	Full/Group	Resilient-to-Phishing
Theft of an Insecurely Stored Physical Copy of the Password	Full/Group	Resilient-to-Theft
Shoulder-surfing	Group	Resilient-to-Physical-Observation
Attacks targeting the user's device as well as remote services		
Compromising the User's Device	Full	Resilient-to-Internal-Observation
Targeted Guessing	Group	Resilient-to-Targeted-Impersonation
Untargeted Guessing	Group	Resilient-to-Throttled-Guessing
Guessing After System Break-In	Group	Resilient-to-Unthrottled-Guessing Resilient-to-Leaks-from-Other-Verifiers Resilient-to-Trusted-Third-Party
Theft of a Digital Copy of the Password	Group	Resilient-to-Theft Resilient-to-Internal-Observation
Attacks targeted at the communication between the user's devices and remote services		
Compromising the network traffic between the users' devices and the remote services	Group	Resilient-to-Internal-Observation

Phishing

The first attack which targets the user directly is sending phishing messages i.e. messages using social engineering techniques or technical subterfuge [7] to lure an unsuspecting user into (a) clicking malicious links that would lead her/him onto a malicious website, or (b) perform actions in the interest of the attacker, e.g. transferring money to their bank account.

As of 2019, this type of attack is still on the rise with more malicious websites being detected in the first quarter of 2019 than in the third and fourth quarter of 2018 [7]. This is also evidenced by successful phishing attacks, such as the ones at the large Internet companies Google and Facebook [105] or even at the security company RSA [170]. Due to the elaborate efforts of the attackers, a combination of technological measures (such as described in [189,216]) and awareness materials (such as described in [128,152]) is advised to thwart attacks based on phishing messages.

An attack using phishing messages is traditionally classified as group compromise, when only one password is entered on a phishing website. However, it must be classified as full compromise, when the phishing attack compromises the user's password of a password manager which is accessible online (e.g. through a web interface).

Theft of an Insecurely Stored Physical Copy of the Password

The second attack, stealing an insecurely stored note of a password, relates to the problem of password memorability and a potentially insecure common coping strategy of users [194]: writing passwords down and not necessarily storing them securely. Such insecurely stored notes can easily be stolen by attackers with physical access. However, even when physical access is not possible, passwords can still leak, e.g. through inadvertent leaks in press images or video footage [146]. The presence of such notes is widely documented and ranges up to cases involving passwords of military systems which were leaked through press photos [48]. Therefore, it is widely given advice to store any written down password in a secure location [183]. This attack is classified as group compromise, since it affects only the password which was written down. However, if the attacker can obtain a password logbook [126] containing all of a user's passwords, then this attack is classified as full compromise.

Shoulder-surfing

In shoulder-surfing attacks, the attacker tries to capture the password by observing the user while they authenticate. Such an observation can be performed by the attacker themselves, but also with technical equipment such as cameras [135] or even after the fact using analyses of smudges on touchscreens [111] or thermal residue [1]. It has been identified as viable attack especially in public spaces [215] and shared spaces [169] and has been studied in particular with respect to mobile authentication (e.g. [12, 59, 111, 121]). Shoulder-surfing attacks are generally classified as *group* compromise attacks.

Due to the breadth of approaches to this attack, Wiese and Roth [227] propose to differentiate four types of attackers: Single Recording (the attacker has access to a small number of recorded authentication procedures), Multiple Recording (the attacker has access to a huge number of recorded authentication procedures), Opportunistic Observer (the attacker can observe a small number of authentication procedures), and Insider Observer (the attacker can observe a huge amount of authentication procedures). In the first two categories (Single Recording and Multiple Recording) the human ability (e.g. memory retention) plays a subordinate role, as the password entries are recorded and can be played back and paused at will. In the second two categories (Opportunistic Observer and Insider Observer) the attackers observe the whole process and try to remember the most important details. Afterwards they depend on their memory retention to try to log in based on the information they observe.

To counteract shoulder-surfing threats, several techniques can be employed. Harbach et al. [94] propose four categories of such techniques: (1) using covert channels to the user, (2) obfuscation of the user's input through distractors, (3) using indirect input, and (4) using additional biometric layers. In addition, DeAngeli et al. [55] propose (5) the portfolio authentication technique, where only a (random) part of the password is needed in every authentication attempt.

2.3.2 Attacks targeted at the user's devices

When moving one step further from the user and their interaction with devices, attacks may target the user's devices (but not the interaction with them).

Compromising the User's Device

A user's device can be compromised by attackers that have physical access to the machine or using malware. Opportunities for physical access may arise when devices are unlocked and unsupervised. In such a case, the attacker might be able to either access all passwords directly or by installing malware [184, 201]. Note that both is potentially possible even when a computer is locked: in case the hard drive is not encrypted and given enough time, an attacker could access the hard drive by means of another operating system and extract the data of interest or place malware on the hard drive such that it is installed on the next startup. When physical access is not possible, attackers must resort to other means, such as (a) tricking users into installing malware or enabling remote access by sending malicious attachments to messages or by tricking them to plug-in a USB stick or (b) exploiting unpatched vulnerabilities in software (e.g. operating system) on the user's device. In particular, the latter can also used for automated attacks using any self-replicating malware such as viruses or worms [201]. Compromising a user's device is classified as full compromise, due to the wide ranging access an attacker achieves with this attack.

Targeted Guessing

In targeted guessing attacks, the goal of the attacker is to compromise the account of one specific user. To that end, the attacker leverages personal knowledge about the victim, trying to guess passwords which this specific user would be likely to choose [26]. While there do not seem to be any studies establishing how successful acquaintances can be in guessing a user's password, other knowledge-based authentication schemes like personal knowledge questions are highly problematic in this regard [25]. Targeted guessing attacks are classified as group compromise.

Untargeted Guessing

In untargeted guessing, the attacker uses a more generic approach and uses as guesses for the password those which are the most likely across all users. The most relevant strategy for untargeted guessing is the so-called trawling [27], where attackers try the most frequently chosen passwords in an attempt to access as many accounts as possible until the lock-out mechanisms of the device takes effect. As evidenced by the fact that year after year the same passwords are found the be the most frequently chosen ones, this attack can be highly effective. Untargeted guessing attacks are classified as group compromise.

Guessing After System Break-In

The most powerful guessing attack is offline guessing after a break-in, i.e. when the password hashes were stolen by the attackers from the device. After e.g. having stolen hashes, the number of guesses an attacker can try is only bound by the resources available to them. For such offline guessing attacks, the security of the password is not only determined by its composition and length, but also by the hash that is used to create the verification information: the weaker the hash, the lower the security. Offline guessing usually employs a combination of brute-force attacks (i.e. trying all possible passwords up to the length where this attack becomes unfeasible) and dictionary attacks (i.e. trying as guesses all entries from a dictionary of likely passwords) is used [210]. Software to mount unthrottled automated guessing attacks on passwords is widely available. Such attacks are classified as group compromise.

Theft of a Digital Copy of the Password

The theft of a digital note of passwords represents the final attack targeted at the user's devices. Some users who are overwhelmed by the number of passwords they have to handle, cope by using unencrypted notes of passwords (i.e. text files) which can be stolen off their devices [194]. Even more dangerous are these when synchronised using cloud services (cf. section 2.3.4). A user might even have such notes on their device unintentionally, when a password manager does not require the user to set a master password or other applications (such as email clients) store the passwords of the accounts used in the application in the clear. Additionally, even when digital notes of passwords are encrypted, they might be at risk, if the key used for the encryption is not strong enough (e.g. derived from a weak password which can be guessed). The theft of a digital note is classified as a group compromise. However, when a (unencrypted or encrypted with weak key) vault of a password manager is stolen, this attack might need to be classified as full compromise.

2.3.3 Attacks targeted at the communication between the user's devices and remote services

Moving anther step away from the user, attackers have the possibility to try and compromise the communication between the user's devices and remote services used by that user.

Compromising the network traffic between the users' devices and the remote services

Compromising the network traffic between the users' devices and the remote services includes the interception and analysis of both, encrypted and unencrypted traffic. The use of unencrypted communication can be especially dangerous and facilitate easily automated interception of passwords or secondary authentication information (such as session cookies) [41]. Therefore, the usage of encrypted communication channels should be preferred. However, even encrypted traffic can be targeted using man-in-the-middle attacks [102, 104]. Thereby, rogue or compromised certificate authorities can represent the most dangerous opportunity to launch man-in-the-middle attacks, as evidence by the compromise of the Dutch certificate authority DigiNotar, which first lead to illicitly issued certificates and subsequently to the certificate authority's removal as root CA from all major browsers [3]. In such cases, users will be unable to recognise an illegitimate certificate, since it is virtually indistinguishable to a legitimate one. Compromising the network traffic between the users' devices and remote services is classified as group compromise.

2.3.4 Attacks targeted at remote services

The attacks farthest away from the user, while being classified as group compromises target the user's devices and remote services. In essence, all attacks described targeting the user's devices also apply in this category: the servers of a remote service can be compromised, passwords guessed using either the service's authentication interface (e.g. login page or API) or after a compromise the hashes stored on the servers, and if electronic notes of a password are stored on a remote service's servers these can be stolen off the servers.

2.4 Information Security Awareness Materials

Information security awareness materials [93] are an important tool to inform users about the attacks outlined in the last section and thereby to keep users and organisations secure. As shown by Hänsch and Benenson [93], even experts use different definitions of what the term *awareness* actually encompasses. Hänsch and Benenson identify three dimensions which are commonly associated with awareness materials: (1) perception, i.e. users recognise a threat after exposure to the materials which represents awareness in the original sense of the word as defined by the oxford dictionary; (2) protection, i.e. users know information security attacks and defences after exposure to the materials; (3) behaviour, i.e. users will always act securely in their daily lives. In the remainder of this thesis awareness materials target the only the first two aspects, i.e. perception and protection[6].

[6] This decision was made in accordance with the goals of the project "KMU AWARE" (`https://secuso.org/kmu-aware`), in which the research pertaining to awareness materials presented in this thesis was conducted

2.4.1 Relevance and Status-Quo of Information Security Awareness Materials

Lin and Kunnathur [130] developed a theory of end user information security competence based a synthesis of information security literature. Their theory comprises three dimensions: “ethics and perception”, “knowledge and skills”, and “behaviour”. They highlight information security awareness as vital part of the knowledge and skills dimension. This is further evidenced by the many studies (e.g. [35, 66, 115, 155, 174]) having identified the importance of advice in awareness materials. Also, the literature review of Lebek et al. [129] identifies information security awareness as antecedent of attitude toward secure behaviour.

Therefore, institutions such as NIST [228] recommend organisations to distribute awareness materials among their employees. Some industry standards even require it from organisations aiming to be compliant with it [160]. However, it has been found, that existing password advice often contradicts current research [149,236]. Zhang-Kennedy et al. [236] present a review of established advice for general-purpose authentication on the web. They conclude that many existing rules (e.g. to change passwords frequently) do not represent the current state of art and propose a new set of password rules to give as advice to users. Murray and Malone [149] present an analysis of actual password advice given by different organisations on the Internet and conclude that the majority of advice contradicts the current state of research.

2.4.2 Design of Information Security Awareness Materials

Due to the importance of information security awareness materials, the literature on the topic highlights several important aspects of their development. Bada and Sasse [15] analysed which aspects are involved in the success and failure of security awareness campaigns. They identified as precedents to successful awareness materials the relevancy of the materials for the users and that the advice in the materials is effective. Their recommendation is that all awareness materials should pay great attention to these aspects. Tsohou et al. [204] add to these recommendations in their review of literature on cognitive and cultural biases influencing users information security perceptions and behaviours by deriving three recommendations for the development of information security awareness materials: (1) using positive stimuli and relative frequencies to overcome affect biases, (2) the design of the material must accommodate for the fact that users tend to rely on the first piece of information they are provided with, and (3) the material should emphasise immediate consequences. Another important concept for the development of awareness materials is intellectual need (also called problem-solution ordering) as described by Fuller et al. [78]. It describes the fact that learners are more motivated and effective at acquiring knowledge, when presented with the problem before the solution is explained to them.

Furthermore, using expert feedback and behaviour has been identified as an important aspect in the development of advice for lay-users. Ion et al. [109] compared expert to lay-user information security behaviour to collect useful information security advice for lay-users. Their study focused on information security behaviour in general. They found that expert and lay-user behaviour differ and summarise their findings by saying that "some promising security advice emerges: (1) install software updates, (2) use a password manager, and (3) use two-factor authentication for online accounts." Likewise, Stobert and Biddle [195] conducted interviews with information security experts specifically in the context of passwords and user accounts. They find that lay-users are in need of consistent strategies to better protect themselves and that the adoption of password managers could help lay-users to manage their passwords more securely.

Focusing on the creation of concrete password security awareness materials, Zhang-Kennedy et al. [235] developed three infographic posters and an online educational comic. They evaluated the posters against a text condition - “the Wikipedia description of how password cracking works” - and found that the posters with explanatory graphics were more effective in the knowledge transfer than the textual Wikipedia description, underlining the importance of graphical elements in awareness materials.

2.5 Portfolio User Authentication

As explained in section 2.3.1, portfolio authentication is a technique to decrease the threat of capturing attacks, such as shoulder-surfing. However, it can be used to mitigate opportunistic attacks in non-shoulder-surfing scenarios as well, e.g. the interception of passwords in network traffic. To achieve this, portfolio authentication schemes regard the password as a set of elements. These elements can have any form (e.g. textual characters, textual strings, images, electronic certificates, etc.). A password P of length n is thus represented as

$$P = \{e_1, e_2, \ldots, e_n\}.$$

During each authentication attempt only a random subset $P' \subseteq P$ of these elements has to be entered by the user. We borrow from the nomenclature of secret sharing and denote such a subset P' as authorised if it has at least t elements, where the parameter t is set by each authentication scheme depending on the desired security properties. In the remainder of this thesis, the magnitude by which the password is larger than the authorised subsets are referred to as portfolio overhead (denoted as o). It is represented as a fraction of the form

$$o = \frac{|P|}{|P'|} = \frac{n}{t}.$$

Note that usually $|P'|$ serves as starting point when determining o and the full password is then chosen accordingly (either by the system or by the user while respective policies guide the user’s choice), since the strength of the authentication scheme against guessing attacks is dependent on $|P'|$ rather than $|P|$.

Portfolio authentication is most useful to prevent opportunistic observation (cf. section 2.3.1) in public spaces or while in the presence of friends or co-workers. It thwarts opportunistic shoulder-surfers that do not observe the user for a longer period of time (i.e. have a large number of observation attempts). An attacker who can capture multiple (possibly any arbitrary number of) authentication attempts or who uses a recording device might still be able to break the portfolio authentication scheme, depending on the actual authentication scheme’s design.

The most common application of portfolio authentication are challenge-response schemes. Thus, the elements e_i are usually challenge-response pairs. In particular, graphical recognition-based authentication schemes have been used. Dhamija and Perrig first introduced the portfolio term for graphical recognition-based authentication schemes [60]. DeAngeli et al. proposed the portfolio approach for graphical recognition-based passwords as high security setting [55]. In their scheme P' is a random subset of the password images (subset size: 4; password length: 6) and is chosen as challenge set for each authentication attempt. Dunphy et al. first proposed portfolio authentication as a measure to mitigate shoulder-surfing risks on mobile phones in public places [62]. As result of their study they report a temporary resistance against shoulder-surfing attacks (on average 5 to 7 observations depending on study condition).

2.6 Cryptographic Secret Sharing

The (t, n)-*threshold verification scheme* proposed in part III of this thesis uses cryptographic secret sharing to address the challenge of secure and efficient storage of passwords in portfolio authentication. In this section the basic principles of secret sharing are explained.

Secret sharing describes a family of cryptographic protocols which can be used to distribute sensitive information (the common secret) among several parties such that only by collaboration of a certain number of these parties the secret can be reconstructed. Such protocols can be denoted (t, n)-threshold secret sharing, where n denotes the number of parties among which the secret is shared and t the number of parties which are required to cooperate in order to reconstruct the secret. The (t, n)-threshold secret sharing protocols consist of two phases: (1) the dealing phase in which each of the n parties is assigned a so-called *share* of the secret and (2) the combination phase in which t or more parties can collaborate to reconstruct the common secret using their shares. Figure 2.4 illustrates these two phases.

An important property in the context of secret sharing is whether the protocol is *perfect*. A secret sharing protocol is perfect if the knowledge of any number of shares smaller than t does not reveal any information about the common secret.

The (t, n)-threshold secret sharing protocols are usually based on finite field (also Galois field or GF) arithmetic. A thorough introduction to finite fields is beyond the scope of this thesis. However, in short, a finite field is a finite set of elements on which multiplication, addition, subtraction, and division are defined as operations. A common way to construct finite fields is using the set of integers modulo p, where p is prime. In the remainder of this thesis, such finite fields of integers modulo p are referred to using the common notation $GF(p)$.

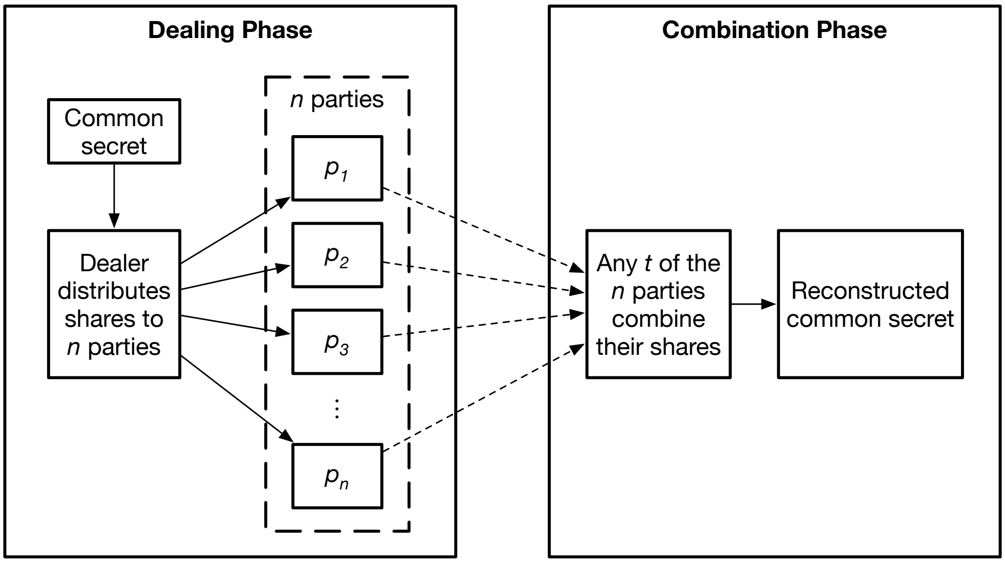

Dealing Phase
Common secret
Dealer distributes shares to n parties
n parties
p1
p2
p3
⋮
pn
Combination Phase
Any t of the n parties combine their shares
Reconstructed common secret

Figure 2.4: Overview of the two phases of (t, n)-threshold secret sharing protocols.

Part I

Effective Password Security Awareness Materials

3 Identifying Misconceptions about Password Security

Many users face problems when choosing, handling, or remembering their text passwords [107, 194, 195]. As a result, users have developed a variety of coping strategies. Some of these coping strategies are beneficial, such as using a master-password protected password manager in order to cope with remembering different passwords for all accounts. In contrast, some of these coping strategies represent insecure behaviour, since they are based on misconceptions about password security[1]. For example a common misconception might lead users to add a '!' to end of the password, because they believe this makes the password more secure [209]. Consequently, users unknowingly find themselves in situations, where they believe they have secure password practices, when in reality they do not: users could be unaware of their insecure password practices due to the prevalence of misconceptions. Clearing up these misconceptions is therefore an important precursor to password security awareness materials.

Misconceptions about password security appear frequently in the results of published literature, e.g. the lack of mental models representing automated attacks and defences against them [209]. However, there does currently not exist an overview of all the different misconceptions which have been reported in the literature. Knowing the prevalent misconceptions and addressing them with effective interventions is vital, when aiming to decrease insecure coping strategies and providing effective advice in awareness materials for lay-users.

Therefore, the overall goal of this chapter is to identify the misconceptions about password security reported in the literature and provide an overview as basis for the creation of effective password security awareness materials for lay-users as described in chapter 4. To that end a methodology for the systematic literature review was developed, including the relevant venues and imposing effective exclusion criteria (section 3.1). As result of the literature review, 23 different misconceptions about password security were identified (section 3.2). The identified set of misconceptions indicates that misconceptions exist in basically all aspects of password security. From the discussion (section 3.3) three aspects arise as next steps (section 3.3.3): assessing the actual prevalence of the identified misconceptions empirically, the development of interventions to decrease the prevalence of the misconceptions, and an investigation of interactions and correlations between the identified misconceptions. Section 3.4 concludes this chapter.

Contributions described in this chapter:

- Overview of the 23 misconceptions about password security reported in the literature, showing that there exist misconceptions with respect to a wide range of aspects of password security.

[1] In the remainder of this part, the term *password security* is used as abbreviation for security aspects affecting passwords and user accounts.

Parts of the results described in this chapter have been published in:

- P. Mayer and M. Volkamer, "Addressing Misconceptions About Password Security Effectively", Workshop on Socio-Technical Aspects in Security and Trust (STAST, co-located with ACSAC), 2017.

3.1 Methodology of the Literature Research

The literature review was conducted from February to May 2017. Only publications published in the last decade before the literature review (i.e. since 2007) were considered, in order to not include outdated findings[2]. Search terms were chosen after consultation with native English speaking experts in the field. As final set of search terms, the term "password" in conjunction with each of the terms "misconception", "misunderstanding", "misperception", "flawed perception", and "flawed understanding", one after the other (e.g. "password misconception") were used. Both terms, i.e. "password" and any one of the other search terms respectively, needed to be present in the publication. As sources for the publications, two approaches were used: (a) the five widely recognised databases of scientific literature in the computer science domain Sciencedirect, ACM, IEEEexplore, SpringerLink, and Scopus; and (b) additional conferences and journals known to publish relevant research on passwords, but not indexed by the aforementioned databases, i.e. Usenix Security Symposium, Symposium On Usable Privacy and Security, Usable Security Workshop, Trustworthy Interfaces for Passwords and Personal Information Workshop, Journal of Computer Security, International Journal of Information Security and Privacy, International Journal of Technology and Human Interaction, Human IT: Journal of information technology studies as a human science, MIS Quarterly, and Journal of Information Systems Security. Overall, 3777 publications meeting the search terms were found in the sources.

From the 3777 publications meeting the search terms, the body of relevant literature was narrowed down using additional criteria as described in the following. Publications not accessible due to a "paywall" (e.g. not licensed by the author's university's library) were excluded. For publications where the respective authors had published the same results multiple times (e.g. extended versions of conference papers in journals), only the latest most up-to-date publication was considered. Also, non-peer-reviewed publications (white-papers, technical reports, etc.) were excluded. To filter out the publications not explicitly dealing with misconceptions in field of passwords, the publications were manually screened based on title, abstract and if necessary a glance on the full text. Overall, 15 publications reporting on misconceptions met these additional criteria. To broaden the results, first a forward, then a backward search was performed, resulting in a final number of overall 20 relevant publications.

3.2 Identified Misconceptions

This section presents the results of the systematic literature review. First, the misconceptions about password security are presented. Then, two further more general information security misconceptions encountered during the literature review are presented, which concern the issues of (a) trust in software and (b) software updates.

3.2.1 Misconceptions about Password Security

In the 20 relevant publications found during the literature search, 23 misconceptions about password security were identified. In the following, each of the identified misconceptions is described.

[2] It must be acknowledged that this time frame is somewhat arbitrary. This is further discussed in section 3.3.

M1: The inclusion of numbers anywhere in passwords makes them automatically more secure

M1 was reported in five publications: [14,171,188,208,209]. The underlying problem with this misconception is that additional character classes (i.e. lowercase, uppercase, numbers, symbols) can make passwords more secure, but this is not automatically the case. Research has shown that (a) when users try to add additional character classes to their passwords, they tend to create very predictable passwords [208] and that (b) forcing users to put these characters in places where they contribute most to the guessing-resistance of the passwords decreases the usability of the created passwords [188].

M2: The inclusion of symbols anywhere in passwords makes them automatically more secure

M2 was reported in the same five publications as M1: [14,171,188,208,209]. The underlying problem is the same as for M1: users tend to put the chosen symbols in predictable places in the password.

M3: The inclusion of uppercase letters anywhere in passwords makes them automatically more secure

M3 was reported in two publications: [14,208]. The underlying problem is the same as for M1 and M2: users tend to put the uppercase letters in predictable places in the password (in particular at the beginning).

M4: Common substitutions (e.g. A $\rightarrow$ 4) make passwords more secure

M4 was reported in two publications: [208, 221]. The underlying problem is similar to M1, M2, and M3: users tend to use predictable substitutions in their efforts to make passwords more secure.

M5: A word from another language than the user's mother tongue is a secure password

M5 was identified in one publication: [195]. The underlying problem with this misconception is that attackers can easily build large dictionaries with words from several languages using e.g. the freely available wikipedia entries in nearly 300 languages. This is illustrated by the research literature on this topic: While Maoneke et al. [137] found that non-English passwords might improve the guessing resistance against attackers using standard dictionaries, investigations into the guessing security of e.g. Chinese [222] passwords and African[3] passwords found that adapted guessing strategies can yield high portions of guessed passwords.

M6: Reusing passwords is OK for secure passwords, but should be avoided for weak passwords

M6 is one of three misconceptions concerning the reuse of passwords, which is a common coping strategy of users [193]. It was reported in one publication: [223]. The underlying problem with this misconception is that users applying this misconception in their handling of passwords might end up with secure passwords

[3] Sibusiso Sishi, "An investigation of the effectiveness and security of passwords derived from African languages", talk at PasswordsCon 2018, `https://youtu.be/ZQYFp4fzpyE` (visited at: 2019-09-07).

at more websites, but at the same time increase the risk of cross-site attacks [110, 223] (e.g. when a website leaks passwords in the clear or the user loses a password in a phishing attack and that password is used by an attacker to break into an account at another website). In particular, this misconception does not include the real metric on which a decision to reuse a password should be based: whether reusing a password would allow an attacker to compromise additional valuable data [74, 150].

M7: Reusing passwords is OK for passwords that are entered more frequently

M7 could be identified in two publications: [154, 223]. The problem underlying this misconception is the same as with M6: it does not consider the real metric on which a decision to reuse a password should be made. Therefore, users might reuse passwords across accounts that give access to different valuable data.

M8: Reusing passwords is more secure behavior than writing them down

M8 was identified in one publication [95]. The underlying problem of this misconception is that while users can control where they store written down passwords (i.e. they can make sure it is stored securely), they have no way of knowing whether a web service is among the many not sufficiently protecting the users' data [28]. While storing passwords in the clear has been known for at least a decade to be a major information security risk [73], it is still a risk users face today as evidenced by breaches of online services revealing bad password storage practices (e.g. [125]). Users' passwords might be stored in the plain on a web service's servers, without them having any way of knowing. If a password is stored insecurely at a web service, this poses a threat to the user's accounts at other services where this password is reused.

M9: Notes of passwords do not need to be particularly protected

M9 regards the handling of paper notes as well as electronic notes [194]. It was reported in three publications: [132, 194, 195]. The problem underlying this misconception is that notes of passwords can be beneficial and have been recommended by security experts (e.g. [183]), but this advice always comes with the condition of secure storage of that note. Thereby, in particular the unprotected storage of cleartext passwords in the cloud (e.g. a text document in Dropbox or in note-taking services such as OneNote) poses a severe risk.

M10: Passwords have to be changed proactively on a regular basis

The existence of M10 among users was reported in one publication: [109]. This is no surprise, since it was a longstanding advice given by institutions like the US NIST [40]. The reasoning behind this advice was that regularly changing a password increases its security because: (a) even if an attacker gets hold of the password, changing it renders it worthless to attackers since they cannot login anymore; and (b) guessing a password that changes frequently is harder for an attacker since its represents a moving target. However, research has shown that both of these lines of thought are flawed and therefore exposed this advice as misconception.

Firstly, Zhang et al. [234] found that new passwords of users that were created under a password policy mandating regular password changes, were often derived from the previous one. Knowledge of a previous password allowed recovering the current passwords for the majority of user accounts in their study. Therefore, any obtained password is valuable to facilitate guessing attacks, even if it has been replaced with a new one.

Table 3.1: Overview of the access control and storage security mechanisms of major browser vendors in their Windows and macOS implementations. Inspected software versions: Windows 10, macOS 10.13.6, Internet Explorer 11.316.17763.0, Edge 44.17763.1.0, Firefox 65.0.1, Chrome 72.0.3626.121, Opera 58.0.3135.107, Safari 12.0.3

Browser	**Windows**	**macOS**
Internet Explorer	Passwords stored in Windows Credential Manager	N/A
Edge	Passwords stored in Windows Credential Manager	N/A
Firefox	Default none, optional master-password	Default none, optional master-password
Chrome	Encryption key stored in Windows Key Store	Encryption key stored in macOS Keychain
Opera	Encryption key stored in Windows Key Store	Encryption key stored in macOS Keychain
Safari	N/A	Passwords stored in macOS Keychain

Secondly, Chiasson et al. [44] formulated a mathematical model to investigate the advantage of regular password changes to protect against guessing attacks. They found that an attacker is only slightly impeded, even in cases where the password is changed much more frequently than the attacker can perform an exhaustive search of the password space. Therefore, they conclude that regular password changes offer “at best partial and minor” [44] security benefits.

Luckily this research has already made its way into the most recent versions of password advice from the US NIST [86], the British NCSC [150], and (very recently) the German BSI [38], who all discourage using mandatory password changes, unless a user account has actually been compromised. Unfortunately, this still bares potential for confusion among users, since not all standardising bodies have adapted their standards with respect to these findings (e.g. PCI-DSS [160]). Since it potentially applies to all of the user’s passwords, this misconception should be considered to be universal.

M11: Storing passwords in the browser does not mean one is using a password manager

M11 indicates that users perceive technologies that are essentially the same as different. It was reported in one publication: [209]. The underlying problem of this misconception is a potential lack of understanding, that the same security requirements apply to dedicated password managers and those integrated in browsers (e.g. setting a master-password in most situations). In particular, if passwords in a password manager are not stored encrypted, they could be easily obtained by an attacker. This holds for both, dedicated password managers and those integrated into browsers. Table 3.1 gives an overview of the security mechanisms used by major browser vendors to protect stored passwords. It becomes clear that no common strategy exists. Internet Explorer, Edge, and Safari are bound to one specific platform (i.e. Windows and macOS respectively) and use the password storage functionality of their platform. Chrome and Opera use the respective key management functionalities of the operating systems to store one encryption key, but store the passwords (encrypted with that key) in an SQLite database. Firefox completely forgoes the key storage capabilities of the platforms and stores the password by default without protection by a master password in an SQLite database, but allows setting a master password if the user so wishes. To clear up this misconception, communicating the general requirement of encrypted storage and how to achieve it (and not the type of password manager) is crucial.

M12: Keyboard patterns are secure passwords

M12 was reported in one publication: [208]. The problem underlying M12 is that while patterns on the keyboard might seem like random strings of characters, the security issues associated with using them for passwords are well documented. About 10-20% of passwords include keyboard patterns of length 3 or more [103]. The most common of these patterns are therefore commonplace in cracking dictionaries [210]. When using more sophisticated cracking methods, exploiting keyboard patterns can yield additional 15-20% cracked passwords [103]. At the same time, users generally overestimate the security of passwords based on keyboard patterns [208].

M13: Using dates of birth that are not the user's makes passwords more secure

M13 was reported in one publication: [209]. The security issues associated with using dates as passwords are well documented [212]: users rely on duplicating a year to create passwords of length 8, dates of holidays are popular choices as passwords, and common non-digits paired with passwords based on dates are single characters or written out months. Analogously to M12, this misconception is relevant in all attack scenarios where guessing attacks are viable.

M14: Attackers do not automate their attacks on passwords, but perform them by hand

M14 was reported in five publications: [82, 117, 136, 154, 209]. The underlying problem of this misconception is that users underestimate the scale of attacks that specialised tools allow. They seem to lack a suitable mental model with respect to automated attacks and how attackers try to guess passwords [208]. This misconception affects attack scenarios that can be automated well, such as mass phishing emails [152] or guessing attacks [210]. However, the field of automated attacks grows steadily as advances in artificial intelligence allow ever more attacks to be automated. Seymour and Tully demonstrated in their work on automated spear-phishing on the Twitter social network [113] that by using a combination of Markov models and neural networks it was possible for them to achieve click-through rates of at least 30%[4]. An overview of exemplary automation techniques for each of the attacks introduced in section 2.3 can be found in table A.1 in appendix A.

M15: All attackers are strangers which are (geographically) far away

M15 was found in three publications [82,117,208]. The problem underlying this misconception is that it limits the perception of possible attacks. However, there are many motives for an attack on a user's passwords or accounts, but not all of them relate to an unknown hacker which is geographically far away, e.g. someone impersonating cleaning personnel in order to get access to an office, co-workers who want that changes to a document in online storage cannot be traced back to them, or a nosy acquaintance who wants to spy on the communication with other users.

[4] The click-through rate of a phishing attack is the percentage of users clicking on links provided in phishing messages. Seymour and Tully observed click-through rates of up to 66% but attributed some of the "clicks" to bots crawling the social network and therefore were not counted.

M16: All attackers are people the users know

M16 represents the opposite assumption of M15. It was reported in two publications: [82, 171]. The underlying problem of this misconception is the same as for M15: a disregard of likely attacks just assuming a reverse point of view as described in M15. Neglected attacks include e.g. trawling attacks, where attackers target as many accounts as possible and the goal is not to get into one specific account, but in as many accounts as they can [27].

M17: Email is security-wise not an important service and therefore does not require a secure password

M17 was reported in two publications: [82, 209]. The underlying problem of this misconception is that insufficiently protected email accounts can compromise other accounts if passwords can be reset using links sent by mail. Thus, a compromise of the user's email account can cause a snowball effect of further compromises. Additionally, attackers can use compromised accounts to send out spam in the user's name.

M18: A SIM-PIN is sufficient to protect the data on a smartphone from unauthorized access

M18 was identified in one publication: [157]. Its underlying problem is that only a system PIN used to lock a device (or a comparable authentication mechanism such as fingerprint readers) protects the data on the respective device. A SIM-PIN can be easily circumvented by removing the SIM from the device. On many smartphones entering the SIM-PIN can be simply skipped, disabling the telephony functionality, but rendering the data on the device accessible.

M19: It is not necessary to set a password to lock the screen of unattended devices

M19 was reported in two publications: [95, 136]. The underlying problem is that attacks aiming at physical access to the user's devices are neglected (e.g. impersonation of help desk staff or cleaning personnel).

M20: The passwords protecting work accounts have lower security requirements because the IT staff is responsible for security

M20 was identified in one publication: [194]. The problem underlying this misconception is that users might perceive the security requirements of their work account differently than those of their private accounts, because of a false sense of security stemming from security measures put in place by the IT staff of the company. However, in reality the work account is exposed to at least the same threats, some of which only the user will be able to fend off in day to day use.

M21: The frequency of use of an account influences the security requirements of the password protecting it

M21 was reported in one publication: [194]. The underlying problem is that it disregards the aspects that actually influence the security requirements of an account (i.e. data accessible in the account and actions which can be performed with the account, e.g. contacts in an email account and the possibility to send phishing emails to all these contacts). Instead frequency of use is associated with importance of the account. This misconception should be assumed to be universal, since it potentially affects all accounts of a user.

M22: Bank accounts do not need to be protected with strong passwords if there is no money in them

M22 was identified in one publication: [209]. The problem underlying this misconception is that while some research reports that financial sites are of high value to users (e.g. [194]), some users voiced that if there was no money in their account, the security requirements would be low. Whether this misconception applies to a user depends on factors such as whether the account can be overdrawn and being left with debts (and the incurring interest) might be a concern or whether additional authorisation of any actions (such as TANs) is needed. Additionally, privacy can be a serious issue, if an attacker cannot perform any actions on the account, but still is able to view all transactions. Such information could be leveraged by the attacker for future attacks through e.g. phishing.

M23: It is possible to be too unimportant to be targeted and therefore to have to choose secure passwords

M23 was reported in three publications: [117,177,208]. The underlying problem of this misconception is that feeling too unimportant to be targeted might lead users to not concern themselves with prevalent attacks on passwords (such as untargeted guessing) and corresponding defences. Then in turn, by being unaware of such attacks users are more likely to choose less secure passwords and thereby render themselves more vulnerable to guessing attacks.

3.2.2 Further Misconceptions

During the literature review, two further misconceptions regarding information security topics were encountered, which will not be regarded in the remainder of this work. The first one states that security updates are not important (reported in [109]). It revolves around one advice deemed most important by security experts [109]: installing security updates. The second one concerns password managers. While using password managers is generally considered to be good advice [109, 195], there also exists a mistrust towards them (reported in [70, 109, 131]). Such a lack of trust is difficult to address with explanatory texts and is potentially affected by the specific password manager used (e.g. through brand recognition). One might try to convince lay-users by explaining to them that experts in the field trust password managers (as found in e.g. [195]). Yet, trust is not easily established by textual communication [29, 172]. Therefore, this lack of trust should not be treated as an issue which can be simply cleared up by knowledge transfer and will not be regarded in the remainder of this work. However, following up on this issue might be an interesting field for future investigations.

3.3 Discussion

This section discusses the results and the limitations of the systematic literature review presented in this chapter, as well as possible next steps to continue this line of work.

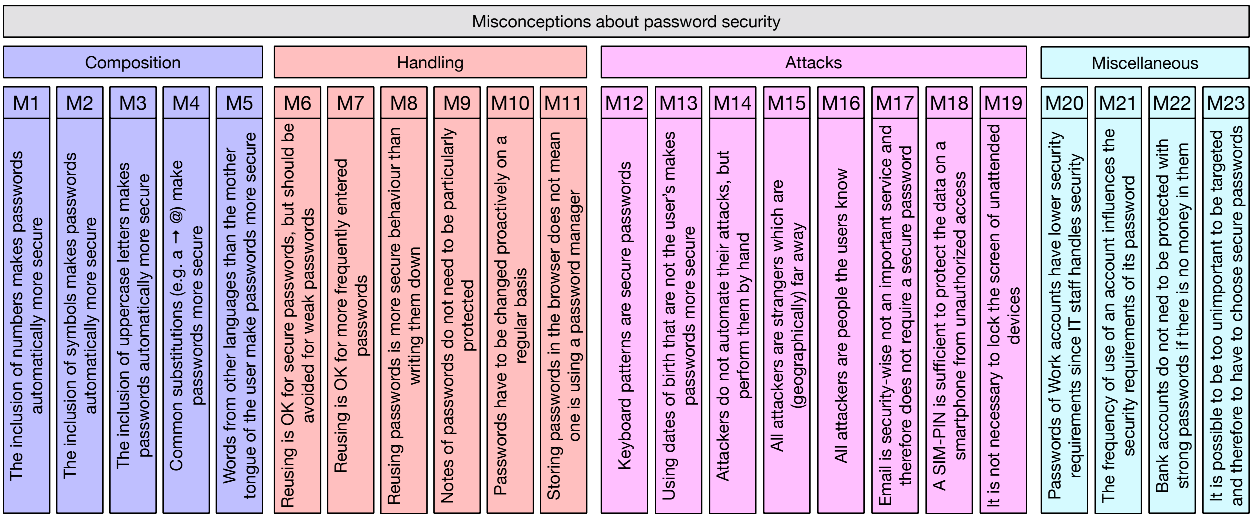

Misconceptions about password security
Composition
Handling
Attacks
Miscellaneous
M1
The inclusion of numbers makes passwords automatically more secure
M2
The inclusion of symbols makes passwords automatically more secure
M3
The inclusion of uppercase letters makes passwords automatically more secure
M4
Common substitutions (e.g. a → @) make passwords more secure
M5
Words from other languages than the mother tongue of the user make passwords more secure
M6
Reusing is OK for secure passwords, but should be avoided for weak passwords
M7
Reusing is OK for more frequently entered passwords
M8
Reusing passwords is more secure behaviour than writing them down
M9
Notes of passwords do not need to be particularly protected
M10
Passwords have to be changed proactively on a regular basis
M11
Storing passwords in the browser does not mean one is using a password manager
M12
Keyboard patterns are secure passwords
M13
Using dates of birth that are not the user's makes passwords more secure
M14
Attackers do not automate their attacks, but perform them by hand
M15
All attackers are strangers which are (geographically) far away
M16
All attackers are people the users know
M17
Email is security-wise not an important service and therefore does not require a secure password
M18
A SIM-PIN is sufficient to protect the data on a smartphone from unauthorized access
M19
It is not necessary to lock the screen of unattended devices
M20
Passwords of Work accounts have lower security requirements since IT staff handles security
M21
The frequency of use of an account influences the security requirements of its password
M22
Bank accounts do not ned to be protected with strong passwords if there is no money in them
M23
It is possible to be too unimportant to be targeted and therefore to have to choose secure passwords

Figure 3.1: Overview of all misconceptions about password security identified in the literature review.

3.3.1 Results

Overall, 23 misconceptions were identified in the published literature. They span a wide variety of aspects, from the inclusion of different character classes to the security trade-offs between reuse and writing passwords down. Thereby, the identified misconceptions can be loosely grouped into four broader categories: composition, handling, attacks, and miscellaneous. Figure 3.1 depicts all the identified misconceptions along those four categories.

Composition. This category comprises the misconceptions which regard the composition of passwords and therefore are most relevant with respect to guessing attacks. These misconceptions regard e.g. the effect of incorporating different character classes (i.e. letters, numbers, and symbols) into the password.

Handling. This category includes all misconceptions with respect to behaviour associated with the handling of passwords, such as reuse of passwords for multiple accounts. The range of attack scenarios these misconceptions are relevant for is broader than those of the composition category, ranging from reuse to taking notes or frequent changes of the passwords.

Attacks. This category contains all misconceptions which concern the perception of the attackers and their strategies, but do not directly relate to password composition or the handling of passwords. These misconception include e.g. potential groups of attackers and the degree to which attacks can be automated.

Miscellaneous. This category consists of further, more general, misconceptions, which were found in the literature to also apply to password security as well.

Also it became apparent that there might exist interactions between some of the identified misconceptions. For example, “Notes of passwords do not need to be particularly protected (M9)” and “It is not necessary to set a password to lock the screen of unattended devices (M19)” require the attacker to have physical access to the device or its surroundings. Therefore, their prevalence should be less likely to coincide with the prevalence of “All attackers are people the users know (M16)” than with the presence of “All attackers are strangers which are (geographically) far away (M15)". Likewise, if a user employs reuse of passwords as a coping strategy, they might base their behaviour on all three misconceptions concerning password reuse (i.e. "Reusing passwords is OK for secure passwords, but should be avoided for weak passwords (M6)", "Reusing passwords is OK for passwords that are entered more frequently (M7)", "Reusing passwords is more secure behavior than writing them down (M8)"). While the literature review does not provide any evidence regarding these links and interactions, an investigation of this aspect could yield valuable insights: knowing which misconceptions appear together might allow to clear them up more effectively together.

3.3.2 Limitations

Certain limitations apply to the methodology of this work. As is the case with every systematic literature review, a usual concern is the so-called "publication bias", i.e. the fact that positive and significant results are more likely being published than negative and non-significant results. While it is very likely, that this work was influenced by this type of bias, it can be argued that the impact is negligible in the context of this specific work. Negative results (i.e. the lack of any misconceptions) would not have contributed to the body of misconceptions identified in this literature review and therefore not changed the results. Yet, it is important to acknowledge that the potential absence of negative results might lead to an overstated perception regarding the prevalence of the identified misconceptions, i.e. many publications report that misconceptions exist, but in reality the portion of users affected by these misconceptions is small. Therefore, an investigation in actual prevalence levels of the identified misconceptions among users is an important next step to properly gauge their prevalence and thereby also their relevance.

Furthermore, non-peer-reviewed publications and those published before 2007 were explicitly excluded. This might have limited the results and led to the exclusion of relevant work. However, it can be argued that in both cases the exclusion generally increases the quality and relevancy of the literature included in the systematic literature review: non-peer-reviewed publications might introduce low quality results and publications which are too old might lead to the inclusion of misconceptions which are not relevant anymore. In particular, publications which are too old might inflate the aforementioned limitation of introducing misconceptions into the results which have a low prevalence among users, since they have been counter-acted since their original appearance.

Last but not least, a broader selection of search terms might have resulted in finding additional relevant publications. Yet, the process applied in this work - i.e. the screening of literature for initial search terms and feedback from native English speaking experts in the field - represents a well founded and systematic way to choose appropriate search terms for the literature review.

3.3.3 Next Steps

Having identified a wide variety of misconceptions covering various aspects of password security, the next steps in this line of work should be:

- The investigation of the actual prevalence of the individual misconceptions among lay-users and thereby the identification of the misconceptions with a high need of being cleared up.
- The development of interventions (e.g. in the form or as part of of password security awareness materials) to decrease the prevalence of the identified misconceptions.

- The investigation of links and interactions between the different misconceptions as well as potential correlations between them.

The next chapter (chapter 4) realises the first two of these steps by developing suitable interventions as part of a password security awareness material and the subsequent investigation of these interventions' performance in a study with lay-users. The third aspect, i.e. the links and interactions between the different misconceptions, is left as future work.

3.4 Conclusion

This chapter represents research lying at the foundation of fighting against a decade of misconceptions about password security. In a systematic literature review 23 misconceptions about password security could be identified. The identified misconceptions cover basically all aspects of password security. This set of misconceptions is an important building block for the development of the password security awareness material presented in the next chapter.

4 Development and Evaluation of Effective Password Security Awareness Materials

Keeping one's passwords secure is no simple task. As outlined in section 2.3, passwords can get into the hands of attackers in numerous ways, e.g. they can be guessed, stolen in phishing attacks, eavesdropped when communication occurs through unsecured channels, or stolen when users are observed while entering them. Many users are not aware how such attacks work and consequently how to defend against them effectively [194, 209]. Moreover, strategies employed by users to cope with the problems they face when composing or handling their text passwords [107, 194, 195] might be based on on misconceptions about password security (cf. Chapter 3).

Consequently, users are at risk of falling victim to attacks. As explained in section 2.4, the prime way to counter this risk is supplying users with password security awareness materials. They are widely considered to help users protecting their accounts and passwords against threats [129,130]. Yet, many existing awareness materials on password security demand from users an impossible task: use only complex passwords, change them frequently, and never write them down. This kind of advice is highly problematic for users [107,194,195] and does not represent the current state of the art [149, 236]: awareness materials must effectively enable users to apply the knowledge contained within them or the time and effort spent working through the materials is spent in vain [15].

This chapter describes the creation of a correct, complete, understandable, and effective password security awareness material specifically for lay-users. The material is aimed at (a) making lay-users aware of the different attacks on passwords and user accounts as well as defences against these attacks and (b) clearing up the misconceptions about password security the lay-users might have.

To that end, a systematic process was developed (section 4.1) to achieve the awareness material's correctness, completeness, understandability, and effectiveness. The initial development of the material (section 4.2.1) is based on (a) the attacks on user authentication described in (section 2.3) and (b) the misconceptions about password security identified in the systematic literature review presented in the last chapter (chapter 3). The material was then refined incorporating feedback from independent information security experts from academia and industry ensuring its correctness and completeness (section 4.2.2) as well as feedback from lay-users ensuring its understandability (section 4.2.3).

The refined material was then evaluated in the field to ensure its effectiveness, following a specifically tailored methodology (section 4.3). The results of the evaluation (section 4.4) show that the developed material was not only received very positively and most participants found it very helpful, but it also contributes to the password-related security in organisations in three ways. Firstly, it improved the participants' ability to assess password-related behaviour with respect to the attacks described in the material as secure or insecure. Secondly, the participants improved their ability to assess the security of passwords. Thirdly, it significantly decreased the overall prevalence of misconceptions about password security. The results of a retention after six months show that all of these improvements remain even several months after the participants have read through the material. The discussion of the study results (section 4.5) outlines the final improvements to the awareness material.

Contributions described in this chapter:

- Providing a systematic iterative process for the development of correct, complete, understandable, and effective information security awareness materials combining the state of the art described in the literature, the expertise of independent experts from academia and industry, feedback from lay-users, and an evaluation of the material in the field.
- Providing an evaluated password security awareness material which improves lay-users' ability to assess password-related behaviour, improves their ability to assess the security of passwords, and clears up the identified misconceptions decreasing the misconceptions' overall prevalence.

Parts of the results described in this chapter have been published in:

- P. Mayer and M. Volkamer, "Addressing Misconceptions About Password Security Effectively", Workshop on Socio-Technical Aspects in Security and Trust (STAST, co-located with ACSAC), 2017.
- P. Mayer, C. Schwartz, and M. Volkamer, "On The Systematic Development and Evaluation Of Password Security Awareness-Raising Materials", Annual Computer Security Applications Conference (ACSAC), 2018, pp. 733–748.

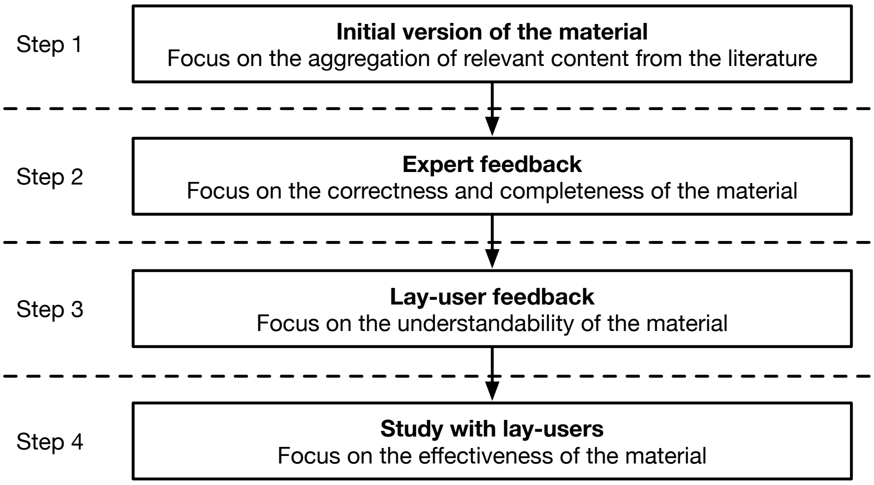

Step 1
Initial version of the material
Focus on the aggregation of relevant content from the literature
Step 2
Expert feedback
Focus on the correctness and completeness of the material
Step 3
Lay-user feedback
Focus on the understandability of the material
Step 4
Study with lay-users
Focus on the effectiveness of the material

Figure 4.1: Overview of the systematic development process, ensuring the correctness, completeness, understandability, and effectiveness of the material.

4.1 Systematic Development Process

The goal of the work presented in this chapter is to develop a correct, complete, understandable, and effective password security awareness material for lay-users. This section presents the systematic development process used to create an awareness material with these four properties. The descriptions of the individual steps in the following explicitly refer to the context of password security for lay-users. However, the process is context-agnostic and could be applied in the same fashion to any other information security context for any target audience. Figure 4.1 shows the systematic development process with all four of its steps.

4.1.1 Step 1 – Initial Version of the Material

The first step of the process focuses on the aggregation of relevant content for the password secrurity awareness material, i.e. compiling descriptions of relevant attacks and defences from the literature on password security as well as the development of interventions clearing up the misconceptions identified in the systematic literature review presented in the last chapter. Basing the awareness material closely on the literature addresses in particular the detachment of password advice and current research literature described by Murray and Malone [149] as well as Zhang-Kennedy et al. [236]. The content for the awareness material should be prepared following the recommendations presented in section 2.4 and since it specifically targets lay-users, the content must be phrased using non-technical terms wherever possible. This holds, of course for the subsequent steps as well.

4.1.2 Step 2 – Expert Feedback

The second step focuses on the correctness and completeness of the awareness material's content from an information security point of view. To that end, feedback from experts from academia and industry is collected to derive improvements for the awareness material. Consequently, the awareness material is considered *correct* in the context of this process, if its contents are free of factual errors to the extent that they cannot be identified by a group of information security experts in a review of the awareness material. Analogously, the awareness material is considered *complete* in the context of this chapter if its contents include all information that will help lay-users to better protect their passwords and user accounts, but does not require expertise in information security, to the extent that missing or superfluous content cannot be identified by a group of information security experts during a review of the awareness material.

4.1.3 Step 3 – Lay-user Feedback

The third step focuses on understandability of the awareness material from the target audience's point of view. To that end, feedback from the target audience of lay-users is collected to derive improvements for the awareness materials. Thereby, the awareness material is considered *understandable* if lay-users find none of the contents hard to understand and do not feel that more detailed explanations are required to understand the awareness material's contents.

4.1.4 Step 4 – Study with Lay-users

The fourth step focuses on the effectiveness of the awareness material. To that end, a study with the target audience of lay-users is conducted. The awareness material is considered *effective*[1] in the context of this chapter if the results of the study show that it statistically significantly (a) improves lay-users' ability to correctly assess password-related behaviour as secure or insecure in different situations, (b) improves lay-users' ability to correctly assess the security of passwords, and (c) clears up common misconceptions about password security and thereby reduces their prevalence.

4.2 Development of the Password Security Awareness Material

This section describes the application of the first three steps of the systematic development process for the creation of a correct, complete, understandable, and effective password security awareness material[2]. The fourth step (i.e. the study with lay-users) is described thereafter in sections section 4.3 to 4.5.

4.2.1 Step 1 – Initial Version of the Material

The password security awareness material developed in this chapter is text-based[3] [197]. The first step in the development of the awareness material was the aggregation of relevant content. The content for the awareness material was prepared following the first recommendation of Tsohou et al. [204], i.e. using positive phrasing and relative frequencies whenever possible. In addition, since the awareness material specifically targets lay-users, the content was phrased using non-technical terms wherever possible, e.g. using the term "fraudulent messages" instead of "phishing". In the remainder of this step's description, the content of the initial version of the material is presented.

Firstly, two introductory sections were inserted at the beginning of the material to give users a short overview of (1) who might attack them and where attacks can be targeted at and (2) the possible consequences of

[1] The definition of effectiveness in the context of the process is intentionally not defined in an abstract manner. It is highly context-specific and needs to be adapted to whichever context the process is applied to. This does not affect the context-agnostic nature of the process itself, but only how the effectiveness of the developed material needs to be assessed.

[2] The final version of the material can be accessed at `https://secuso.org/material-thesis-final-version`

[3] This decision was made in accordance with the goals of the project "KMU AWARE" (`https://secuso.org/kmu-aware`), in which the research pertaining to awareness materials presented in this thesis was conducted.

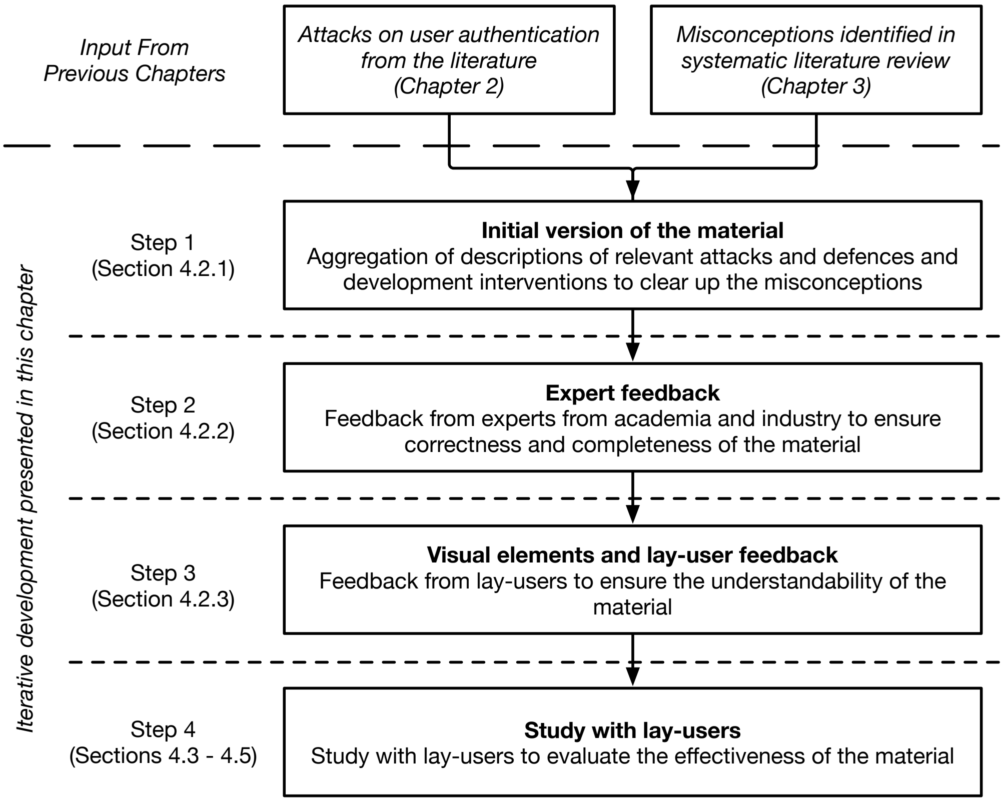

Input From
Previous Chapters
Attacks on user authentication from the literature (Chapter 2)
Misconceptions identified in systematic literature review (Chapter 3)
Iterative development presented in this chapter
Step 1
(Section 4.2.1)
Initial version of the material
Aggregation of descriptions of relevant attacks and defences and development interventions to clear up the misconceptions
Step 2
(Section 4.2.2)
Expert feedback
Feedback from experts from academia and industry to ensure correctness and completeness of the material
Step 3
(Section 4.2.3)
Visual elements and lay-user feedback
Feedback from lay-users to ensure the understandability of the material
Step 4
(Sections 4.3 - 4.5)
Study with lay-users
Study with lay-users to evaluate the effectiveness of the material

Figure 4.2: The four step process underlying this work in the context of the previous chapters.

successful attacks. The latter was thereby intended to addresses the third recommendation of Tsohou et al. [204] (i.e. emphasising immediate consequences). However, this recommendation was carefully balanced with the first, focusing in the awareness material on the positive phrasing instead of risk and fear.

Secondly, the description of the actual attacks and defences were inserted after the two introductory sections. The selection of attacks covered in the material represents the attacks on user authentication described in section 2.3. In addition, the exploitation of reset-mechanisms was considered as attack for the material. This attack is highly relevant in the context of text passwords, since commonly used reset-mechanisms such as "personal security questions" have been shown to offer low security [25]. The explanation of each attack was divided into three parts: a description of the attack, a description of the defences, and further hints. This order was chosen to have a proper problem-solution ordering in the awareness material. The described defences were based on the literature on password security and included in particular the recommendations identified as important advice, e.g. by Ion et al. [109] to use password managers and two-factor authentication (cf. section 2.4). Additionally, the explanations in the awareness material included descriptions of the following technologies: fingerprint readers, graphical passwords, hardware tokens, privacy filters, and single-sign-on. The dedicated *further hints*-section included only information that was anticipated to be relevant or interesting for few users (e.g. hints for specific software).

Thirdly, to clear up the misconceptions, interventions in the form of short texts were created for each of the misconceptions identified in chapter 3. These interventions explain the misconception itself, the situations in which it is relevant, and the respective underlying problem (e.g. the underlying problem with M1 is that numbers only improve the guessing resistance of a password, if they are put in an unpredictable place in the password). The wording and contents of this first version of the interventions were iteratively improved using informal feedback from information security and psychology experts as well as lay-users. The interventions were placed appropriately in the material to complement the descriptions of the attacks and defences (i.e. misconceptions relevant in the context of guessing attacks were placed in the description of the guessing attacks, etc.). The interventions were placed alongside the relevant portions of the descriptions and in case several placements would have been possible (i.e. as part of each of the descriptions of the three guessing attacks), they were placed in all appropriate positions. The full intervention was placed in the most appropriate of all possible positions and further instances were shortened (if appropriate) in order to not unnecessarily inflate the material's length. During the iterative design, it became clear that in three instances the interventions of multiple misconceptions converged to very similar texts and only differed slightly in the wording. On multiple occasions throughout the development, feedback indicated that these misconceptions might be better cleared up together. Therefore, in three instances multiple interventions were integrated into one text: (1) M1, M2, M3, and M4, (2) M6 and M7, and (3) M15 and M16 (cf. figure 3.1). An overview of the intervention texts can be found in table A.2 in appendix A.

4.2.2 Step 2 – Expert Feedback

Following the systematic development process, a round of expert feedback was held. Its goal was to ensure the awareness material's completeness and correctness from an information security point of view (cf. section 4.1). For this purpose, a PDF-file of the initial version of the awareness material was created which had an additional dedicated feedback page inserted after each of the two introductory sections as well as after the descriptions of each attack. Each feedback page had two free text questions asking (a) if the expert felt that any aspects would be missing from the respective section of the awareness material, and (b) if the expert thought that the section of the awareness material should be altered to be clearer.

Overall 30 independent information security experts from academia and from industry were contacted (i.e. researchers, information security consultants, auditors, administrative staff, and developers of security solutions) to give feedback. They received via email the awareness material with instructions to give feedback on the descriptions and interventions in each of its sections using the dedicated feedback pages. The experts were contacted based on their expertise in the password security domain. Only German native-speakers were contacted, since the awareness material was created in German. From the 30 experts who were contacted, 13 sent back their feedback: three researchers, four IT security consultants, three IT administrators, two people working in the IT security department of their companies, and one person working in a company testing and developing security solutions. Thus, the feedback stemmed from a diverse set of information security experts. In the following, the major improvements to the awareness material derived from the feedback are outlined.

More detailed information about possible consequences

Due to the focus on positive phrasing, the initial version of the awareness material comprised only relatively few, but broad examples of possible consequences of successful attacks. All experts noted that more concrete examples would be beneficial. Therefore, the focus on positive phrasing was adjusted based on the expert feedback and the respective introductory section was expanded with concrete examples of consequences of breaches for different attacks and types of user accounts (e.g. banking, email, social networks, etc.).

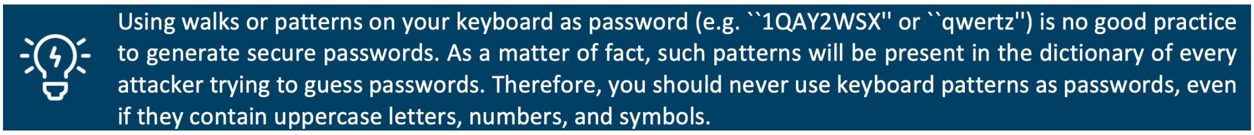

Using walks or patterns on your keyboard as password (e.g. ``1QAY2WSX" or ``qwertz") is no good practice to generate secure passwords. As a matter of fact, such patterns will be present in the dictionary of every attacker trying to guess passwords. Therefore, you should never use keyboard patterns as passwords, even if they contain uppercase letters, numbers, and symbols.

Figure 4.3: Example of the design of the interventions after the expert evaluation.

Split of the attacks on network communication

The majority of experts felt that the section on attacking the network communication would benefit from a split in two sections: one covering attacking unencrypted communication and one covering attacking encrypted communication. Therefore, the section was split up accordingly into one attack named “Eavesdropping on unencrypted communication” and one attack named “Eavesdropping on encrypted communication”. For these two attacks the wording was improved based on the expert feedback as well, replacing the term “compromising” with “eavesdropping” since it was frequently mentioned that in the context of network communication the term might be better suited for lay-users.

The aspect of physical access in order to compromise devices

Several experts noted that the description on how devices can be compromised should not mainly focus on malware, but include physical access to the devices more prominently than in the initial version. While the awareness material already instructed users to set a password lock on their devices and lock the devices whenever they leave them unattended, the experts felt further scenarios were of importance (e.g. access to data on unencrypted hard drives by removing the drives from an unattended device). Consequently, the respective section was reworked to include these aspects.

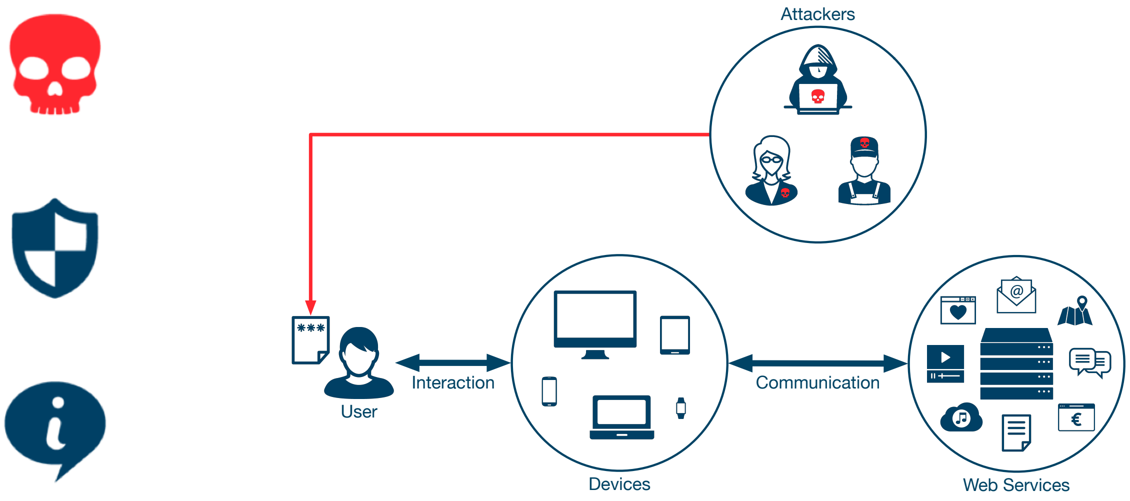
Improvements of the interventions to address the misconceptions

In addition to the aforementioned aspects, the expert feedback lead to several minor refinements of the interventions (such as additional explanations in the texts), but no larger changes had to be introduced. The wording of all intervention texts in table A.2 in appendix A already reflects these improvements.

Additionally, there was consistent feedback from the experts on the presentation of the interventions, i.e. that they should be visually separated to allow recognising the texts as interventions even when integrated in longer texts on password security. Therefore, the presentation of all interventions was adapted accordingly by adding a dedicated memorable light bulb icon identifying the interventions as such and by visually separating the intervention text from the other elements in the awareness material by using a coloured box. This design is illustrated in figure 4.3.

4.2.3 Step 3 – Lay-user Feedback

The third step of the development focused on the understandability of the awareness material. Also, visual elements were added to the awareness material at the beginning of this step. Firstly, dedicated icons signified the different types of content: a red skull signifying the attack description, a blue shield signifying the defence description, and a blue speech bubble with an "i" on it to signify further hints. Secondly, images illustrating each of the attacks were added to the attack's descriptions. Figure 4.4 depicts these visual elements.



(a)

(b)

Figure 4.4: The visual elements in the awareness material: (a) the icons signifying different types of content; (b) an example of the images included for each of the attacks (here: illustrating the theft of a note of a password). The image intentionally includes different types of attackers.

Qualitative feedback of lay-users was then used to improve the awareness material with respect to both, the textual descriptions and the visual elements. To this end, several lay-users from the Technische Universität Darmstadt[4] (i.e. secretaries, designers, and project coordinators) were invited to a lab. They were given the awareness material and asked to point out any aspect they had problems understanding or found visually unappealing. With respect to the textual descriptions, only minor changes (e.g. wording) were necessary. Regarding the interventions, the lay-users had only minor feedback as well. Thus, the final design remained virtually unchanged from the one depicted in figure 4.3.

4.3 Methodology of the User Study

The study was conducted with 90 lay-users employed in three SMEs (30 participants in each SME)[5]. Following the recommendation of Haeussinger and Kranz [91] the awareness material was evaluated in the participating employees real work environments. The goals of this study is to ensure the effectiveness of the developed password security awareness material as defined in section 4.1.4. Additionally, qualitative feedback is collected from the lay-users with respect to the usefulness of the awareness material and the images added in the third iteration of its development. The study methodology conforms to all requirements of the university's ethics commission[6]. In the following, the hypotheses, procedure, and questionnaire items of the study as well as some important aspects of the analysis methodology are explained.

4.3.1 Hypotheses

This section presents the hypotheses of the study along the three aspects of the awareness materials effectiveness as defined in section 4.1.4: (a) improvement of lay-users' ability to correctly assess password-related

[4] The research presented in this part of the thesis was conducted at Technische Universität Darmstadt.

[5] This decision was made in accordance with the goals of the project "KMU AWARE" (`https://secuso.org/kmu-aware`), in which the research pertaining to awareness materials presented in this thesis was conducted.

[6] The research in this part of the thesis was conducted at Technische Universität Darmstadt.

behaviour as secure or insecure in different situations, (b) improvement of lay-users' ability to correctly assess the security of passwords, and (c) clearing up the common misconceptions about password security of lay-users.

The first aspect of the awareness material's effectiveness is whether lay-users recognise a situation of attack (i.e. they can assess a situation as secure or insecure) and know how to behave in such situations. To that end, typically pre-treatment and post-treatment questionnaires with the same items are used to measure the difference in performance of the participants (where the treatment is the awareness material). The respective hypothesis is:

> H_{1a}**:** *The awareness material significantly increases lay-users' ability to discern secure from insecure password-related behaviour in different situations **known** before reading through the material.*

However, a concern with such evaluations is that the pre-treatment questionnaires might prime the participants with respect to the treatment. It remains therefore unknown if an improvement which is measured after the participants have read through the awareness material can be transferred to new situations, i.e. situations unknown to the participants before reading through the material. Therefore, in the post-treatment questionnaire it is not only investigated whether the participants improve their ability to discern secure from insecure password-related behaviour in situations known from the pre-treatment questionnaire, but also how they performed in new situations unknown before reading through the awareness material. The corresponding hypothesis is:

> H_{1b}**:** *The awareness material significantly increases lay-users' ability to discern secure from insecure password-related behaviour in different situations **unknown** before reading through the material.*

The second aspect of the awareness materials's effectiveness is the lay-users' ability to correctly assess the security of passwords. It is essential to the assessment of behaviour related to password security in the face of guessing attacks. Even when relying on password managers, users most often have to choose some passwords themselves, in particular the master password for their password manager [236], but also the password for unlocking their devices and other uses [107]. The respective hypothesis is:

> H_{2a}**:** *The awareness material significantly increases lay-users' ability to correctly assess the security of passwords.*

The third aspect of the awareness materials's effectiveness is whether the developed intervention texts clear up the misconceptions and thereby reduce their prevalence. The respective hypothesis is:

> H_{3a}**:** *The intervention texts in the awareness material significantly decrease the overall prevalence of the identified misconceptions in lay-users.*

The above hypotheses H_{1a}, H_{1b}, H_{2a}, and H_{3a} pertain to the effect of the awareness material observed directly after the treatment (i.e. reading through the material). In addition to this direct effect, it is important that the effect of the awareness material does not decline even after longer periods of time, since companies will usually distribute awareness materials not continuously, but rather in intervals (e.g. annually, biannually, or quarterly). Therefore, the effects of the awareness material were also investigated in a retention after six months. The respective hypotheses are

H_{1c}: *The awareness material significantly increases lay-users' ability to discern secure from insecure password-related behaviour in different situations* ***known*** *before reading through the awareness material even six months after reading it.*

H_{1d}: *The awareness material significantly increases lay-users' ability to discern secure from insecure password-related behaviour in different situations* ***unknown*** *before reading through the awareness materialeven six months after reading it.*

H_{2b}: *The awareness material significantly increases lay-users' ability to correctly assess the security of passwords even six months after reading it.*

H_{3b}: *The intervention texts in the awareness material significantly decrease the overall prevalence of the identified misconceptions in lay-users even six months after reading through the awareness material.*

4.3.2 Procedure

To investigate the hypotheses outlined in the last section, a study procedure consisting of four phases was designed: (1) a pre-treatment questionnaire measuring the baseline for the hypotheses in the participant sample; (2) exposure of the participants to the treatment (i.e. the awareness material), (3) a post-treatment questionnaire measuring the effect of the treatment with respect to the aforementioned hypotheses and gathering the qualitative feedback as well as collecting basic demographics data, and (4) a retention questionnaire measuring the effect of the treatment with respect to the aforementioned hypotheses after six months. Figure 4.5 depicts an overview of these phases in the context of the hypotheses presented in the last section.

The evaluation was conducted with employees at three SMEs in Germany. Consequently, the user study was conducted in German, i.e. the awareness material and the questionnaires were given to the participants in German. The participants were explicitly selected as lay-users with respect to information security and from a wide range of professions by a contact person in each of the three SMEs. The contact person also sent out and collected the questionnaires. Using a contact person as intermediary in each organisation ensured that participants remained anonymous, despite answering the questionnaires in their real work environment. Participants received the questionnaires and the awareness material as PDF-files via email one after the other as per the four phases outlined before (i.e. one PDF-file per phase). The PDF-file in the fist phase comprised the pre-treatment questionnaire with the respective instructions. The PDF-file in the second phase comprised only the awareness material. The PDF-file in the third phase comprised the post-treatment questionnaire with the respective instructions. The PDF-file in the fourth phase comprised the retention questionnaire and the respective instructions. Only upon sending the completed pre-treatment questionnaire, the participant received the awareness material with the instruction to take their time to read it. Once the participants confirmed that they had read the awareness material, they received the post-treatment questionnaire. After all participants in an organisation had completed the post-treatment questionnaires, the contact person sent

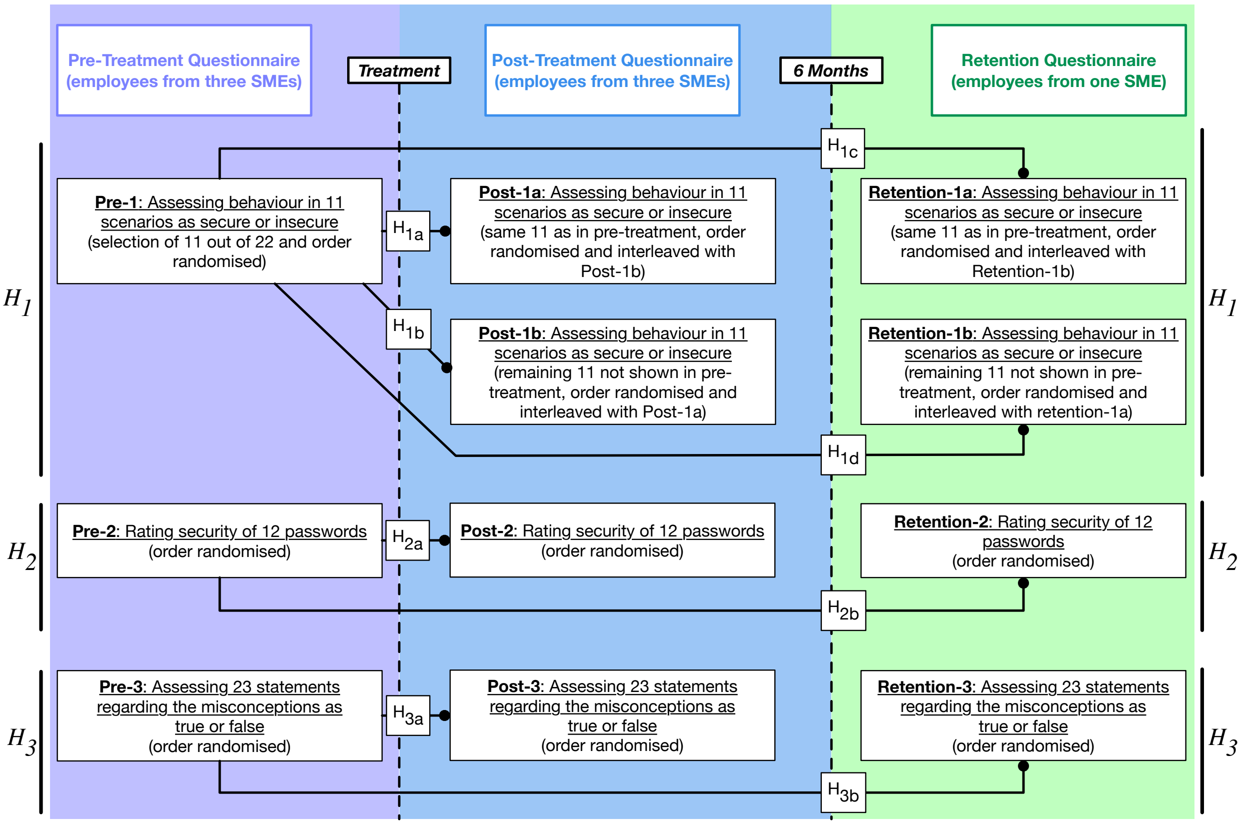

Pre-Treatment Questionnaire (employees from three SMEs)
Treatment
Post-Treatment Questionnaire (employees from three SMEs)
6 Months
Retention Questionnaire (employees from one SME)
H1
H1c
Pre-1: Assessing behaviour in 11 scenarios as secure or insecure (selection of 11 out of 22 and order randomised)
H1a
Post-1a: Assessing behaviour in 11 scenarios as secure or insecure (same 11 as in pre-treatment, order randomised and interleaved with Post-1b)
Retention-1a: Assessing behaviour in 11 scenarios as secure or insecure (same 11 as in pre-treatment, order randomised and interleaved with Retention-1b)
H1b
Post-1b: Assessing behaviour in 11 scenarios as secure or insecure (remaining 11 not shown in pre-treatment, order randomised and interleaved with Post-1a)
Retention-1b: Assessing behaviour in 11 scenarios as secure or insecure (remaining 11 not shown in pre-treatment, order randomised and interleaved with retention-1a)
H1d
H2
Pre-2: Rating security of 12 passwords (order randomised)
H2a
Post-2: Rating security of 12 passwords (order randomised)
Retention-2: Rating security of 12 passwords (order randomised)
H2b
H3
Pre-3: Assessing 23 statements regarding the misconceptions as true or false (order randomised)
H3a
Post-3: Assessing 23 statements regarding the misconceptions as true or false (order randomised)
Retention-3: Assessing 23 statements regarding the misconceptions as true or false (order randomised)
H3b

Figure 4.5: Overview of the study design with respect to the hypotheses in the analysis. Note that H_{1a}, H_{1b}, H_{2a}, and H_{3a} pertain to the data collected from the participating employees of all three SMEs. Since only one SME participated in the retention after six months, H_{1c}, H_{1d}, H_{2b}, and H_{3b} pertain only to the data collected from the participants in that one SME. The ● indicates for each hypothesis where better performance is expected.

the filled-out questionnaires to the authors. Then, after six months the contact person received the retention questionnaires for all participants and again first distributed them and then sent the filled ones back for analysis. Only one of the three SMEs agreed to participate in the fourth phase (i.e. the retention session). This is further discussed in section 4.5.

4.3.3 Questionnaire Items

This section presents the questionnaire items used to test the hypotheses described in section 4.3.1. All items were developed in an iterative process using feedback from psychologists and from two rounds of pre-tests with lay-users.

Items for $H_{1a\text{-}d}$

Hypotheses H_{1a-d} evaluate the awareness material's effectiveness, i.e. whether the knowledge regarding the attacks and defences is conveyed to the participants and they are able to distinguish secure and insecure behaviour in relevant situations. To this end, 22 scenarios were developed. Each scenario was modelled to reflect a situation employees might encounter in their daily lives. Two scenarios were developed relating each of the 11 attacks described in the awareness material: one representing secure behaviour and one representing insecure behaviour. The scenarios were developed with additional feedback from information security consultants, in order to increase the real world relevancy of the scenarios for SME employees. The

Table 4.1: The passwords used in the study. The interested reader finds alongside them the strength estimates according to the Password Guessing Service (PGS; unguessed passwords are represented by negative guess numbers), and zxcvbn (FH = offline cracking fast hashing algorithm; SH = offline cracking slow hashing algorithm).

Password	**Scores**		
	PGS (guesses)	***zxcvbn FH (sec)***	***zxcvbn SH (sec)***
1q2w3e4r$	18855448	< 0.001	1.422
Amazon1	83224531	< 0.001	5.775
tIG3R	81261755508	< 0.001	0.168
karina5!	48797777197	< 0.001	12.040
Q1w2e3r4%	91785526330	< 0.001	1.647
samsung!!!	48796663503	< 0.001	2.140
@Q1w2e3r4#	91785528114	0.010	10021.344
EimerKnirpsGoldSchelmerei	-5	$1.23 * 10^{14}$	$1.23 * 10^{20}$
HofJahreszahlAnfangskurs	-5	$2.70 * 10^{14}$	$2.70 * 10^{20}$
RechtsordnungKiesigHaushoch	-5	$1.00 * 10^{17}$	$1.00 * 10^{23}$
EssayRiesenrolleProbabilistisch	-5	$7.87 * 10^{18}$	$7.87 * 10^{24}$
SchießenAußenwandUltraschallgeber	-5	$3.78 * 10^{22}$	$3.78 * 10^{28}$

goal was to create challenges for the participants which did not simply test the participant's declarative knowledge about the attacks and defences. Instead, the created scenarios required the participant to judge password-related behaviour aligned to the attacks and defences as secure or insecure in the same way they would have to judge their own behaviour when applying the newly gained knowledge in different situations of their daily lives. For example, the scenario relating to the *secure* behaviour with respect to *storing a physical note of the password* is:

> *Mr. Schmidt has to change the password for one system in the company every 90 days. He already had to call the help desk of his company multiple times to have them reset a password he could not remember after a mandatory change. When changing the password for the next time, he makes a note of it and stores the note under his mouse pad on his desk.*

An overview of all 22 scenarios can be found in table A.3 in appendix A. In the questionnaires, each scenario was accompanied by two questions: (1) a binary question where the participants had to indicate whether they believed the scenario represents secure or insecure behaviour and (2) an open text question offering the participants the possibility to justify their decision.

To allow testing H_{1a-d} using the developed scenarios, the pre-treatment questionnaire comprised only 11 of the 22 scenarios chosen at random for each participant (one for each of the attacks, balanced in terms of secure/insecure behaviour). The post-treatment questionnaire comprised all 22 scenarios. This allowed to assess not only the performance in scenarios known before the treatment (H_{1a}), but also the participants' ability to transfer the gained knowledge to new scenarios (H_{1b}). The retention questionnaire also included all 22 scenarios allowing to test H_{1c} and H_{1d} respectively. The order of the scenarios was randomised for each participant in all questionnaires.

Items for $H_{2a,b}$

For $H_{2a,b}$ the participants had to rate 12 passwords according to their security on a 5-point Likert scale. For the Likert items, only the two poles of the scale were labelled as *very insecure* and *very secure*. Of the 12 passwords, seven were chosen to be guessable within seconds using Hashcat with the best64 and generated2 rule sets in conjunction with Mark Burnett's wordlist [39] which he specifically released for academic research (in the following "insecure passwords"). These passwords resembled examples of very insecure passwords given in the awareness material (e.g. keyboard walks, single common word with number and '!' at the end, etc.). The remaining 5 passwords were chosen using a German diceware list (about 80'000 entries, created from the German Mozilla Firefox dictionary) to be not guessable with reasonable effort (in the following "secure passwords"). These passwords resembled the advice for secure passwords given the material, i.e. concatenating words to reach a length of 20 characters or more. All twelve passwords were included in all three questionnaires (i.e. pre-treatment, post-treatment, retention). Table 4.1 lists all twelve passwords along three popular strength estimates to allow comparing their strength to published literature: their guess number as determined using the Password Guessing Service [210] and two estimates using the zxcvbn heuristic (offline cracking fast hashing algorithm and offline cracking slow hashing algorithm) [224].

Items for $H_{3a,b}$

To evaluate the awareness material's effectiveness in clearing up the misconceptions (i.e. H_{3a} and H_{3b}) 23 items were developed (one for each misconception). For each of those items, the participant had to state whether they believe the statement was correct or incorrect. The items were developed with feedback from psychologists familiar in item design. Following this feedback, seven out of the 23 items were formulated as the inverse of the respective misconception, in order to give the participants both correct and incorrect statements to respond to. Table A.4 in appendix A lists all 23 items. All 23 items were included in all three questionnaires (i.e. pre-treatment, post-treatment, retention). The order of the items was randomised for each participant in all three questionnaires.

Qualitative and Demographic Items

The qualitative questions inquiring the participants' opinion on the awareness material's content and visual elements as well as the demographics questions were only present in the post-treatment questionnaire. With respect to the qualitative feedback, the participants were asked free text questions regarding four aspects: (1) the relevancy of the included information (item: "Was the content of the awareness material relevant for you?"); (2) additional information the participants would have hoped for in the awareness material (item: "Which additional information would you have hoped for in the awareness material?"); (3) helpfulness of the images for understanding the content of the awareness material (item: "Were the images helpful in understanding the content of the awareness material? How could they be improved?"); and (4) whether the awareness material will have an effect on how the participants manage their passwords and, if so, what effect it is (item: "Will the content of the awareness material influence the way you currently manage your passwords?"). As demographics, only the participants' gender and age were collected and analysed.

4.3.4 Analysis

For the analyses pertaining to the assessment of behaviour in scenarios related to password security (i.e. H_{1a-d}), the participants' responses are aggregated into ratios of correct responses for each scenario. Thereby, it is important to note that H_{1a} and H_{1c} are thus based on a paired test design: all responses pertaining to scenarios not seen by the respective participant in the pre-treatment questionnaire are excluded from the post-treatment questionnaire data in this analysis. Consequently, paired hypothesis tests are used in this case. In contrast, the analyses of H_{1b} and H_{1d} are based on an unpaired design: the responses in the post-treatment questionnaire stem solely from participants that have not seen the respective scenarios in the pre-treatment questionnaire. Consequently, unpaired hypothesis tests are used in this case. Since the data is not normally distributed the non-parametric Wilcoxon signed rank test for paired samples and the non-parametric Wilcoxon rank sum test for unpaired samples are used. Where appropriate, Bonferroni-Holm-corrected α-levels are used to correct for multiple testing. Effect sizes are interpreted according to [49] as small ($r \geq 0.10$), medium ($r \geq 0.30$) or large ($r \geq 0.50$).

4.4 Results

This section presents the analysis of the data collected in the course of the user study. First, the participant sample is described. Then, the analysis with respect to the assessment of the scenarios with password-related behaviour as secure or insecure is presented. Thereafter, the analysis regarding the ratings of passwords as secure or insecure is presented. Then, the analysis pertaining to the prevalence of the misconceptions about password security is presented. Last but not least, the qualitative results are described.

4.4.1 Participants

The data pertaining to the pre-treatment and post-treatment questionnaires (i.e. for hypotheses $H_{1a,b}$, H_{2a}, and H_{3a}) was collected from overall 90 employees of three SMEs in Germany. Six participants had to be excluded from the analysis for the the pre-treatment and post-treatment questionnaires. Their answers to the free text questions showed detailed knowledge of information security (e.g. different encryption algorithms) and therefore raised doubts as to whether they would qualify as lay-users. Of the remaining 84 participants, 56 are male, 27 are female, and one participant chose to not answer this question. The participants' age ranged from 19 years to 43 years (M: 30.0 years; SD: 5.4 years).

The data pertaining to the retention questionnaires was collected from only 30 employees, since only one of the three SMEs agreed to participate in the retention after six months. Any differences between the SME partaking in the retention (SME-r in the following) and the other two SMEs are highlighted before presenting the analyses of the respective hypotheses (i.e., $H_{1c,d}$, H_{2b}, and H_{3b}). Participants who had been excluded from the analysis of the pre-treatment and post-treatment questionnaires were also excluded from the analysis of the retention. Overall, the sample for the retention questionnaire comprised 26 participants (after exclusions). Of those 26 participants, 16 were male, 9 were female, and one participant chose to not answer this question. The participants' age ranged from 19 years to 43 years (M: 32.2 years; SD: 5.5 years).

4.4.2 Assessment of Scenarios

Figure 4.6 shows an overview of the participants' responses for each individual scenario. It becomes apparent that for most scenarios, the participating employees assessed the described behaviour correctly in the pre-treatment questionnaire. In particular, some scenarios exhibit ceiling effects. A ceiling effect appears when the number of correct responses in the pre-treatment questionnaire is already so high that a significant increase of correct responses cannot occur with the sample size of this study, i.e. even if all incorrect responses in the pre-treatment questionnaire would be affected by the treatment and resulted in correct responses in the post-treatment questionnaire, this would not result in a significant difference. Such ceiling effects occur for S1, S4, S5, S6, S7, S8, S11, S14, S15, S18, and S21.

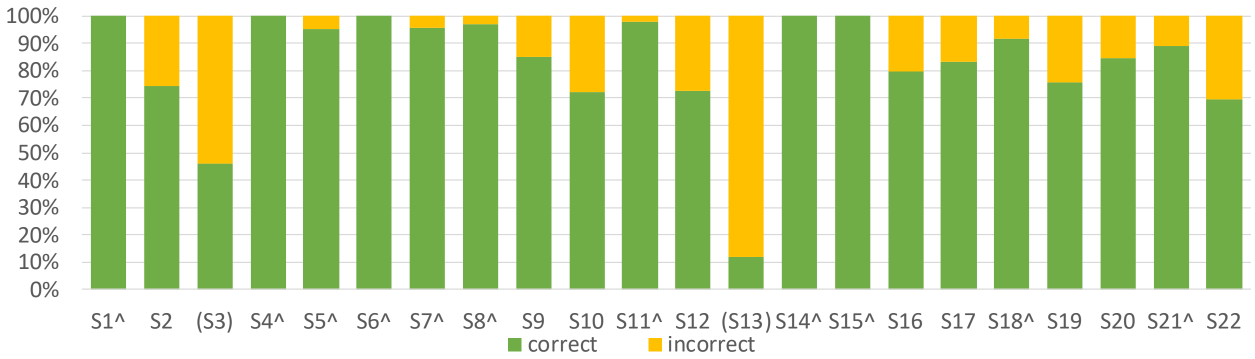


Figure 4.6: Overview of the participants' responses for each individual scenario in the pre-treatment questionnaire. The scenarios affected by ceiling effects are marked with a '^'. The two scenarios with methodological problems are set in parentheses.

In contrast, some scenarios exhibit relatively large portions of incorrect responses. Particularly scenario 3

> *Mr. Schmidt finds it difficult to remember his passwords. Therefore, he keeps a note of his private passwords at home in a locked drawer of his desk, which only he can open.*

and scenario 13

> *Mr. Schmidt takes his private smartphone to work (but does not use it for business purposes). He does not want that friends or colleagues can access the phone by guessing his PIN. Since he shares the phone with his wife in their free-time, she should be able to easily re-member the PIN. Therefore Mr. Schmidt uses as PIN the birthday of the family dog, which is only known to him and his wife.*

stand out due to large numbers of incorrect answers. The reason for this discrepancy from the other scenarios could be found in the free text answers pertaining to these two scenarios. They indicated that specific formulations in the scenarios caused participating employees to misinterpret them. This indicates methodological problems with the scenarios (as opposed to problems with the content of the awareness material), which were not uncovered in the pre-tests. For scenario 3, the majority of participants perceived a locked drawer at home not as secure storage, despite the scenario explicitly stating that only the legitimate owner can open it. Likewise, for scenario 13, the majority of participants perceived that a dog's birthday would not be a secret, despite the scenario explicitly stating this as fact. Therefore, these two scenarios were excluded from any further analysis and only the responses to the remaining 20 scenarios were considered.

Pre-Treatment to Post-Treatment Analysis

Figure 4.7 gives an overview of the portions of overall correct responses for the *pre-treatment* (Pre-1), *post-treatment known scenarios* (Post-1a), and *post-treatment unknown scenarios* (Post-1b) measurements, i.e. the measurements pertaining to the analyses of H_{1a} and H_{1b}. A substantial difference between Pre-1 and both of Post-1a and Post-1b becomes apparent. The hypothesis tests pertaining to these differences are described in the following.

Analysis of $\mathbf{H_{1a}}$. The awareness material leads to an overall mean increase in correct responses for the scenarios from 88.2% before the treatment to 93.4% afterwards, when only considering the responses with

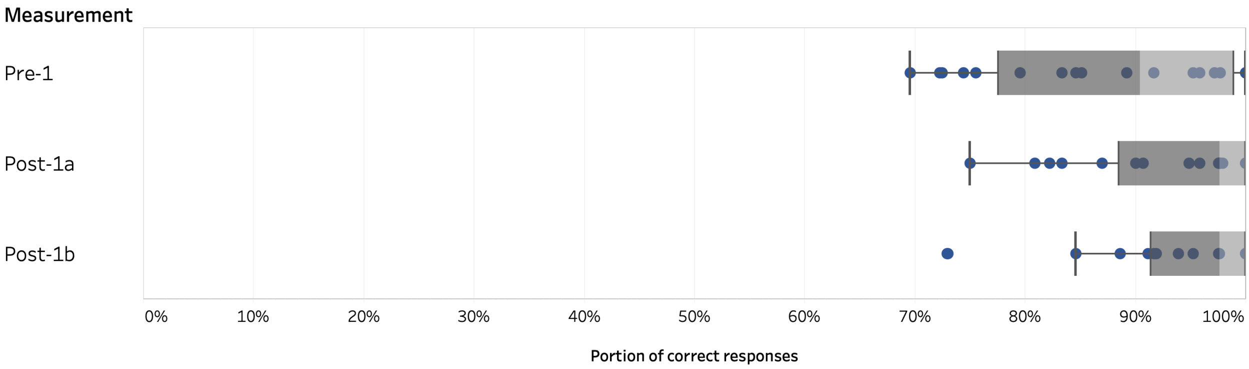


Figure 4.7: Overview of the ratios of correct responses to the scenarios for the pre-treatment and post-treatment questionnaires.

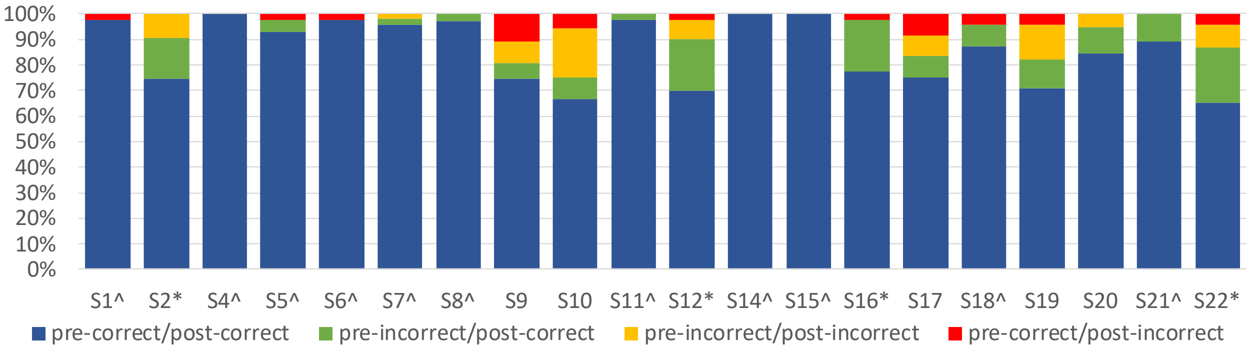


Figure 4.8: The change in participants' responses for each individual scenario from the pre-treatment to the post-treatment questionnaires. Each colour represents the respective number of participants with "*pre-treatment/post-treatment*" responses, e.g. "pre-incorrect/post-correct" is the number of participants having responded incorrectly in the pre-treatment questionnaire and correctly in the post-treatment questionnaire. The scenarios with significant improvements are marked with a '*'. The scenarios affected by ceiling effects are marked with a '^'.

respect to the scenarios participants saw before and after the treatment (Pre-1 and Post-1a in Figure 4.5). A Wilcoxon signed rank test shows this increase to be significant ($V = 17.5$, $p = .010$). The effect size $r = 0.409$ reaches the .3 threshold, i.e. indicates a medium effect. This indicates that working through the awareness material leads to a significant improvement in the employees' ability to assess behaviour as secure or insecure in scenarios known before the treatment. Therefore, the results of the study support H_{1a}.

Since the data pertaining to H_{1a} is paired, a detailed analysis of the changes for each individual scenario is possible using McNemar's χ^2 test. Figure 4.8 shows the change in responses (i.e. correct or incorrect) from each individual scenario between the pre-treatment and post-treatment questionnaires. Scenarios S2 ($\chi^2(1) = 5.14$, $p = .023$), S12 ($\chi^2(1) = 4.00$, $p = .046$), S16 ($\chi^2(1) = 4.90$, $p = .027$), and S22 ($\chi^2(1) = 4.08$, $p = .043$) show a significant improvement between the pre-treatment questionnaire and the post-treatment questionnaire. The remaining scenarios, however, do not. This is to be expected for the 11 scenarios affected by the ceiling effect, but less so for the remaining five scenarios, i.e. S9, S10, S17, S19, and S20. Notably, despite not representing a significant difference, S9 exhibits an increase in *incorrect* responses in the post-treatment questionnaire (+2 total incorrect responses). This will be further discussed in section 4.5.

Analysis of H$_{1b}$. Additionally, it was investigated whether the participants can transfer the knowledge gained by reading the awareness material to scenarios they only saw in the post-treatment questionnaire (93.8% mean correct responses for the scenarios in Post-1b). A Wilcoxon rank-sum test does not find a significant difference ($W = 138$, $p = .089$) between the portions of correct responses in the pre-treatment

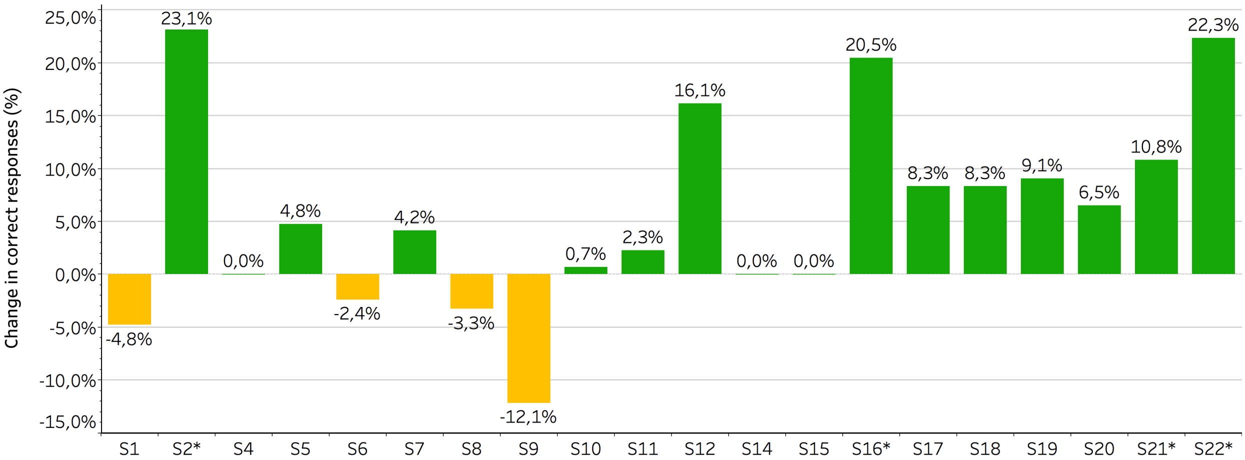


Figure 4.9: Overview of the change of correct responses by the participants for each individual scenario from the pre-treatment to the post-treatment questionnaires. The scenarios with significant improvements are marked with a '*'.

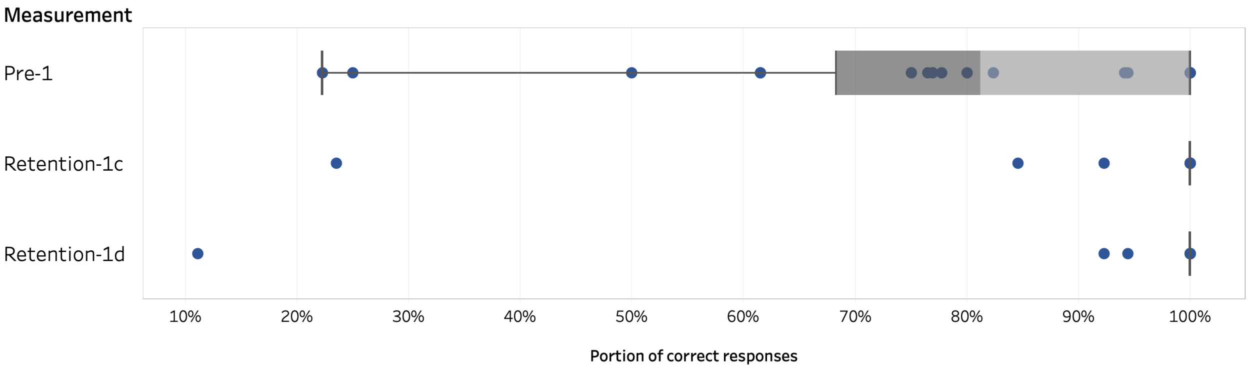


Figure 4.10: The ratios of correct responses for the scenarios in the pre-treatment and retention questionnaires. Note the different values for *Pre-1* due to the inclusion of only SME-r.

questionnaire and the responses in the post-treatment questionnaire corresponding to the scenarios only present in the post-treatment questionnaire (Pre-1 and Post-1b in Figure 4.5). While the test only closely fails significance, this result indicates that working through the awareness material might not improve the employees' ability to assess information security behaviour as secure or insecure in new scenarios, which are unknown before reading the awareness material. Therefore, the results do not seem to support H_{1b}.

The unpaired nature of the data pertaining to H_{1b}, does not allow to trace the change in response for each participant. Instead, the ratios of correct and incorrect answers must be compared, when analysing the scenarios individually. Again differences in the participants' performance between the scenarios become apparent. Figure 4.9 shows the changes for each individual scenario. Most of the scenarios show an increase in correct responses. Fisher's exact tests show that the increase is significant for scenarios S2 (FET: p = .003), S16 (FET: p = .027), S21 (FET: p = .034), and S22 (FET: p = .014). For scenarios S4, S14, and S15 no change in responses occurs: all responses pertaining to these three scenarios are correct in the pre-treatment and the post-treatment questionnaires. Unfortunately, four scenarios, i.e. S1, S6, S8, and S9, exhibit decreased ratios in correct responses (albeit none of these changes are statistically significant). This will be further discussed in section 4.5.

Pre-Treatment to Retention Analysis

The portion of correct responses increased from 81.7% in the pre-treatment questionnaire to 95.0% in the retention questionnaire, when only considering SME-r. An overview of the results is depicted in figure 4.10. For the analysis of the retention, scenarios 3 and 13 are again excluded, due to their issues outlined at the beginning of section 4.4.2.

Differences between SME-r and the other two SMEs. The employees of SME-r did not perform significantly different in the pre-treatment questionnaire than the employees of the other two SMEs with respect to the assessment of scenarios. A Wilcoxon rank-sum test does not indicate a significant difference (W = 150, p = 0.173). This holds for the post-treatment questionnaire as well, where a Wilcoxon rank-sum test also does not indicate a significant difference (W = 222, p = .549).

Analysis of H$_{1c}$. The participants of SME-r perform better in the retention questionnaire than in the pre-treatment questionnaire with respect to the scenarios seen in the pre-treatment questionnaire: The portion of correct responses increased from 81.7% in the pre-treatment questionnaire to 94.0% in the retention questionnaire, when only considering SME-r. A Wilcoxon signed rank test indicates a significant difference (V = 15.5, p = .011) and r = 0.341 indicates a medium effect for this difference. Therefore, it seems the results of the study support H_{1c}.

Figure 4.11 depicts the change in correct and incorrect responses of the participants from the pre-treatment to the retention questionnaire. These individual changes were again tested using McNemars χ^2 test. Only three scenarios (i.e. S1, S9, and S21) do not achieve 100% correct responses in the retention questionnaire. Overall, four scenarios show significant differences: S9 ($\chi^2(1)$ = 5.82, p = .016), S10 ($\chi^2(1)$ = 5.14, p = .023), S12 ($\chi^2(1)$ = 4.17, p = .041) and S22 ($\chi^2(1)$ = 4.17, p = .041). Most notably, for scenario S9 this represents a significant decrease in correct responses. This will be further discussed in section 4.5.

Analysis of H$_{1d}$. For the scenarios not seen in the pre-treatment questionnaire, assessments in the retention (96.0% correct responses) improved as well when compared to the pre-treatment questionnaire. A Wilcoxon rank-sum test indicates a significant difference (W = 96.0, p = .002). An effect size of r = 0.707 indicates a large effect. Therefore, the results of this study seem to support H_{1d}.

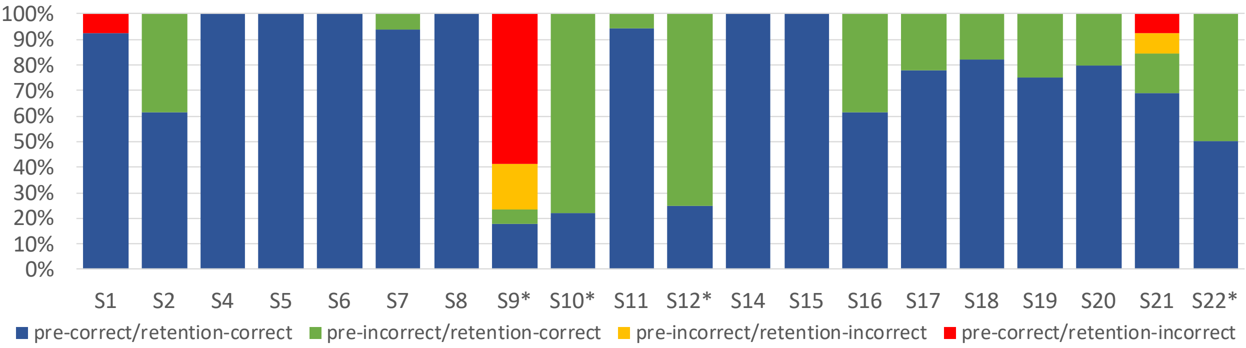


Figure 4.11: The change in participants' responses for each individual scenario from the pre-treatment to the post-treatment questionnaires. Each colour represents the respective number of participants with "*pre-treatment/post-treatment*" responses, e.g. "pre-incorrect/post-correct" is the number of participants having responded incorrectly in the pre-treatment questionnaire and correctly in the post-treatment questionnaire. The scenarios with significant improvements are marked with a '*'. The scenarios affected by ceiling effects are marked with a '⌒'.

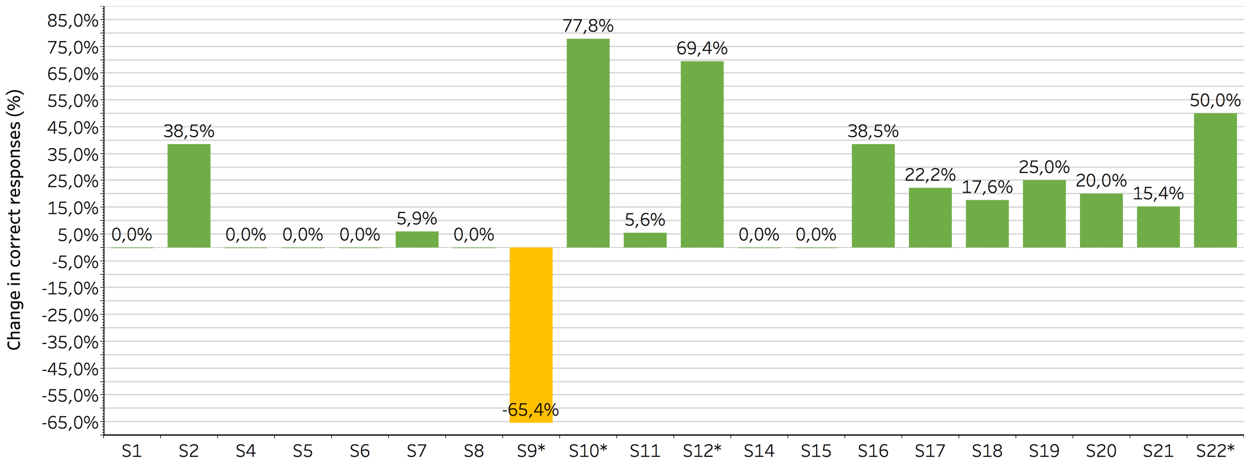


Figure 4.12: The differences in correct responses by the participants for each individual scenario from the pre-treatment to the post-treatment questionnaires. The scenarios with significant improvements are marked with a '*'.

Figure 4.12 shows the differences between the pre-treatment and retention questionnaires. In contrast to H_{1c}, only one scenario, i.e. S9, shows an increase in incorrect responses. Seven scenarios do not show any difference (i.e. S1, S4, S5, S6, S8, S14, and S15). Analogously to H_{1b}, a traced comparison (i.e. tracing whether individual participants changed their response) of the responses is not possible due to the unpaired nature of H_{1d}. Therefore, comparisons of the portions of correct responses was conducted using Fisher's exact test. The same four scenarios as for H_{1c} exhibit significant differences (i.e. S9, S10, S12, and S22). With a substantial increase of 65.4% in *incorrect* responses, S9 again stands out due to its bad performance.

4.4.3 Password Security Ratings

Only 81 participants could be included in the analysis regarding H_{2a}. In addition to the participants excluded due to the doubts regarding their status as lay-users as outlined before, three participants had to be excluded, since they did not complete the ratings of all passwords. The same problem arose in the retention questionnaire, where only 22 participants could be included in the analysis regarding H_{2b}, since four of the 26 participants did not complete the ratings for all passwords.

Pre-Treatment to Post-Treatment Analysis

Figure 4.13 shows the participants' ratings of the passwords in the pre-treatment and the post-treatment questionnaires. The analyses pertaining to these ratings are described in the following.

Analysis of H_{2a}. After the treatment, the assessment of all passwords improved, i.e. the insecure passwords were perceived as insecure by more participants and the secure passwords were perceived as secure by more participants. A Wilcoxon signed rank test showed a significant difference between the accumulated Likert scores of the 81 participants ($V = 285.5$, $p < .001$). The effect size $r = 0.426$ is above .3, i.e. a medium effect. Therefore, the results seem to provide supporting evidence with respect to H_{2a}.

Further Findings. In the pre-treatment questionnaire, the security of most of the insecure passwords was assessed correctly by the participants. In contrast, the security of the secure passwords was mostly assessed

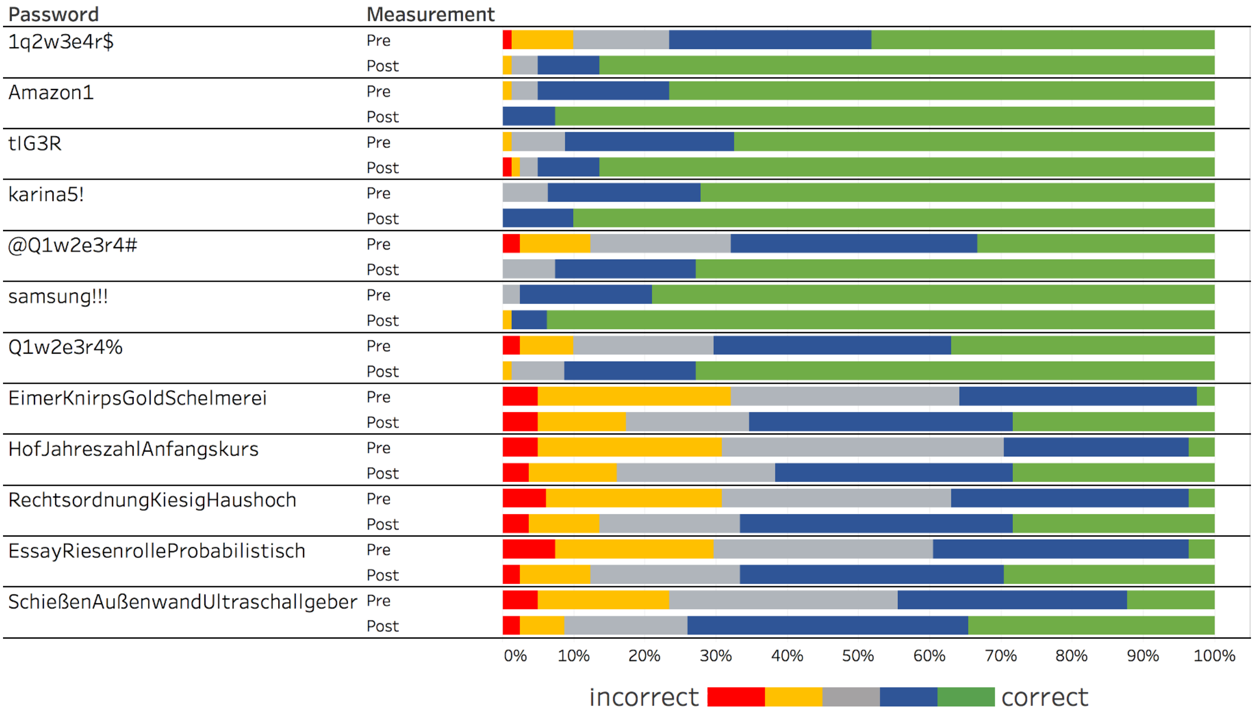


Figure 4.13: The responses on the 5-point Likert scale with respect to the perceived security. In this chart, the participants' responses are equalised in terms of correctness: the higher the value, the more correct (i.e. insecure for the easy to guess passwords and secure for the diceware passwords) is the participants' assessment.

incorrectly. A Wilcoxon signed rank test showed a significant difference in the correctness of the assessment between the secure and insecure passwords ($V = 3321$, $p < .001$). The effect size $r = 0.614$ indicates a large effect. This difference remains in the post-treatment questionnaire: a Wilcoxon signed rank test showed a significant difference ($V = 3240$, $p < .001$). The effect size $r = 0.612$ again indicates a large effect.

Pre-Treatment to Retention Analysis

Differences between SME-r and the other two SMEs. The employees of SME-r performed worse in the pre-treatment questionnaire than the employees of the other two SMEs. A Wilcoxon rank-sum test showed a significant difference ($W = 405$, $p = .004$). An effect size of $r = 0.321$ indicates a medium effect. This difference between SME-r and the other two SMEs reverses in the post-treatment questionnaire: the employees of SME-r rate the security of passwords more correctly than the employees of the other two SMEs. A Wilcoxon rank-sum test indicates this difference to be significant ($W = 977.5$, $p = .002$). An effect size of $r = 0.340$ indicates a medium effect.

Analysis of H_{2b}. A Wilcoxon rank-sum test indicates that the performance in the pre-treatment questionnaire is significantly worse than in the retention questionnaire ($V = 0$, $p < .001$). An effect size of $r = 0.603$ indicates a large effect. Thus, the results seem to support H_{2b}. An overview of the results is depicted in figure 4.14.

Further Findings. Analogously to the results of the pre-treatment and post-treatment questionnaires, a Wilcoxon signed rank test showed a significant difference in the correctness of the assessment between the

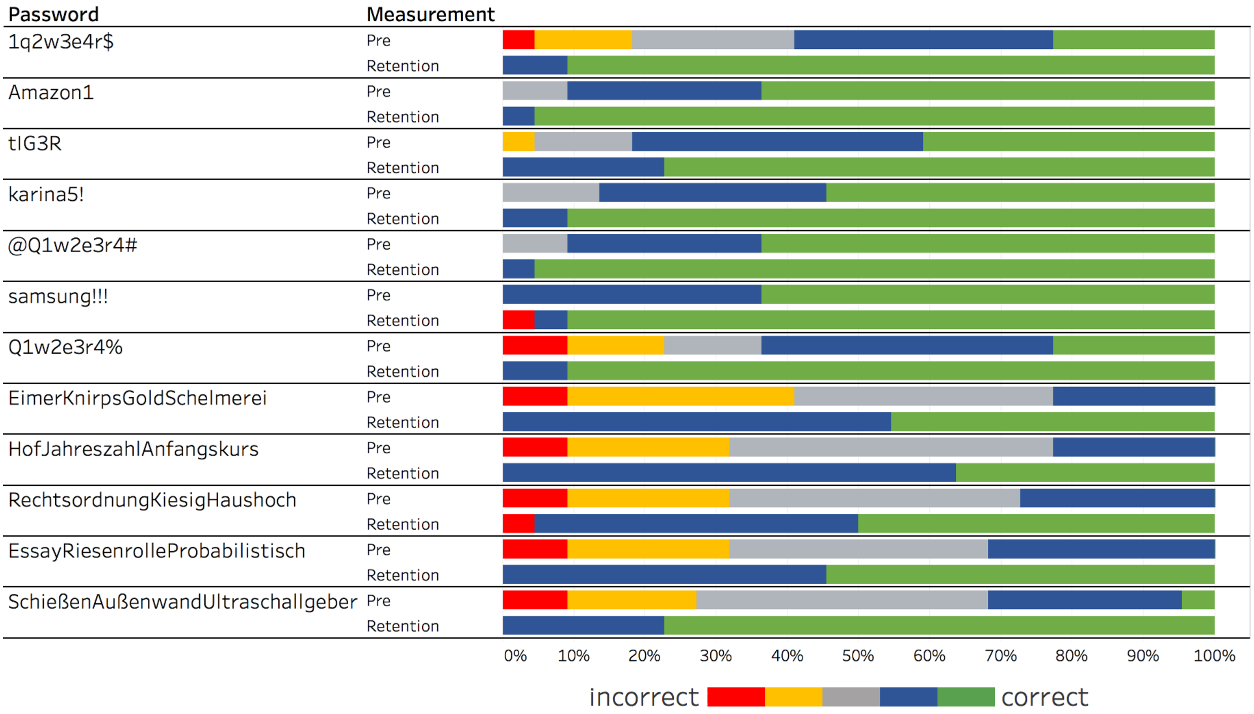


Figure 4.14: The responses on the 5-point Likert scale with respect to the perceived security. In this chart, the participants' responses are equalised in terms of correctness: the higher the value, the more correct (i.e. insecure for the easy to guess passwords and secure for the diceware passwords) is the participants' assessment.

secure and insecure passwords for the retention questionnaire ($V = 253$, $p < .001$). An effect size of $r = 0.621$ indicates a large effect.

4.4.4 Clearing Up the Misconceptions

Overall 72.8% of the responses pertaining to the misconceptions were correct in the pre-treatment questionnaire. However, as becomes apparent from figure 4.15, most of the misconceptions were prevalent in the sample of SME employees during the first phase of the study (i.e. before the treatment). Some of the misconceptions appeared in the majority of the participants. A visual inspection of the data indicates that misconceptions M1, M2, M3, M4, M10, and M11 show low portions of correct responses and seem to be especially prevalent in the sample before the intervention. In contrast, ceiling effects (i.e. no significant improvement possible) appear for the misconceptions M5, M9, M13, M15, M16, M19, and M20.

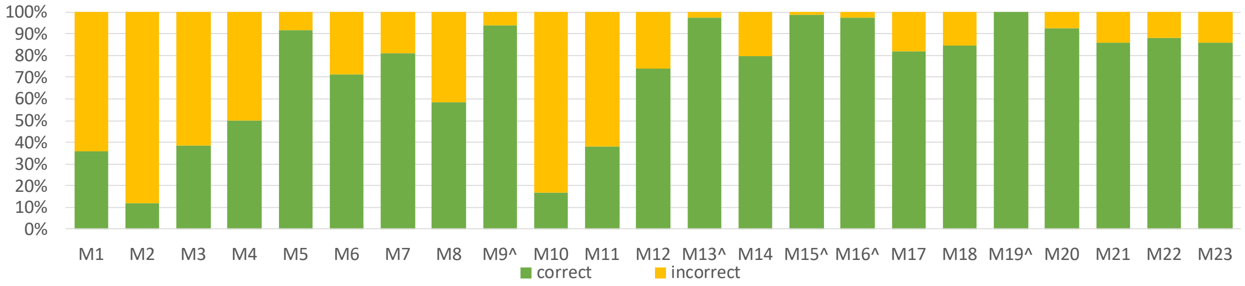


Figure 4.15: The correct and incorrect responses for each of the misconceptions *before* reading through the awareness material. The misconceptions affected by ceiling effects are marked with a ‘^’.

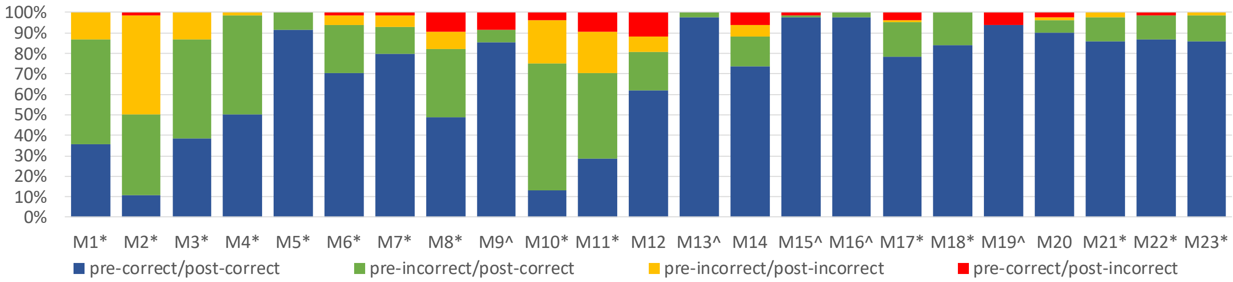


Figure 4.16: The correct and incorrect responses for each of the misconceptions before and after reading through the awareness material. Each colour represents the respective number of participants with "*pre-treatment/post-treatment*" responses, e.g. "pre-incorrect/post-correct" is the number of participants having responded incorrectly in the pre-treatment questionnaire and correctly in the post-treatment questionnaire. The misconceptions with significant improvements are marked with a '*'. The misconceptions affected by ceiling effects are marked with a '^'.

Pre-Treatment to Post-Treatment Analysis

After the intervention, participants performed better regarding the prevalence of misconceptions. The overall portion of correct responses increases from 72.8% in the pre-treatment questionnaires to 89.3% in the post-treatment questionnaires.

Analysis of H_{3a}. A Wilcoxon signed rank test with continuity correction shows a significantly higher number of correct responses per participant in the post-treatment questionnaire than in the pre-treatment questionnaire ($V = 35$, $p < .001$). An effect size of $r = 0.562$ indicates a large effect. Therefore, the results of this study support H_{3a}.

Further Findings. Figure 4.16 shows for each misconception the individual differences in correct and incorrect responses between the pre-treatment questionnaire and the post-treatment questionnaire. The individual differences were evaluated with McNemar's test.

On the one hand, all of the misconceptions which stood out with high numbers of incorrect responses before the treatment show a significant improvement: M1 ($\chi^2(1) = 41.02$, $p < .001$), M2 ($\chi^2(1) = 28.27$, $p < .001$), M3 ($\chi^2(1) = 38.03$, $p < .001$), M4 ($\chi^2(1) = 39.02$, $p < .001$), M10 ($\chi^2(1) = 41.89$, $p < .001$), and M11 ($\chi^2(1) = 15.72$, $p < .001$). Yet, it is of note that despite showing a significant improvement, M2 still exhibits 50.0% incorrect answers in the post-treatment questionnaire. The misconceptions with higher numbers of correct pre-treatment responses showing significant improvements after the treatment are M6 ($\chi^2(1) = 15.43$, $p < .001$), M7 ($\chi^2(1) = 6.75$, $p = .009$), M8 ($\chi^2(1) = 10.03$, $p = .002$), M17 ($\chi^2(1) = 5.88$, $p < .015$), M18 ($\chi^2(1) = 11.08$, $p < .001$), M21 ($\chi^2(1) = 8.10$, $p = .004$), M22 ($\chi^2(1) = 5.82$, $p = .016$),and M23 ($\chi^2(1) = 9.09$, $p = .003$).

On the other hand, despite the decrease of the overall prevalence of misconceptions, the number of correct responses did not increase for all misconceptions from the pre-treatment questionnaire to the post-treatment questionnaire. Namely, the two misconceptions M9 and M19 exhibit more incorrect answers in the post-treatment questionnaire than in the pre-treatment questionnaire. However, for both no significant differences were found.

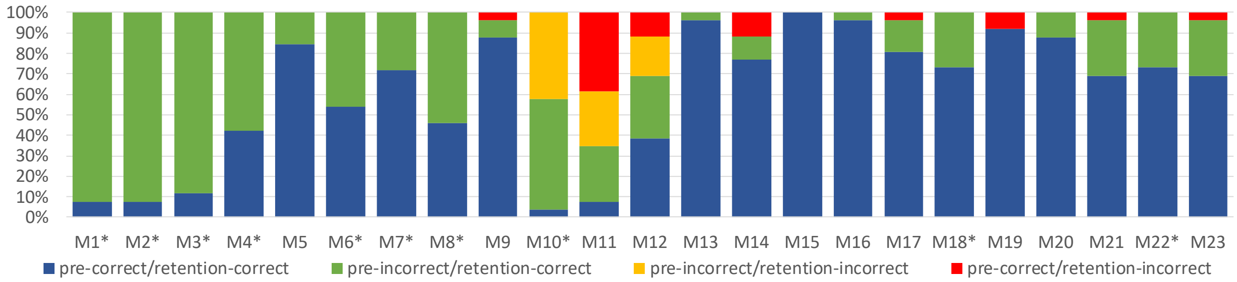


Figure 4.17: The changes in correct and incorrect responses for each of the misconceptions from the pre-treatment questionnaire to the retention questionnaire. Each colour represents the respective number of participants with "*pre-treatment/retention-treatment*" responses, e.g. "pre-incorrect/retention-correct" is the number of participants having responded incorrectly in the pre-treatment questionnaire and correctly in the retention questionnaire. The misconceptions with significant improvements are marked with a '*'.

Pre-Treatment to Retention Analysis

This section describes the results pertaining to the prevalence of the misconceptions in the retention questionnaire.

Differences between SME-r and the other two SMEs. The misconceptions were more prevalent in the employees of SME-r than in the employees of the other two SMEs. A Wilcoxon rank-sum test showed a significant difference ($W = 422.5$, $p = .001$). An effect size of $r = 0.351$ indicates a medium effect. This difference between SME-r and the other two SMEs remains in the post-treatment questionnaire, albeit smaller. A Wilcoxon rank-sum test again indicates the difference to be significant ($W = 999.5$, $p = .0.015$). However, an effect size of $r = 0.265$ indicates a small effect.

Analysis of H$_{3b}$. The portion of correct responses increased from 63.5% in the pre-treatment questionnaire to 92.4% in the retention questionnaire. A Wilcoxon signed rank test indicates that this difference is significant ($V = 0$, $p < .001$). An effect size of $r = 0.605$ indicates a large effect. Therefore, the results seem to provide supporting evidence for H_{3b}.

Figure 4.17 depicts the changes in the prevalence of the individual misconceptions. Most of the misconceptions (14 out of 23) exhibit no incorrect responses in the retention questionnaire[7]: M1 ($\chi^2(1) = 22.04$, $p < .001$), M2 ($\chi^2(1) = 22.04$, $p < .001$), M3 ($\chi^2(1) = 21.04$, $p < .001$), M4 ($\chi^2(1) = 13.07$, $p < .001$), M5, M6 ($\chi^2(1) = 5.14$, $p = .001$), M7 ($\chi^2(1) = 5.14$, $p = .023$), M8 ($\chi^2(1) = 12.07$, $p < .001$), M13, M15, M16, M18 ($\chi^2(1) = 5.14$, $p = .023$), M20, and M21. Not all of these misconceptions show significant differences due to ceiling effects. However, in particular M1, M2, and M3 stand out in SME-r with a high prevalence in the pre-treatment questionnaire, but a perfect score in the retention questionnaire and therefore a very strong (significant) improvement. In addition to the 14 misconceptions achieving perfect scores, five misconceptions achieve a prevalence of less than 10 percent (i.e. less than ten percent incorrect responses) in the retention questionnaire: M9, M17, M19, M21, and M23. None of these achieve a significant difference. Notably, the difference for M19 is an increase in incorrect responses (albeit not significant), bringing it down from a perfect score in the pre-treatment questionnaire. Unfortunately, there is another misconception which exhibits more incorrect responses in the retention questionnaire than in the pre-treatment questionnaire: M11 (again not significant). This will be further discussed in section 4.5. Of the remaining misconceptions only M10 exhibits a significant difference ($\chi^2(1) = 12.07$, $p < .001$).

[7] In the following details are only given for the significant differences.

4.4.5 Qualitative Results

While most participants answered the qualitative questions, the participants' responses to these questions were very concise. However, for all four aspects under investigation clear themes emerged.

Relevancy of Included Information. Most participants who answered the respective free text questions found the content of the awareness material relevant and helpful (90.6%). Participants were unexpectedly explicit with respect to their positive opinion about the material, e.g.: *"It was very helpful. I have learned a lot!"* (P7) and *"I believe the [content of the material] is relevant for everybody in today's world. I have never seen such good education materials, all information was very helpful."* (P62).

Three topics were perceived as particularly helpful by the participants: (1) the information regarding password composition and guessing attacks, (2) the information with respect to regular changes of passwords, and (3) the information on password managers. The most frequently voiced concern (stated by 6 participants) was that the awareness material was perceived as too long.

Additional Information the Participants Would Have Hoped For. The most frequently mentioned aspect participants would have hoped for (20.8% of participants who answered this free text question), was more concrete information with respect to password managers and software which can be used to generate passwords, e.g.: *"Concrete suggestions regarding software which can be used to generate secure passwords, about good password managers."* (P78). The participants also would have liked more concrete rules on how to choose passwords in addition to the composition advice already present in the awareness material (16.7%).

Helpfulness of the Images. All of the participants found the images in the awareness material helpful, e.g.: *"The images were very helpful in understanding the [awareness material's content]."* (P53) and *"The images were very helpful."* (P89). No further improvements to the images could be derived from the participants' free text answers.

Effect of the Awareness Material on the Users' Password Management. The dominant theme in the answers to this free text question were password managers. 20.3% of participants answering this question stated their intention to start using a password manager in the future instead of their original strategy. In addition, 61.0% stated to continue using a password manager. An additional 9.8% stated to create their passwords differently from now on (without explicitly specifying in which way). The remaining 8.9% of participants answering this question stated that the awareness material would have no impact on their behaviour and did not mention the use of password managers.

4.5 Discussion

This section presents the discussion of the study results and limitations. First the results are discussed and proposals for improvements[8] of the awareness material derived from them to create the final version of the material for this work[9]. Then the limitations of the study are discussed.

[8] Note that the empirical validation of these refinements is still future work.

[9] Future work might still further improve the material.

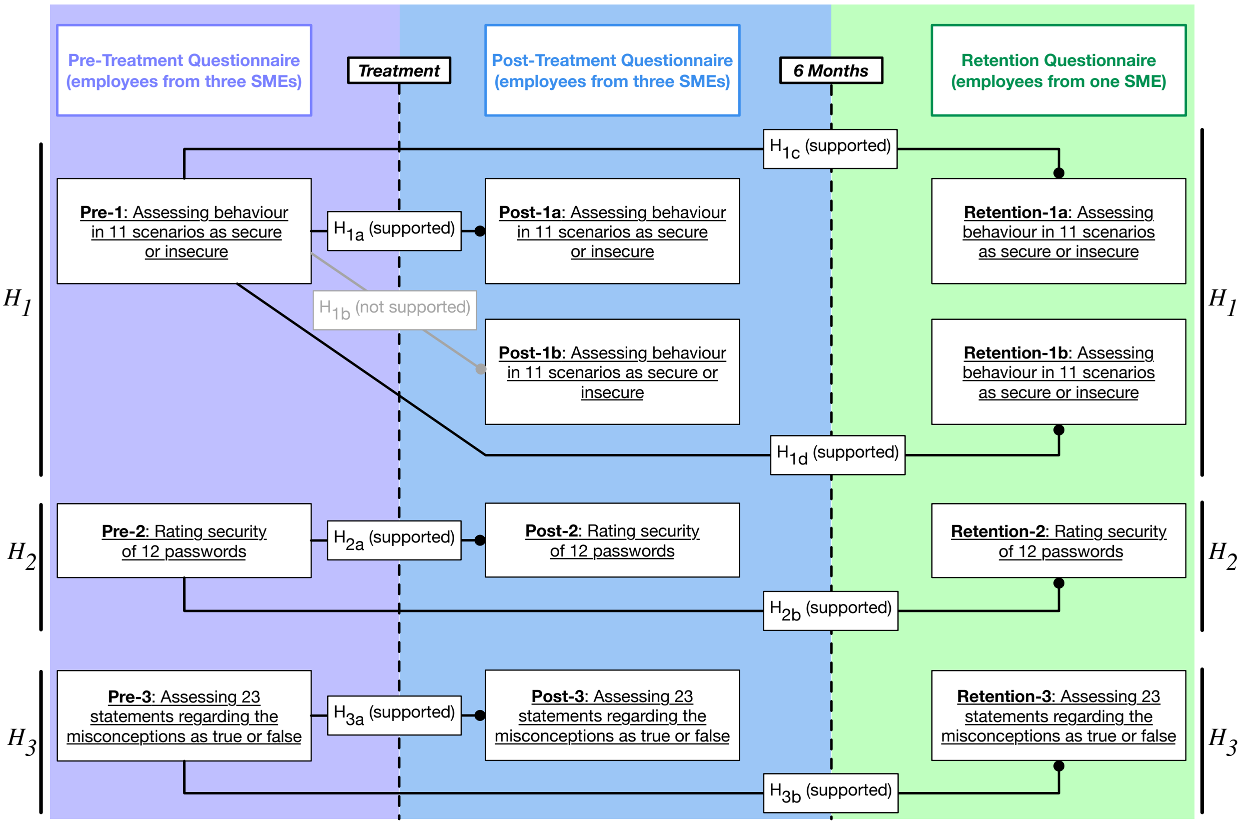

Pre-Treatment Questionnaire (employees from three SMEs)
Treatment
Post-Treatment Questionnaire (employees from three SMEs)
6 Months
Retention Questionnaire (employees from one SME)
H1
H1c (supported)
Pre-1: Assessing behaviour in 11 scenarios as secure or insecure
H1a (supported)
Post-1a: Assessing behaviour in 11 scenarios as secure or insecure
Retention-1a: Assessing behaviour in 11 scenarios as secure or insecure
H1b (not supported)
Post-1b: Assessing behaviour in 11 scenarios as secure or insecure
Retention-1b: Assessing behaviour in 11 scenarios as secure or insecure
H1d (supported)
H2
Pre-2: Rating security of 12 passwords
H2a (supported)
Post-2: Rating security of 12 passwords
Retention-2: Rating security of 12 passwords
H2b (supported)
H3
Pre-3: Assessing 23 statements regarding the misconceptions as true or false
H3a (supported)
Post-3: Assessing 23 statements regarding the misconceptions as true or false
Retention-3: Assessing 23 statements regarding the misconceptions as true or false
H3b (supported)

Figure 4.18: Overview of the results of all hypothesis tests. The ● indicates for each hypothesis where better performance is expected.

4.5.1 Results

The goals of the study could be successfully attained. The awareness material was received positively by all participating employees and addresses all attacks deemed relevant by the literature and by independent information security experts from academia and industry. Figure 4.18 summarises the results with respect to all hypotheses tested in the course of the analysis. All but one of the hypotheses are supported by the study results, with at least medium effect sizes. For the retention session after 6 months the results seem to indicate a trend towards large effect sizes. This can be attributed to the reduced participant sample (due to only SME-r participating in the retention) while retaining a similar magnitude of mean difference in the hypothesis tests. In the following, the results pertaining to each of the three study goals are discussed.

Assessment of Scenarios

The awareness material significantly increased the participants' ability to correctly assess whether password-related behaviour in different information security scenarios known before reading through the material is secure or insecure (H_{1a} supported). The transfer of the gained knowledge by the participants to new contexts is still in need for improvement: the test whether the material improves the ability to correctly assess password-related behaviour in previously unknown scenarios closely fails significance (H_{1b} not supported). This again shows how difficult it is to develop effective awareness materials, even when following a thorough methodology during their creation. One possible enhancement of the awareness material might be to include more examples, in order to illustrate more clearly how the described attacks and defences work in practice and might transfer to the context of the participants' daily lives. Also, it is of note that using a traditional methodology with the same scenarios in the pre-treatment and post-treatment questionnaires would not have been able to detect this issue of transferring gained knowledge to unknown scenarios. In the design of future studies and further improvements of the material it should be considered that the pre-treatment questionnaire will undoubtedly affect the results. Therefore, the additional complexity of the methodological design seems to be justified. Yet, a larger participant sample might have allowed for a design potentially reducing the introduction of bias, by allowing multiple groups receiving the questionnaires at different points in time (i.e. pre-treatment, post-treatment, retention). This limitation is further discussed in section 4.5.2.

Regarding the results of the retention, participants improved their ability to assess the "unknown" scenarios[10] (H_{1c} and H_{1d} supported). This might indicate that between the post-treatment questionnaire and the retention questionnaire, the participants talked about the different scenarios with colleagues also participating in the study or that they revisited the awareness material. Unfortunately, these two aspects could not be controlled, as discussed in section 4.5.2. Another potential source of improvement in the retention questionnaire might be that there was a learning effect among the previously unknown scenarios where the later seen scenarios in the post-treatment questionnaire helped the participants to correctly assess other scenarios in the retention. This effect needs more studies to be further explored since the data of this study does not allow to draw conclusions in this respect. The introduction of multiple groups receiving questionnaires at different points in time (as described in the last paragraph) in such future studies would help to establish the effect of priming effects more reliably.

When looking at the performance of the participants in the individual scenarios, despite the overall improvement in correct responses, only few scenarios show significant differences after the treatment. Out of the 20 analysed scenarios, 11 were affected by ceiling effects. While this might offer an opportunity to reduce the overall size of the awareness material, by using the scenario items presented in section 4.3.3 to measure among the individuals to which the password security awareness material is to be distributed which attacks need to be addressed (and in turn if some descriptions can be shortened or left out), it also indicates that the initial baseline of correct responses is very high. This limitation will be further discussed in section 4.5.2. In contrast, only four scenarios show significant improvements from the pre-treatment questionnaire to the post-treatment questionnaire: S2, S12, S16, and S22. This changes slightly in the retention session. While the overall number of individually significant differences remains the same in the retention questionnaire, S2 and S16 do not exhibit significant changes anymore and the difference for S10 becomes significant. However, this can be attributed to the different sample in the retention. S2 and S16 exhibit ceiling effects when only SME-r is considered in the pre-treatment questionnaire. Thus, no significant effects can be expected in the retention session. For S10 on the other hand, the improvements in performance from the pre-treatment questionnaire to the post-treatment questionnaire also largely stem from SME-r: while the other two SMEs have overall 86.2% correct responses in the post-treatment questionnaire, SME-r has overall 46.2% correct responses. Unfortunately, no participant took the opportunity to explain their choice in the free text question pertaining to S11 in the retention questionnaire. Therefore, there is no data to offer an explanation for the increase in correct responses by the employees of SME-r from 46.2% in the post-treatment questionnaire to 100% in the retention questionnaire. Potential reasons which are not specific to S11 (such as conversations between the participants about the questionnaires in the 6 months retention period) are discussed in section 4.5.2.

From the results pertaining to the assessment of the individual scenarios, improvements can be derived to create the final version of the material. The analysis of H_1a and H_1b indicated a potential need to improve

[10] Of course they were not truly unknown during the retention, since the participants had seen them in the post-treatment questionnaire. However, the priming with respect to these scenarios did not occur before working through the material.

scenarios S9, S10, S17, S19, and S20. The analysis of the retention session ($H1_c$ and $H1_d$) indicated a potential need to improve scenarios S1, S9, and S21. In the following, these scenarios are discussed along the attacks they represent in ascending order (i.e. S1, S9 & S10, . . .). The final version with all improvements was made available to the participants after the retention session.

S1 (Fraudulent Messages). S1 is among the scenarios showing a decrease below of correct answers in the retention. However, only one participant answered incorrectly (after having answered correctly before) and this participant did not use the free text question to explain why they changed to the incorrect response. Therefore, the data obtained in the study is insufficient to derive improvements pertaining to S1 and no change was introduced in the final version of the password security awareness material.

S9 & S10 (Eavesdropping on unencrypted communication). Both scenarios pertaining to *eavesdropping on unencrypted comunication* showed a potential for improvements. Thereby, S9[11] is the one scenario which clearly stands out in terms of individual changes. While achieving more than 85% correct responses in the pre-treatment questionnaire, performance drops to about 77% in the post-treatment questionnaire and to about 19% in the retention questionnaire. The (non-significant) drop in the post-treatment questionnaire already indicates a need of improvement in the awareness material. However, the steep drop in the retention questionnaire is very alarming. While only few participants made use of the possibility to explain their choice in the free text question, the few responses of this scenario from the post-treatment questionnaire still offered one potential reason[12]: the open wifi in the scenario was equated with a generally insecure wifi. Consequentially, the respective formulation was reworked for the final version of the password security awareness material to better explain what behaviour in open networks is secure (using encrypted connections) and which behaviour might lead to security risks. The need for improvement with respect to scenario S10[13] is less pronounced and only one participant who answered incorrectly used the opportunity to justify their answer. While this special case pointed toward an ambiguity in one specific formulation, the data obtained from the study does not provide a broad reason for the incorrect answers. Consequently, the only change in the final version of the material pertaining to scenario S10 regards this one ambiguity.

S17 (Guessing after a break-in). The free text answers pertaining to S17[14] showed one clear theme: the concatenation of words was not perceived as leading to secure passwords, even when more than 20 characters long. Instead, participants named dictionary attacks as possible way to guess the password. Therefore, an additional explanation explicitly mentioning long passwords created by concatenation of words as well as explanations outlining the limitations of dictionary attacks were added to the final version of the password security awareness material.

[11] Text of S9: Mr. Schmidt is on his way to a client. Unfortunately, the train is delayed. Therefore, he sits down in a café at the train station. There he uses an open wifi. He uses his laptop as usual, but pays attention that he visits all websites using an encrypted connection.

[12] (Analogously to S11, no participant filled the free text question pertaining to S9 in the retention questionnaire.)

[13] Text of scenario S10: Mr. Schmidt is on a business trip visiting a client in a different city. There he stays in a hotel and uses its charged premium unencrypted wifi to work in his room. To login to the wifi, he has to enter a user name and a password.

[14] Text of S17: Mr. Schmidt is frequently on business trips and once he had his laptop almost stolen at the airport. Therefore, he encrypts its hard drive and chooses to encrypt the hard drive and to login a password with more than 20 characters which he creates concatenating multiple words to one another.

S19 & 20 (Theft of a digital note of the password). With respect to scenario S19[15] the same theme as for S17 emerges: the participants did not believe a password created by concatenating words to a length of more than characters was secure. Therefore, a reference to the explanations made for S17 was added to the final version of the password security awareness material. Unfortunately, for scenario S20[16] the data collected in the study did not provide a reason for the lack of significant improvement. Consequently, no improvements with respect to S20 were added to the final version of the password security awareness material.

S21 (Exploiting a weak reset-mechanism). Scenario S21[17] is (beside S1 and S9) the last scenario not scoring 100% correct answers in the retention. Unfortunately, none of the participants answering to this scenario incorrectly used the opportunity to explain their answer in the free text question. Consequently, no improvements could be derived from the study results with respect to S21.

Password Security Ratings

The participants' ability to rate the security of passwords improved significantly after reading through the material (H_{2a} supported). This improvement is retained even after six months (H_{3b} supported). Yet, participants seemed to be hesitant to rate the security of long passwords composed of multiple concatenated words correctly, which reflects the findings with respect to the assessment of scenarios outlined in section 4.5.1 and also findings reported in the literature [186]. In contrast, the participants could significantly better identify insecure passwords. Interestingly, this difference in the ratings is significant in all three measurements (i.e. pre-treatment, post-treatment, and retention). This seems to indicate that the awareness material was more effective in improving the ability of the participants to recognise insecure passwords than the ability to recognise secure ones. Therefore, one focus of further improvements of the awareness material must be the inclusion and teaching of good creation strategies. As a first step, additional explanations and examples were added to the respective sections of the material as outlined in section 4.5.1.

Clearing Up the Misconceptions

The prevalence of the different misconceptions varies greatly. Despite being reported in the literature review, several misconceptions did not seem to be prevalent in the sample, as evidenced by the ceiling effects found for M9, M13, M15, M16, and M19. Consequently, in awareness materials with space constraints or which are distributed in environments with severe constraints on the time and effort spent working through awareness materials, these five misconceptions could potentially be left out. In particular, in such cases the questionnaire items presented in section 4.3.3 could be used to measure prevalence of all misconceptions among the individuals to which the password security awareness material is to be distributed and only include those misconceptions in the awareness material that appear prevalent.

[15] Text of S19: Since Mr. Schmidt has problems remembering the many passwords he needs for his job, he asks the IT-department whether they can install a password manager on his work laptop. Since he wants to synchronise the passwords to his business smartphone, he chooses a master password with more than 20 characters, which he creates by concatenating multiple words.

[16] Text of S20: Mr. Schmidt has to use many passwords in his daily job to log on to all the different systems needs to access. Since he also works on his business smartphone, Mr. Schmidt saves all the passwords in a Word document and synchronises this document through a public third-party cloud storage provider between his laptop and his smartphone.

[17] Text of S21: Mr. Schmidt uses different external web services, as is usual in his company. For one of the web services, the password can be reset using personal security questions. Instead of answering the questions truthfully, Mr. Schmidt chooses a random character sequence as answers, writes this sequence down, and stores it where only he can access it.

Overall, the developed interventions significantly reduced the prevalence of the misconceptions (H_{3a} supported). This overall effect still holds in the retention session six months after the original treatment (H_{3b} supported). However, the results regarding multiple misconceptions warrant closer inspection. When looking at the prevalence of the individual misconceptions, most of the misconceptions not affected by ceiling effects (15 out of 18) could be cleared up by the interventions, leading to a significant increase between the pre-treatment questionnaires and the post-treatment questionnaires. This changes only slightly in the retention questionnaire, where five misconceptions do not show a significant improvement anymore: M5, M11, M17, M21, and M23. However, this effect can be attributed to the different sample in the retention for four of these misconceptions: for M5, M17, M21, and M23 ceiling effects occur when only SME-r is considered in the pre-treatment questionnaire. In contrast, M11[18] proves more problematic. Despite showing an improvement from the pre-treatment to the post-treatment questionnaire, the employees of SME-r perform significantly worse in the retention questionnaire. Unfortunately, the data collected during the study does not provide reasons for this observation, since no free-text questions regarding the misconceptions were included in the questionnaire. Therefore additional studies are needed to investigate M11 and its intervention in more depth and future studies should include the respective free text questions analogously to the scenarios.

The analysis also clearly indicated a need for improvement for some of the misconceptions. Firstly, the intervention texts for the three misconceptions M12, M14, and M20 did not result in a significant improvement with respect to the prevalence of the respective misconception. Secondly, despite exhibiting more than 90% of correct responses in the post-treatment questionnaire, the two misconceptions M9 and M19 misled more participants to an incorrect answer after the treatment, than they resulted in additional correct answers. This as well indicates a need for improvements. Finally, the (in comparison to the other misconceptions) relatively high number of incorrect responses in the post-treatment questionnaire regarding misconception M2 also indicate a need for improvement of the intervention. In the following, each of the aforementioned six misconceptions is discussed individually in ascending order (i.e. M2, M9, M12, ...).

M2: The inclusion of symbols anywhere in passwords makes them automatically more secure. While M2 exhibits a significant improvement from the pre-treatment questionnaire to the post-treatment questionnaire, it reaches only 51.1% of correct responses. This is the lowest value among all misconceptions. Consequently, further refinements are required. The following wording with more concrete examples – based on the studies reporting this misconception and common mangling rules used with password cracking software[19] – was used in the final version of the password security awareness material:

Using specialized software, attackers try to mimick human behaviour when guessing passwords. They use a long list of words (from dictionaries, but also passwords from past breaches) and apply different common modifications to these words to generate additional words they will also use as guesses for the password. Such modifications are:

- *Appending or prepending numbers and symbols (e.g. adding an "!" to the end is a popular choice). Examples of passwords that can be guessed very quickly using such rules are "brooklyn16", "bubblegum1!", "1proudmom" or "Mamamia!!!".*

[18] Text of M11: Internet browsers often have an integrated password manager, which allows saving passwords entered on websites. Saving passwords in a browser is the same as saving them in a dedicated password manager. If the hard drive of your device is not encrypted and your passwords are saved in a password manager, which is not protected by a master password (no matter if in a browser or as dedicated program), attackers can easily copy your passwords off your hard drive, if they have physical access to it or your device is infected with malware.

[19] Based on the winning contribution to the Best64 Challenge, aiming to find the most effective rules: `https://hashcat.net/forum/thread-1002.html`

- *Substituting letters with numbers (e.g. E → 3) or with symbols (e.g. a → @). Examples of passwords that can be guessed very quickly using such rules are "p@ssw0rd", "L0vemetal", "m0nkeyl10n", or "4n4belle".*
- *Substituting lowercase letters with uppercase letters (in particular at the beginning of words). Examples of passwords that can be guessed very quickly using such rules are "pAsswOrd", "Thisismypass", "NOTSOSURE", "daywalker", or "Lovemetal".*

Therefore, using numbers, symbols, or uppercase letters in your password will not automatically make it harder to guess.

M9: Notes of passwords do not need to be particularly protected. M9 misled seven participants (i.e. seven participants having answered correctly in the pre-treatment questionnaire answered incorrectly in the post-treatment questionnaire). However, all five participants who answered incorrectly in the pre-treatment questionnaire, changed their answer to the correct one in the post-treatment questionnaire and the portion of correct post-treatment responses was 92.1%. Furthermore, the decrease in correct answers between the pre-treatment questionnaire and the post-treatment questionnaire was not found to be a significant difference. Thus, it seems that more testing of the intervention is required, before any further changes should be proposed. Consequently, no changes were introduced into the final version of the password security awareness material with respect to M9.

M12: Keyboard patterns are secure passwords. The analysis did not show a significant increase in correct answers for M12. Additionally, it exhibits the largest number of misled participants of all misconceptions (i.e. 10 participants), although the number of participants having answered incorrectly in the pre-treatment questionnaire and correctly in the post-treatment questionnaire is still larger (i.e. 16 participants). The intervention text already included an explanation of the misconception and how an attacker can use it against the participants. To increase the clarity of this misconception's formulation, the following rephrased version is used in the final version of the password security awareness material:

Using walks or patterns on your keyboard as password (e.g. "1QAY2WSX" or "qwertz") is no good practice to generate secure passwords. As a matter of fact, such patterns will be present in the dictionary of every attacker trying to guess passwords. Therefore, you should never use keyboard patterns as passwords, even if they contain uppercase letters, numbers, and symbols.

M14: Attackers do not automate their attacks on passwords, but perform them by hand. The analysis did not show a significant increase in correct answers for M14 as well. Due to the fact that this misconception applies to many possible attacks on passwords and user accounts, the intervention was formulated in an abstract manner with just two concrete examples. In order to make the underlying problems more tangible for lay-users, the following refinement adding an example and a more concrete wording was used in the final version of the password security awareness material:

Attackers can easily automate their attacks and do not have to perform attacks manually by hand. Specialized software to e.g. test billions of different passwords in just one second after a successful break-in, build phishing websites that look just like the original, or snoop on passwords in unencrypted network traffic are readily available.

M19: It is not necessary to set a password to lock the screen of unattended devices. The results with respect to M19 were very surprising: before the treatment all responses were correct, but after the treatment some participants changed to incorrect responses. This indicates that there might be an issue with the intervention. However, since the decrease in correct responses was not found to be a significant difference and the portion of correct post-treatment responses was 94.4%, it seems that further investigations are required before further refinements to the intervention text can be proposed. Consequently, no changes with respect to M19 were introduced into the final version of the password security awareness material.

M20: The passwords protecting work accounts have lower security requirements because the IT staff is responsible for security. M20 did not show a significant improvement after the participants read the intervention texts. The important aspect with this misconception is that users realise that they themselves play a vital role in protecting their devices even at work (e.g. checking attachments for potential malware). Thus, the following proposed improvement with a greater emphasis on the users role was used for the final version of the password security awareness material:

You are responsible for the security of your devices, even at work when there is a dedicated IT department. The staff of the IT department can help you, but in the end it is your job to keep the device secure. Many attacks on organizations target the employees first, in order to get access to internal systems. Therefore, you and the other employees are an important line of defence for your organization.

Further Improvements

The qualitative answers show that the participating employees desire additional concrete advice with respect to two aspects: (1) how to create secure passwords and (2) password managers. The former should be addressed in the final version of the awareness material to some degree by the information added with respect to the strategy of concatenating multiple words as outlined in section 4.5.1. Yet, a section dedicated to password creation strategies should be considered as future work. Need for the latter (which was not addressed in the final version version of the awareness material and instead remains future work) is also supported by the large number of participants who stated that they would start using a password manager after having worked through the material. This uptake of password managers is a positive effect, as usage of password managers is widely considered a good advice to lay-users by information security experts [109,195]: even when not directly changing all passwords to unique randomly generated ones, password managers can help users mitigate shoulder-surfing and phishing attacks by auto-filling passwords only on the correct domain in the browser.

4.5.2 Limitations

The participant sample represents the first larger limiting aspect of the study presented in this chapter. Firstly, six participants had to be excluded from the study due to their apparent thorough knowledge of information security, despite the recruiting done by contact persons in each SME who were instructed to make sure all participants they selected were lay-users. Future studies should not rely on the ability of such contact persons alone, but rather include additional check items such as the RSeBIS [179] or the HAIS-Q [158]. Albeit, depending on the country in which the study takes place a validated translation might not be readily available. Secondly, all participants are employed in German SMEs. Consequently, it is unclear whether the findings of this study fully translate to different groups of users and to different countries. Therefore, future work should include the validation of the results in various contexts. Thirdly, recruiting participants was a major hurdle. Concerns in the SMEs focused mostly on the potential impact on productivity, due to the participants filling out the questionnaires in their working environment. This was a huge factor deterring SMEs from participating in the study, ultimately deterring two of the three SMEs from taking part in the retention session after six months. While addressing this issue is difficult, since thorough investigations rely on thorough questionnaires, any researcher planning similar endeavours is advised to consider this issue. As a result of the low number of participants, testing H_{1a} and H_{1b} (as well as H_{1c} and H_{1d} respectively) had to be done with the same participants, potentially introducing a statistical bias into the study. Future studies should try to minimise this bias by recruiting enough participants to test these hypotheses with distinct groups of participants receiving all scenarios in all questionnaires, but receiving questionnaires only at one measurement point per group (i.e. pre-treatment, post-treatment, retention). As a second result of the relatively low number of participants in the sample, it was not possible to include a control condition in the study design, although this would have allowed to present stronger evidence with respect to the effectiveness of the awareness material and whether the questionnaires have an influence on later measurements (i.e. whether there is a so-called carryover effect [47] and participants improve in later questionnaires due to having thought about questions from previous questionnaire, having discussed them with others, etc.).

Also, the evaluation pertaining to H_2 (i.e. the assessment of the security of passwords) and H_3 (i.e. clearing up the misconceptions) is based on a less complex study design than H_1: all items were shown in all three questionnaires (i.e. pre-treatment, post-treatment, retention). This enabled a more straight forward statistical analysis and rendered the study design for these two hypotheses more approachable. Yet, since the analysis of H_1 found a difference in performance between known and unknown items, future studies should (a) either apply a study design similar to H_1 to the other two hypotheses as well H_2 and H_3 or preferably (b) follow the study design with a fully between subject design as described in the last paragraph.

The second larger limiting aspect of this study is introduced by following the recommendation of Haeussinger and Kranz [91] to conduct the study in the real work environment of the participating employees in three SMEs. The most reliable option in this regard would have been to monitor the password-related behaviour of the employees in their organisation and check whether the awareness material influences this behaviour [67, 203]. However, such a design has severe issues in a real world setting in organisations: Gathering the necessary data might pose security risks (e.g. collecting passwords created by the employees to test whether creation strategies change after the treatment) or have legal and privacy implications (e.g. surveilling employees at their desks in order to see whether they store notes of passwords insecurely). Consequently, a different study design was necessary, allowing to retain the anonymity and privacy of the participants as well as avoid any security risks, while delivering the study materials to the participants' real work environments through a contact person in each SME. As a result, the participants were unsupervised throughout all three stages of the study, i.e. filling the pre-treatment questionnaire, reading through the interventions, and filling the post-treatment questionnaire. Therefore, a number of limitations arise: (a) participants might have used the material while filling out the questionnaire, (b) participants might have filled out the post-treatment questionnaire after reading the material only partially, (c) participants who work in the same SME might have worked (partially) together, and (d) participants might have spent very different amounts of time reading through the material which might have impaired consistency not only between organisations, but also between the participants of each organisation. To counteract these issues, participants received instructions during each of the four phases in the study. Participants were instructed to read the awareness material carefully and in its entirety. Also, they were told that the post-treatment questionnaire would be sent out only after they had explicitly acknowledged having read the entire material. Last but not least, participants were instructed to fill out the questionnaire by themselves.

Another limitation relates to the high baseline in the participant sample. Even before working through the awareness material, over 88% of the responses pertaining to the scenarios were correct. Arguably, two aspects might have contributed to this finding. On the one hand, despite being developed with feedback from independent experts, multiple scenarios exhibited ceiling effects, rendering them not well suited for the assessment of the awareness material's effectiveness. This might indicate a methodological problem resulting from misaligned difficulty levels of the scenarios, not uncovered during the pre-tests. In particular for scenarios S1, S4, S5, S6, S7, S8, S11, S14, and S15 more than 95% of participant responses were correct even before the treatment. Therefore, an important line of future work is the development of improved, more difficult scenario questions. On the other hand, the participants in the sample might have already received education with respect to password security before the study. To investigate both aspects, studies with more difficult scenarios and different samples are needed.

Even though the study was conducted in the employees' real work environment, it might be that the participation in the study has motivated the employees more than they would have been otherwise. In particular, the awareness material's effectiveness outside the study setting and whether the significant improvements found in the study lead to more secure behaviour might depend on how the awareness material is presented to the users and what information is provided alongside it.

Currently the methodology as presented in section 4.3 mostly includes feedback from information security experts and lay-uers on the material. While experts from the field of psychology gave feedback on the questionnaires and on the interventions, no expertise other than information security was used for feedback on the description of the attacks and defences. Future development efforts of awareness materials might benefit from including additional feedback of experts from fields such as psychology, education science, or visualisation and design on the complete awareness material.

Furthermore, the awareness material currently includes only the following technologies: password managers, two-factor authentication, fingerprint readers, graphical passwords, hardware tokens, privacy filters, and single-sign-on. Future versions of the awareness material might need to include additional alternatives to text passwords should they gain widespread adoption (e.g. Face ID [8] or palm vein authentication [24]).

Last but not least, the set of misconceptions covered by the interventions might change over time. In particular, research published since the systematic literature review presented in chapter 3 might report on new misconceptions. To gauge the magnitude of this issue, a forward search based on the 20 relevant publications identified in chapter 3 was conducted using the Google Scholar index. The same exclusion criteria as applied in chapter 3 were used and only publications in the time frame since the original systematic literature research (i.e. since 2017) were considered. Overall 344 publications were found using the forward search (an additional 143 were excluded due to the exclusion criteria). Of those, 23 contained information with respect to misconceptions of users regarding password security. This seems to imply an uptake in research on misconceptions when compared to the original systematic literature review. With respect to the results, on the one hand, several of these 23 publications provide further evidence for the prevalence of different misconceptions identified before (M1/M2/M3: [90,92,161,205,207]; M4/M5: [92,207]; M6: [92]; M7: [50,77]; M8: [153]; M10: [90]; M14: [84]; M15: [79]; M16: [153]; M19: [37,142]; M23: [79,84]). On the other hand, some publications provide evidence contradicting the prevalence of two of the previously identified misconceptions (M5: [137]; M6: [126,134,161]). This implies conflicting evidence for these two misconceptions (M5 and M6) arising from the forward search. Also, both were prevalent among the participants of the study presented in this chapter. Consequently, further investigations into these misconceptions might prove a valuable field for future research. In addition to the existing misconceptions, the aspect of password reuse arose in a more differentiated form from the literature review [46, 83, 83, 84, 90, 106, 144, 196, 220], e.g. Stobert and Biddle [196] find similar reuse rates for lay-users and experts, Choong et al. [46] report that a substantial fraction of school children believe it is viable to "use the same password for everything", and Golla et al. [84] report on the aspect of users not believing that leaked reused passwords would be used in automated attacks and that users believe changing all affected passwords to the same new password is a viable defence strategy. These findings expand M14 to the aspect of password reuse, but might also be seen as additional misconceptions. Another aspect which could be considered to be an additional misconception pertaining to password reuse was reported by Habib et al. [90], who found that users believe it is a viable strategy to reuse passwords between work and private accounts. Beside the area of password reuse, additional misconceptions about password security seem to have emerged in the literature: Hayes et al. [96] as well as Luna [133] report that users might think that a lack of memorability in passwords generally indicates strength, Seitz and Hussmann [186] report that users believe passphrases to be less secure than they are (which is in line with the results presented in this chapter), and Habib et al. [90] find that users only think about guessing attacks when it comes to safeguarding passwords. Additionally, beside the area of text passwords, Zimmermann and Gerber [239] find that users might believe that fingerprints cannot be stolen from them in same way as text passwords (despite media reports about fingerprint forgeries based on photos). These results of the forward search show, that regular updates to the awareness material are necessary and that future revisions of the password security awareness material might need to adjust its scope by covering misconceptions about other authentication schemes as well.

4.6 Conclusion

This chapter presented the development of a correct, complete, understandable, and effective password security awareness material along a systematic process. It addresses all attacks deemed relevant by the literature and independent information security experts from academia and industry and increases lay-users' ability to correctly assess (a) whether specific password-related behaviour in different information security scenarios is secure or insecure and (b) the security of passwords. Furthermore, it significantly decreased the prevalence of misconceptions about password security. In particular, these abilities are retained or even improved six months after reading through the developed password security awareness material. At the same time, the awareness material was received positively by all participants.

The results of this study also point out areas for future work. The participating employees expressed the desire to learn more about password managers and the composition of secure passwords. Thus, it might be warranted to expand the awareness material accordingly or create additional awareness materials for these topics. Also, it might be worthwhile to investigate how to make the transition towards using a password manager easier for users.

The systematic development process presented in this chapter and applied to create the password security awareness material can help create other awareness materials beyond the password context. It could be easily applied to other information security contexts and other target audiences. Therefore, applying the process in areas other than password security represents another important line of future work.

Part II

Shoulder-surfing Resistant Text Password Entry on Gamepads

5 Requirements and Status Quo of Authentication in the Gamepad Context

Gamepad-driven devices such as game consoles are an important part of many people's lives. A 2017 report by the entertainment software association found that about half of all American households own a dedicated game console [68]. It is common to use accounts for e.g. video streaming like Netflix, music streaming like Spotify, or game networks like Playstation Network on consoles. The text passwords protecting these accounts are entered on game consoles almost exclusively using *on-screen keyboards* in combination with gamepads as input devices, which are far more constrained (e.g. regarding the available buttons and input precision) than the traditional combination of mouse and keyboard. At the same time, for many users time spent on their consoles is also a social activity and therefore occurs in shared spaces: 53% of the users play on average five hours with others in person per week (as opposed to online multiplayer games) [68]. Considering that Renaud et al. [169] found in their survey that 90.9% of their participants would authenticate when not alone, opportunistic shoulder-surfing [227] is a real threat, leaving users in a dilemma: either show mistrust of people by asking them to look away [58], behave insecurely by letting them observe, or store the password on the device, which enables purchases by every person with access to the device.

To address the challenge of shoulder-surfing resistant text password entry on gamepads, this chapter presents the first investigation of this topic. To that end, it first describes the requirements of authentication which specifically apply to the gamepad context (section 5.1) and all authentication schemes must fulfil in order to be deemed suitable for the gamepad context. Overall six requirements across the three categories security, technical, and usability are identified. One of the defining requirements of this scenario is the resistance to opportunistic shoulder-surfing. It is important to note though that these requirements specifically applying to the gamepad context must be fulfilled in addition to any general requirements [26] that might are from a specific application scenario in the gamepad context (e.g. resistance against guessing attacks [27, 119]),

Then, as second step building on the identified requirements, the authentication schemes currently deployed in the gamepad context as well as a representative set of shoulder-surfing resistant schemes proposed in the literature are assessed along the requirements. The results of this assessment (section 5.2) show that none of the currently deployed and only four of the proposals in the literature fulfil all six requirements. From the discussion of these assessment results (section 5.3) it becomes clear that the grid-based scheme by Kim et al. [123] is the proposal which seems to be best suited for being adapted from a non-gamepad context to the gamepad context. Form the work presented in this chapter two important next steps arise which are then addressed in the next chapter: (1) an empirical assessment of the baseline performance in terms of shoulder-surfing as well as usability of the on-screen keyboard (as the de-facto standard in the gamepad context) and (2) an empirical evaluation of alternatives to the on-screen keyboard. Section 5.4 concludes this chapter.

Contributions described in this chapter:

- The requirements of authentication in the gamepad context are identified and outlined. They can be grouped into three categories: security, technical, and usability. These requirements can inform the design of new authentication schemes in the gamepad context.
- The authentication schemes currently deployed in the gamepad context as well as a representative set of shoulder-surfing resistant authentication schemes proposed in the literature are assessed along the identified criteria.

Parts of the results described in this chapter have been published in:

- P. Mayer, N. Gerber, B. Reinheimer, P. Rack, K. Braun, and M. Volkamer, "I (don't) see what you typed there! Shoulder-surfing resistant password entry on gamepads", Conference on Human Factors in Computing Systems (CHI), 2019.

5.1 Requirements

This section describes the requirements that specifically apply to authentication in the gamepad context. To ease referring to the individual requirements in the remainder of this part of the thesis, they are each assigned both, a numeric reference of the form *Rx* (where *x* is the number) and a short mnemonic descriptor. The requirements are grouped into three categories: security, technical, and usability.

5.1.1 Security Requirements

Providing a shoulder-surfing resistant text password entry for usage with gamepads is the core motivation behind the work presented in this part of the thesis, since it represents a key requirement of the gamepad context as explained in the remainder of this section.

Using gamepad-driven devices such as game consoles is for many users a social activity: 53% of users play on average five hours with others in person per week (as opposed to online multiplayer games) [68]. Therefore, usage of these devices occurs in a so-called shared space [169]. Another defining aspect of authentication on these devices is that they are usually used in conjunction with large displays such as TVs. Such large displays have been found to be perceived qualitatively different to smaller displays [199]. Tan and Czerwinski [202] found that users were more likely to read sensitive content on large screens and note that since such devices are usually outside a user's "personal zone", they might be perceived as less private. Yet, research indicates that 90.9% of users would authenticate in such a setting when not alone [169]. Together, these aspects indicate a large potential for shoulder-surfing threats. Due to the threat model of usage in shared spaces, the opportunistic observer (cf. section 2.3.1) is the most likely attacker in the gamepad scenario, since any recording of the authentication procedure by a friend sitting on the user's couch right next to them is likely to draw attention. Since the definition of opportunistic observers [227] does not set a specific threshold to achieve resistance, but rather states that the attacker has access to only "a small number of observations", the threshold is assumed to be "not more observations than there are characters in the password"[1]. In conclusion, the requirement regarding shoulder-surfing resistance is:

> ***R1:*** *Authentication schemes used on gamepad-driven devices must resist shoulder-surfing attacks by opportunistic observers, i.e. resist observers with not more observations than there are characters in the password they want to observe. (resistant-to-opportunistic-observers)*

5.1.2 Technical Requirements

In addition to the security requirement *R1*, several technical requirements must be fulfilled by authentication schemes in the gamepad context. Gamepads are very constrained input devices. In comparison to a keyboard, gamepads offer far less buttons. However, they often also have output capabilities, such force-feedback through vibration motors. Figure 5.1 depicts the input capabilities of a typical gamepad as considered for this work. Modern gamepads are usually equipped with two small joysticks commonly referred to as analogue sticks. They are used to capture directional input (i.e. input that moves something on the screen) and are usually operated using the thumbs (i.e. one is operated with the left thumb one with the right thumb).

[1] It is necessary to acknowledge that this threshold is somewhat arbitrary.

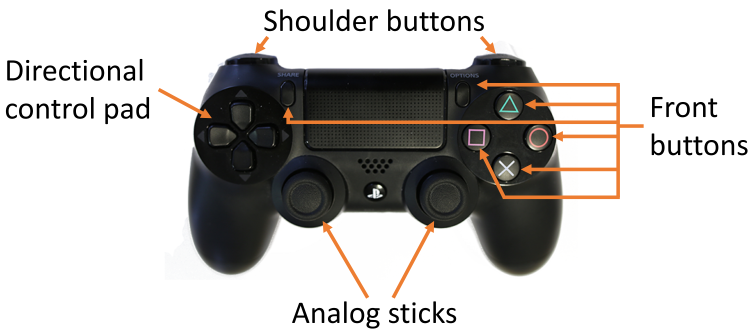


Figure 5.1: The controls available on a typical gamepad.

In addition, gamepads have buttons which can be pressed. Firstly, there is one array of four buttons representing a directional control pad. It consists of a cross-shaped rocking button with actuators underneath and its working principle is that of a switch-activated joystick (minus the stick part) [10]. Secondly, there is usually an array of four buttons on the other side of the gamepad. Last but not least, there are four so-called shoulder buttons, which are positioned on the top of the gamepad. On the majority of modern gamepads two of these shoulder buttons (one on the left side and one on the right side) are implemented as analogue triggers (i.e. one dimensional analogue controls).

The technical requirements reflect these available controls of gamepads as well as their sensitivity and precision. In particular, authentication schemes in the gamepad context must not require more different buttons than available on a gamepad:

> ***R2:*** *Authentication schemes used on gamepad-driven devices must not require as controls more than eight freely programmable buttons, one directional control pad, two analogue sticks, and two analogue triggers if they are to be compatible with the gamepads of most modern gamepad-driven devices. (compatible-with-gamepad-controls)*

The accounts used on gamepad-driven devices usually require text password authentication (e.g. common services such as Netflix or Xbox Live). Therefore, remaining compatible to these accounts is of the essence. Such compatibility is trivial if the authentication scheme is explicitly designed for text password entry. Otherwise compatibility can be achieved e.g. through an extension of a traditional on-screen keyboard, a direct mapping of the characters to the gamepad controls, or scaling up the procedures of the schemes. However, a crucial aspect to compatibility with test passwords is that no changes to the backend of the verifier are necessary. The respective requirement is:

> ***R3:*** *Authentication schemes used on gamepad-driven devices must be compatible with text passwords. (compatible-with-text-passwords)*

Most operating systems of gamepad-driven devices do not allow installation of drivers for additional hardware. Therefore, the following requirement arises:

> ***R4:*** *Authentication schemes used on gamepad-driven devices must not require support for additional hardware such as biometric readers or token devices. (no-additional-hardware)*

5.1.3 Usability Requirement

As is the case with any controls, the interaction with the gamepad is bounded by the capabilities of the human interacting with it. Those restrictions must be followed or the user might not be able to engage in the authentication procedure. Consequently, the layout of a typical gamepad (see Figure 5.1) in conjunction with human anatomy poses restrictions on the controls which can be used simultaneously to enter a password. For example, all controls on the front of gamepads (i.e. the analogue sticks, directional control pad, and front buttons) are actuated using the thumbs when the gamepad is held as intended (i.e. only two controls on the front can be actuated at the same time). This is reflected by the following requirement:

> ***R5:*** *Authentication schemes used on gamepad-driven devices must be operable within human anatomic constraints*[2]*. (anatomically-compatible)*

Furthermore, the analogue movement input controls used for directional input on gamepads (i.e. analogue sticks) are generally less precise than a mouse or a trackball [69]. This is reflected by the following requirement:

> ***R6:*** *Authentication schemes used on gamepad-driven devices must not require mouse-equivalent precision directional input. (no-mouse-equivalent-precision-input)*

5.2 Assessment of Existing Schemes

This section presents the assessment of both, the schemes currently deployed in the gamepad context as well as existing shoulder-surfing resistant schemes proposed in the literature in non-gamepad contexts[3] (i.e. schemes meeting $R1$), with respect to the requirements outlined in the last section.

5.2.1 Schemes Deployed in the Gamepad Context

Entry of text passwords in the gamepad context is mostly based on traditional text entry methods. Therefore, all schemes for text password entry already deployed in the gamepad context (e.g. on game consoles) fulfil the technical requirements $R2$-$R4$[4] and the usability requirements $R5$ and $R6$. The compromise is made in terms of security, since text entered this way is easily observable. Therefore, in the remainder of this section the deployed schemes' properties regarding $R1$ are examined in detail and explanations provided why the schemes are not likely to meet the security requirement $R1$.

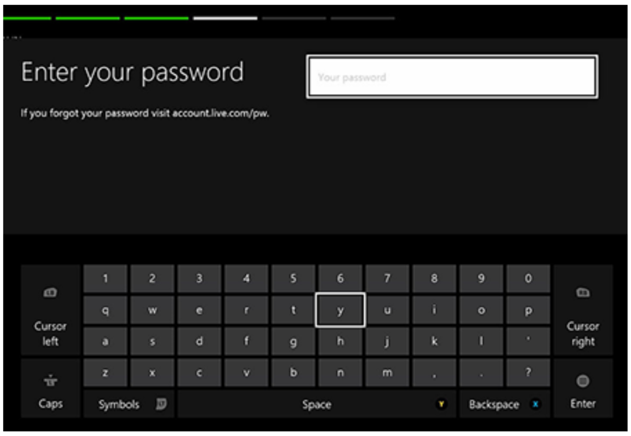
On-screen Keyboards

All current gamepad driven devices (e.g. game consoles) offer *on-screen keyboards* for all text entry, including text passwords. The *on-screen keyboards* consist of a grid of buttons corresponding to the characters which

[2] The measure for this requirement is that the author would be able to actuate all required controls on a gamepad while holding it in the vendor-intended way. While this is somewhat arbitrary, it would be difficult to define and check this requirement based on an "average-sized human hand" instead.

[3] Shoulder-surfing resistant schemes proposed for the gamepad-context in the literature would have been also included, but to the author's knowledge there are none.

[4] There are also deployed schemes with are used specifically for PIN entry. These schemes map each number 0-9 directly to one button of the gamepad. These schemes are not considered here, since they are not *compatible-with-gamepad-controls* ($R2$) when scaled up for a full alphanumeric alphabet.



Enter your password

If you forgot your password visit account.live.com/pw.

Your password

Figure 5.2: The *on-screen keyboard* used on the Xbox game console. The white cursor is currently positioned on the letter “y”.

can be entered (see figure 5.2). By default these buttons display all lowercase letters and numbers as well as frequently used special characters. Uppercase letters and more special characters can be accessed by pressing buttons on the gamepad which serve to switch the characters in the shown grid.

When used with a gamepad (opposed to e.g. touch input), all *on-screen keyboards* can be controlled by using a cursor. The user moves this cursor (white rectangle in figure 5.2) to the desired character using the analogue sticks or the directional pad and confirms the input by pressing a specific (platform dependent) button on the gamepad. Since the cursor highlights the selected character at any time, observation of the path taken on the on-screen keyboard is relatively easy. Therefore, this scheme is assumed to be highly prone to shoulder-surfing threats (i.e. it does not meet $R1$).

Daisy Wheel

The only other scheme to enter text passwords currently deployed for the usage with gamepads, is *Daisy Wheel*. This scheme is specifically developed for usage with gamepads and is based on a so-called pie menu

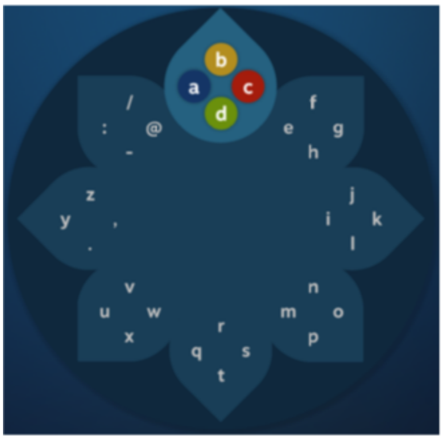


Figure 5.3: The interface of the *Daisy Wheel* scheme. The upper most petal is selected using the analogue stick. The four characters in the selected petal are coloured reflecting the buttons the users needs to press on the gamepad to enter the respective character.

structure [42]. However, it is only available on the Steam[5] platform with a special gamepad (i.e. most users of the Steam platform still use a traditional on-screen keyboard [211]).

Its interface (see figure 5.3) looks like a daisy blossom and contains eight petals, each displaying four different characters. First the petal is selected using one of the analogue sticks. Then, only in the selected petal, the four characters are highlighted in different colours. The desired character on that petal is selected using the array of four buttons on the front of the gamepad (cf figure 5.1), where the colour of the letters reflects the colours of the buttons of the specific gamepad *Daisy Wheel* is designed for. Different sets of characters (e.g., lowercase letters, uppercase letters, special characters) can be accessed analogously to the *on-screen keyboard* by pressing the shoulder buttons on the gamepad and thereby changing the set of vsible characters. Since the selected character in the petal briefly blinks as visual feedback, this scheme seems highly prone to shoulder-surfing threats.

5.2.2 Existing Shoulder-Surfing Resistant Proposals from the Literature

As outlined in section 2.3.1, the literature describes five techniques used to counteract shoulder-surfing threats: covert channels to the user, obfuscation of the user's input through distractors, indirect input, additional biometric layers, and portfolio authentication. In the following, a representative selection of shoulder-surfing resistant authentication schemes proposed in the literature covering all five categories is presented and assessed with respect to the requirements[6] described in section 5.1. Table B.1 in appendix B lists all assessments presented in the following.

Covert channels

Bianchi et al. [19] propose *Secure Haptic Keypad*, an authentication scheme using haptic feedback on a keypad with three vibrating buttons as covert channel. Secure Haptic Keypad is proposed as alternative to traditional keyboards for the entry of passwords. It completely foregoes visual feedback and instead relies completely on the haptic feedback of three vibrating buttons. The buttons vibrate in special patterns (termed tactons) which the user has to identify and subsequently press the button corresponding to the input they want to make. Figure 5.4 shows an overview of the scheme and its components. Since the scheme does not require directional input, $R6$ (no-mouse-equivalent-precision-input) is fulfilled and since it is specifically developed as replacement for alpha-numeric keyboards it is compatible-with-text-passwords ($R3$). $R2$ (compatible-with-

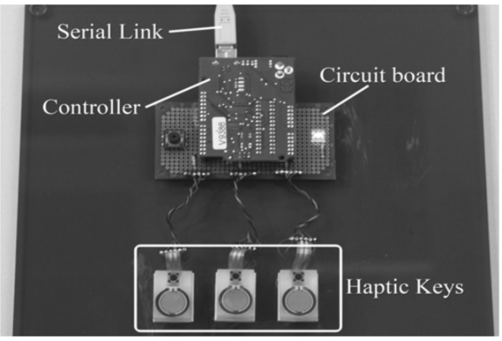


Figure 5.4: Overview of the *Secure Haptic Keypad* scheme as depicted in [19] on page 1090.

[5] https://store.steampowered.com/

[6] Requirement $R1$ is not considered in the following, since all schemes are described as shoulder-surfing resistant in the literature.

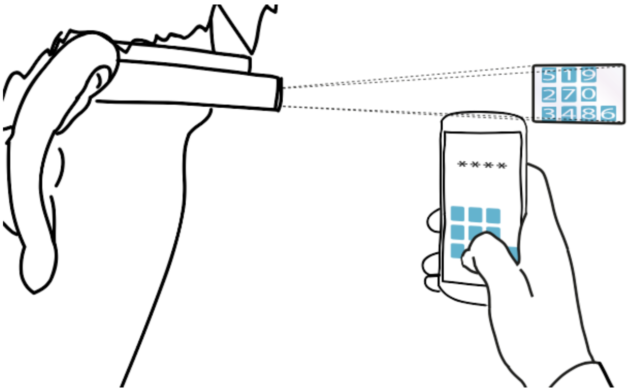


Figure 5.5: Concept of the *Glass Unlock* scheme as depicted in [229] on page 1407.

gamepad-controls) and $R5$ (anatomically-compatible) are technically also fulfilled, as the scheme requires only three buttons. However, while most modern gamepads offer haptic feedback through vibration motors, this feedback is not available on a per button basis as intended for Secure Haptic Keypad, but only for the gamepad as a whole. Additional haptic hardware could remedy this, but would lead to a violation of $R4$ (no-additional-hardware). Therefore, *Secure Haptic Keypad* is not suitable for deployment in the gamepad context.

Winkler et al. [229] propose *Glass Unlock*. This authentication scheme uses smart glasses to display additional information required for successful authentication. Glass Unlock is proposed as alternative for unlocking smartphones and therefore geared towards usage with PINs. In such a scenario, the phone would display a grid of empty buttons and the layout would be visible only on a personal display in the smart glasses. Figure 5.5 depicts this concept. Given a personal display with high resolution, this concept can be adapted for usage with text passwords on gamepads – e.g. as extension of an on-screen keyboard – leading to fulfilment of $R2$ (compatible-with-gamepad-controls), $R6$ (no-mouse-equivalent-precision-input), $R3$ (compatible-with-text-passwords), and $R5$ (anatomically-compatible). However, the smart glasses represent additional hardware which leads to a violation of requirement $R4$ (no-additional-hardware). Thus, *Glass Unlock* is not suitable for usage in the gamepad context.

Krombholz et al. [127] propose the *force-PIN* authentication scheme, which uses the applied pressure in two levels (termed “shallow pressure” and “deep pressure” by the authors) on touch-screens as covert channel. Figure 5.6 illustrates this working principle. This scheme is specifically designed as replacement for PINs on

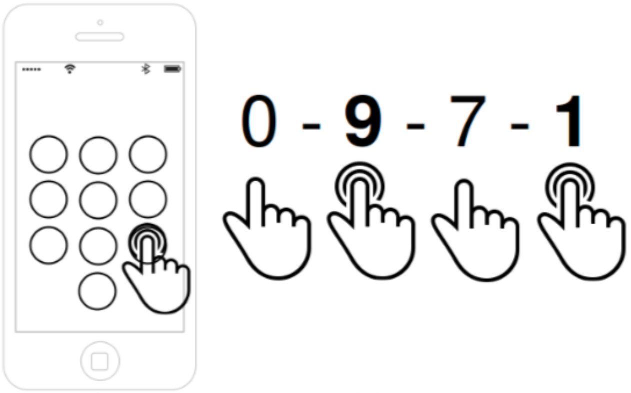


Figure 5.6: Working principle of the *force-PIN* scheme as depicted in [127] on page 207. Normal font digits indicate “shallow pressure” and bold font digits indicate “deep pressure” touch input.

mobile devices but it could theoretically be scaled up to using a full on-screen keyboard in conjunction with pressure-dependant input used on all buttons. While gamepads do not offer touch-screens, they offer other force-sensitive controls in the form of the analogue triggers on the top of the gamepad and the two analogue sticks. These analogue controls could be used to trigger a “deep pressure” input after a certain threshold is reached and a “shallow pressure” input otherwise. Overall this amounts to 18 potentially force-sensitive inputs: the two one dimensional analogue triggers and overall eight directions for each of the analogue sticks. Using not more than eight directions per analogue stick is a conservative choice motivated by the study prototype used in [42] and also the number of directions used by the deployed scheme Daisy Wheel. Increasing the number of directions might easily violate requirement $R6$ (no-mouse-equivalent-precision-input). However, even using more directions on the analogue sticks, the number of analogue controls available on the gamepad would still be less than the number of printable ASCII characters – which can be seen as a lower bound of characters needed to be compatible-with-text-passwords ($R3$) – rendering the scheme problematic with respect to requirement $R2$ (compatible-with-gamepad-controls). Also, using the plainly visible analogue sticks instead of the much more subtle application of force on a touchscreen might render the channel easier to observe and therefore not resistant-to-opportunistic-observers ($R1$). Thus, even while the two requirements $R4$ (no-additional-hardware) and $R5$ (anatomically-compatible) might be fulfilled, *force-PIN* is not easily adapted for shoulder-surfing resistant usage with gamepads.

Sasamoto et al. [176] propose the authentication scheme *Undercover*, which uses a trackball covered by the user’s hand as covert channel to communicate a button layout to the user. The user places their hand on the trackball (on the left of the device in figure 5.7), which is rotated by the system to indicate through its rotation direction which of the five possible button layouts is active (on the right of the device in figure 5.7). Thus, only the user knows which button to press. The Undercover scheme is designed to be used as replacement for PINs at ATMs. The canonical implementation in [176] proposes to use the five buttons as input for a graphical recognition-based authentication scheme. Therefore, to be compatible-with-text-passwords ($R3$), the scheme would have to be adapted by increasing the number of buttons and consequentially the layout displayed on the device has to be scaled up to accommodate a complete keyboard. This might compromise the fulfilment of $R1$ (resistant-to-opportunistic-observers). While recognition-based graphical passwords can be implemented using completely random passwords and will still exhibit favourable usability results [141]. As a matter of fact this is considered best practice due to security considerations [53]. Therefore, user inputs in

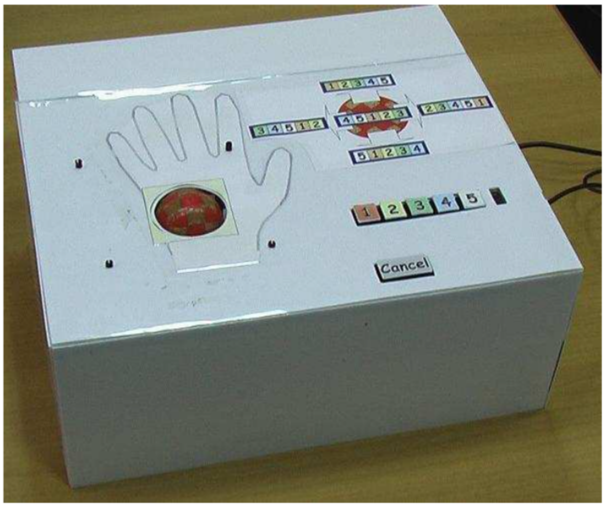


Figure 5.7: Prototype of the *Undercover* scheme as depicted in [176] on page 187.

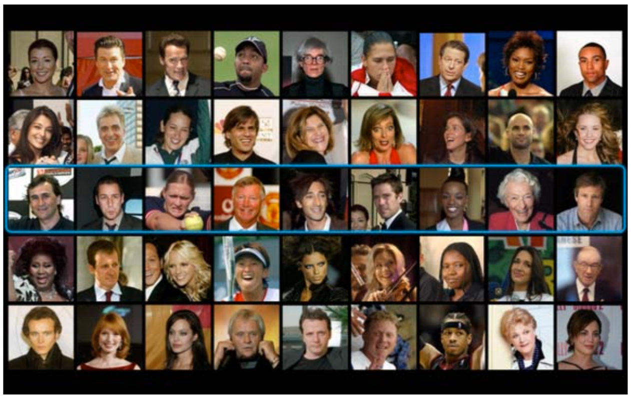


Figure 5.8: Interface of the *Tetrad* scheme as depicted in [169] on page 394. The currently selected row is highlighted with a blue border.

graphical recognition-based schemes are independent of each other. In contrast, users are unlikely to choose random passwords [210], introducing dependence of the individual input. Therefore, having an unconcealed input as in Undercover might allow an attacker to limit the search space of the secret substantially. In addition, gamepads do not offer a trackball to convey the additional information to the user. They only offer haptic feedback, which might also be easier observed (e.g. through the sound of the haptic actuation motors). Therefore, *Undercover* does not seem to be suitable for the deployment with gamepads.

Obfuscation of Input

In the *Tetrad* scheme by Renaud et al. [169] the password is comprised of images. These are placed on the screen in a grid among distractor images and have to be aligned either horizontally, vertically or diagonally (see figure 5.8). Every interaction manipulates whole columns or rows in the grid, i.e. the distractors are moved at the same time, making it hard to discern which images are part of the password. While *Tetrad* is - to the author's knowledge - the only authentication scheme explicitly designed with the limitations of shared space and usage with TVs in mind and therefore easily fulfilling $R2$ (compatible-with-gamepad-controls), $R6$ (no-mouse-equivalent-precision-input), $R4$ (no-additional-hardware), and $R5$ (anatomically-compatible), it is based on a graphical authentication approach, leading to non-fulfilment of $R3$ (compatible-with-text-passwords). Thus, *Tetrad* is not suitable for usage in the gamepad context.

In the *Draw-A-Secret* scheme by Jermyn et al. [112], the user draws their secret on a chequered canvas. Zakaria et al. [233] propose three shoulder-surfing resistant variants of this scheme using different techniques to obfuscate the input. The first variant automatically "draws" distractor strokes while the user is entering the actual password. In the second variant, the password strokes disappear from the screen once the user has finished drawing each one. The third variant builds upon the second, but the strokes disappear even sooner

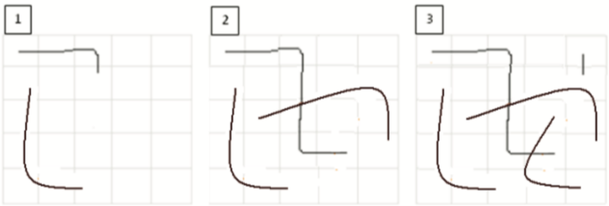


Figure 5.9: Interface of the first variant of the *Draw-A-Secret* scheme by Zakaria et al. as depicted in [233] on page 3. The distractor strokes are here highlighted using the darker colour.

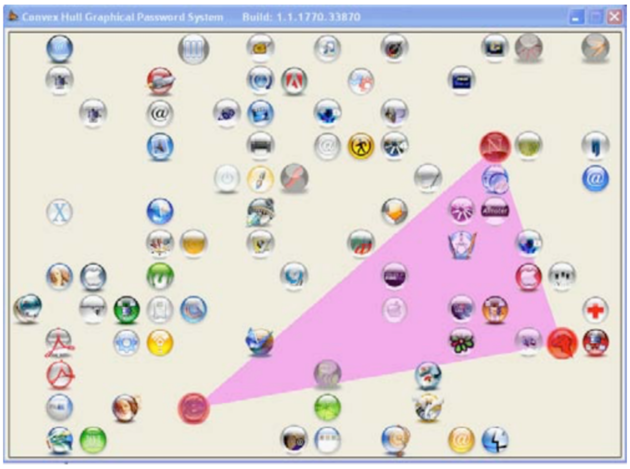
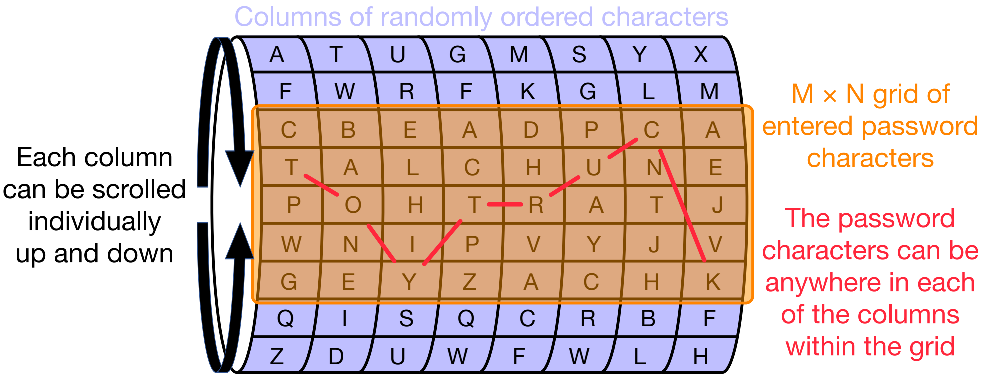


Figure 5.10: Interface of the *Convex Hull Click* scheme as depicted in [226] on page 180.

from the screen: A snaking effect is used where the strokes begin disappearing after a short delay while user is still drawing each stroke. Figure 5.9 depicts the first variant of the scheme. As graphical password, these variants are not compatible-with-text-passwords ($R3$). While it would theoretically be possible to replace the original drawing grid with a grid offering more cells to accommodate a full keyboard and have the user draw strokes from one character in the password to the next (similar to the Swype keyboard [200] for touch-screen devices), this (a) might increase the necessary precision of the input due to the smaller targets in case the size of the drawing canvas is not increased, rendering the scheme unsuitable with respect to $R6$ (no-mouse-equivalent-precision-input) and (b) might reveal the characters of the password on the screen when a change of direction in the swipe is necessary to reach the next character, rendering the scheme not-resistant-to-shoulder-surfing ($R3$). Therefore, these variants of *Draw-A-Secret* are not suitable for usage with gamepads, despite fulfilling $R2$ (compatible-with-gamepad-controls), $R4$ (no-additional-hardware), and $R5$ (anatomically-compatible).

In the *Convex Hull Click* scheme proposed by Wiedenbeck et al. [226], the password is composed of multiple icons. These icons are placed on the screen among distractor icons and the user has to envision the convex hull spanned by the icons of their password and click inside it. As a graphical authentication scheme it is not compatible-with-text-passwords ($R3$). The *S3PAS* scheme by Zhao et al. [237] uses in a similar approach a grid of all available characters (where the characters not comprised in the password serve as distractors) and the user has to click into the convex hull spanned by the characters in their password, thereby being compatible-with-text-passwords ($R3$). Yet, both of these schemes are designed for mouse input and require the user to navigate a cursor on-screen, hindering fulfilment of $R6$ (no-mouse-equivalent-precision-input) when the cursor has top be moved with the analogue sticks. Therefore, *Convex Hull Click* and *S3PAS* do not seem to be suitable for the usage with gamepads.

The *grid-based scheme* by Kim et al. [123] uses a $M \times N$ grid of characters, where N is the length of the password, i.e. there is one column in the grid for each character in the password, and M is the number of characters visible at the same time in each column, i.e. the number of rows. Each column contains all characters in a random order. To login, the user has to scroll into view the character of the password in the respective column (i.e. has to scroll the column until the character is one of the M characters visible for that column). The characters of the password do not have to be aligned in the same row. Figure 5.11 illustrates this concept. The non-password characters visible in the $M \times N$ grid serve as distractors to obfuscate the input. Therefore, the number of rows is essential for the shoulder-surfing resistance of the scheme. A

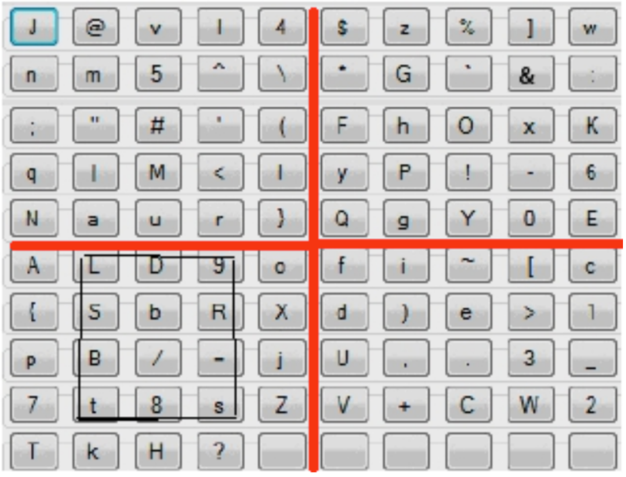

Columns of randomly ordered characters
Each column can be scrolled individually up and down
M × N grid of entered password characters
The password characters can be anywhere in each of the columns within the grid

A	T	U	G	M	S	Y	X
F	W	R	F	K	G	L	M
C	B	E	A	D	P	C	A
T	A	L	C	H	U	N	E
P	O	H	T	R	A	T	J
W	N	I	P	V	Y	J	V
G	E	Y	Z	A	C	H	K
Q	I	S	Q	C	R	B	F
Z	D	U	W	F	W	L	H

Figure 5.11: The basic working principle of the grid-based scheme. The password characters have to be scrolled into the $M \times N$ grid (marked in orange) for each column. Each column contains all characters in a random order. In the case of this example, the password “TOYTRUCK” was entered. The characters are not aligned in one row. Input for this scheme is ambiguous and all combinations in the grid have to be hashed and tested.

peculiarity of this scheme is that the user input is ambiguous and all combinations in the grid have to be hashed and tested (unless the password is stored in the clear, which is contrary to best practices). The scheme is intended for text input and is therefore compatible-with-text-passwords ($R3$) and while originally developed for mouse input, the *grid-based scheme* can be adapted for usage with gamepads in a way that it fulfils all remaining requirements. The buttons intended to scroll the columns up and down are replaced by the y-axis input of one of the analogue sticks or the directional control pad. The column to be scrolled can be selected using the x-axis input respectively, highlighting the currently selected column in colour. Only one additional button is needed to indicate to the system when all characters of the password have been scrolled into the grid. Therefore, the scheme only requires one analogue stick or the directional control pad and one button, thereby fulfilling $R2$ (compatible-with-gamepad-controls), $R4$ (no-additional-hardware), and $R5$ (anatomically-compatible). The simple directional input fulfils $R6$ (no-mouse-equivalent-precision-input). Thus, the *grid-based scheme* scheme can be used in the gamepad context.

The *PairPasswordChar* scheme by Rao et al. [166] uses a 10×10 grid in which all characters available for input are randomly distributed in the grid's cells (see figure 5.12). The centre of the grid is considered to be its coordinate origin for the x-axis and y-axis. As a first step in the login procedure the user has to



J	@	v	l	4	$	z	%	]	w
n	m	5	^	\	•	G	`	&	:
;	"	#	'	(	F	h	O	x	K
q	l	M	<	I	y	P	!	-	6
N	a	u	r	}	Q	g	Y	0	E
A	L	D	9	o	f	i	~	[	c
{	S	b	R	X	d	)	e	>	1
p	B	/	-	j	U	,	.	3	_
7	t	8	s	Z	V	+	C	W	2
T	k	H	?						

Figure 5.12: Interface of the *PairPasswordChar* scheme as depicted in [166] on page 167. Illustrated here is the application of the third rule, where the two characters in the password are ‘L’ and ‘s’.

construct the tuples of each character p in the password and the subsequent character in their mind, i.e. $(p_1, p_2,), (p_2, p_3), ...(p_n, p_1)$ (where n is the length of the password). Then the user has to click[7] into certain areas of the grid determined by the coordinates of the characters in the grid and the following set of rules:

1. If the cells of both characters are in the same column of the grid, the user has to click in the area spanned by those cells and the cells with mirrored coordinates on the x-axis.
2. If the cells of both characters are in the same row of the grid, the user has to click in the area spanned by those cells and the cells with mirrored coordinates on the y-axis.
3. If the cells of both characters are in different rows and columns, the user has to click in the rectangle spanned by the coordinates of those two cells (see figure 5.12).
4. If the two characters are the same, the user has to click in the area spanned by the cell of this character and the cell with mirrored coordinates on the x-axis and y-axis.

PairPasswordChar is intended for use with text passwords and therefore fulfils $R3$ (compatible-with-text-passwords). The navigation in the grid can be implemented in the same manner as for a traditional on-screen keyboard, fulfilling $R2$ (compatible-with-gamepad-controls), $R6$ (no-mouse-equivalent-precision-input), $R4$ (no-additional-hardware), and $R5$ (anatomically-compatible). Thus, the *PairPasswordChar* scheme is suitable for usage in the gamepad context.

Indirect Input

The *cognitive trapdoor game* scheme by Roth et al. [173] separates the keypad for PIN input into two coloured sets and the user has to enter only the correct colour instead of the actual PIN, where for each digit in the PIN multiple such game rounds are performed. The interface of the scheme is depicted in figure 5.13. Since the *cognitive trapdoor* scheme requires only two buttons to be operated, it is compatible-with-gamepad-controls ($R2$) and anatomically-compatible ($R5$). It also requires neither high-precision input nor additional hardware, thereby fulfilling $R6$ (no-mouse-equivalent-precision-input) and $R4$ (no-additional-hardware). When made compatible-with-text-passwords ($R3$), each character of the password would require $n = \lceil \log_2 |96| \rceil = 7$ rounds (when following the descriptions in [173] for the determination of the number of rounds required for each character with $n = \lceil \log_2 |A| \rceil$, where A is the alphabet and assumed to be the 96 printable ASCII characters). Assuming the NIST recommendation [85] of at least 8 characters for memorised

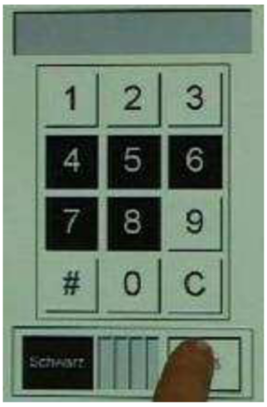


Figure 5.13: Interface of the *cognitive trapdoor game* scheme as depicted in [173] on page 238.

[7] The scheme is designed for mouse input.

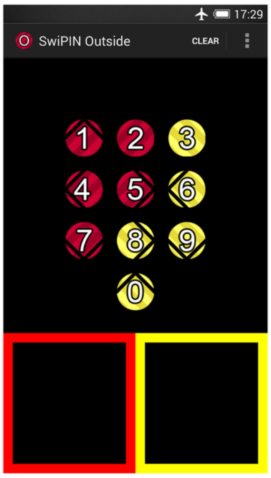


Figure 5.14: Interface of the *SwiPIN* scheme as depicted in [218] on page 1403.

secrets, the user would need to perform 56 rounds of the cognitive trapdoor game to authenticate. Using this adaptation, the *cognitive trapdoor game* fulfils all six requirements.

Similarly, the *SwiPIN* scheme by von Zezschwitz et al. [218] separates the keypad for PIN input into two coloured sets, but the user has to perform a gesture on a touch surface where the starting point is determined by the colour of the number the user wants to enter and the swipe direction (left, right, up, down, none/tap) by the specific number. For each character in the secret, one swipe gesture has to be performed. SwiPIN fulfils requirement $R4$ (no-additional-hardware). For this scheme, issues arise when scaling it up to work with text passwords instead of PINs. When keeping the directional input at four directions plus one button press, an alphabet of 96 ASCII characters would require 20 coloured sets of characters and equally many directional inputs. When using 8 directions (as e.g. used in the *Daisy Wheel* scheme described in section 5.2.1) the scheme would still require 11 directional controls. This is well beyond the threshold of controls available on typical gamepads, thereby rendering it not compatible-with-gamepad-controls ($R2$), and also beyond what users could operate on a gamepad, thereby rendering it not anatomically-compatible ($R5$). Alternatively, when using the two available analogue sticks, 48 directions would have to be mapped to each, thus violating $R6$ (no-mouse-equivalent-precision-input). Therefore, while the scheme theoretically is compatible-with-text-passwords ($R3$) and can accommodate text password input, its implementation would violate other requirements. Therefore, the *SwiPIN* scheme seems not suitable for usage with gamepads.

DeLuca et al. [57] proposed the *XSide* authentication scheme which uses input on both, the front and the back of the device. The secret in XSide is made up of horizontal and vertical strokes on touch surfaces on the front and back of the device. The same stroke cannot be used twice in a row and diagonal strokes are not possible. XSide does not require high precision input (only four directions: horizontal left/right, vertical up/down) and therefore meets $R6$ (no-mouse-equivalent-precision-input). It also does not require two or more controls on the front of the gamepad to be operated at the same time, therefore rendering the scheme anatomically-compatible ($R5$). Adapting XSide to be compatible-with-text-passwords ($R3$) is non-trivial, but could be achieved using techniques similar to the *cognitive trapdoor game* scheme. However, the scheme requires a touch sensitive surface control at the back of the device, rendering it not compatible-with-gamepad-controls ($R2$). Alternatively, $R4$ (no-additional-hardware) is violated, if touch surfaces on additional hardware should be used as complement to the gamepad. Therefore, the *XSide* scheme is not suitable for usage with gamepads.

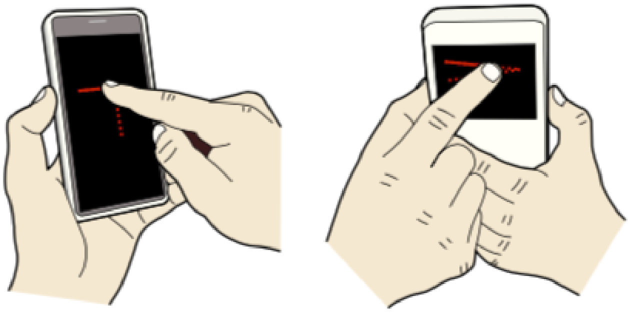


Figure 5.15: Working principle of the *XSide* scheme as depicted in [218] on page 1403. Users can use the touchscreen on the front an the touch sensitive area on the back to enter their secret.

Shirazi et al. [175] found the *MagiSign* scheme [120] to be shoulder-surfing resistant. The secret in this scheme consists of hand motions with a magnet (e.g. ring on one finger of the hand) which are captured by magnetic sensors (see figure 5.16). The scheme does neither require mouse-equivalent precision input nor a lot of different controls (thus fulfilling $R2$ and $R6$). It can be adapted to be compatible-with-text-passwords ($R3$) by assigning different motions to letters and it is also anatomically-compatible ($R5$). However, gamepads do not offer magnetic sensors which would have to be added as additional hardware, leading to a violation of $R4$ (no-additional-hardware). Therefore, the *MagiSign* scheme is not suitable for usage with gamepads.

Additional Biometric Layers

Over the last years, biometric authentication is on the rise in particular due to its market penetration in the mobile sector [191, 192]. However, the presence of biometric sensors (e.g. fingerprint readers) on gamepad-driven devices is rare. The typical gamepad has no such sensors (cf. section 5.1.2). However, *behavioural biometrics* do not need any additional sensors. They identify users based on the characteristics of their input during normal operation, e.g. the speed and variance in inputs [178]. In particular, behavioural biometrics based on the user's specific gesturing patterns [56] and accelerometer based gait data [101] have been proposed. Figure 5.17 illustrates the usage of parameters such as input speed and pressure as additional layer to the Android unlock pattern.

Behavioural biometrics could be implemented based on the user's input characteristics of the controls available on a gamepad during entry of a text password. Such an implementation would need neither additional controls, thereby fulfilling $R2$ (compatible-with-gamepad-controls), nor high-precision directional input, thereby fulfilling $R6$ (no-mouse-equivalent-precision-input), nor additional hardware, thereby fulfilling $R4$ (no-additional-hardware). Since behavioural biometrics are based on the users input during normal tasks they also do not require the users to operate more controls at one time than they can handle, thereby fulfilling $R5$ (anatomically-compatible). Still, the additional biometric information would have to be veri-

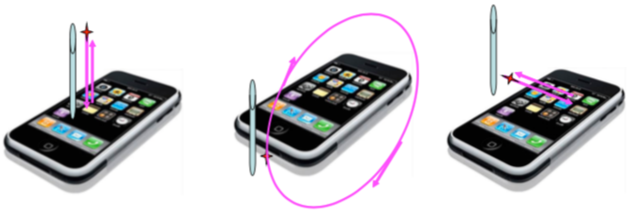


Figure 5.16: Working principle of the *MagiSign* scheme as depicted in [175] on page 4. The magnet required in the scheme is symbolised by the pen.

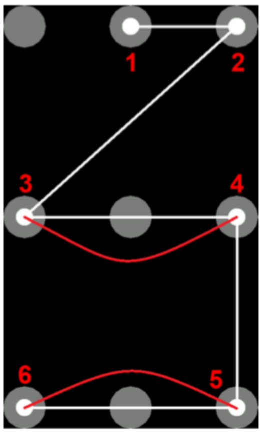


Figure 5.17: Users can be identified by the individual way they enter their Android unlock pattern. Illustration from [56], page 992.

fied, thereby not being compatible-with-text-passwords ($R3$). Therefore, the *behavioural biometrics* are not suitable for usage with gamepads.

Portfolio Authentication

DeAngeli et al. [54, 55] first proposed the portfolio authentication approach as a higher security variant for recognition-based graphical passwords. These schemes have the same basic working principle as usual recognition-based authentication. The password is a set of images. During authentication, a grid of images is shown and the user has to select the image(s) which belong to their password (this selection process is potentially repeated multiple times, depending on the concrete implementation). The difference portfolio authentication introduces is that only a (random) subset of the images in the password have to be entered during each authentication attempt (e.g. in [54] the password consist of 8 images and only four have to be recognised during each authentication attempt). Dunphy et al. [62] investigated the shoulder-surfing resistance of *recognition-based graphical portfolio authentication*. Their prototypes were intended for use on mobile phones (see figure 5.18) and need neither additional controls, thereby fulfilling $R2$ (compatible-with-gamepad-controls), nor high-precision directional input, thereby fulfilling $R6$ (no-mouse-equivalent-precision-input), nor additional hardware, thereby fulfilling $R4$ (no-additional-hardware)1. It also requires just one

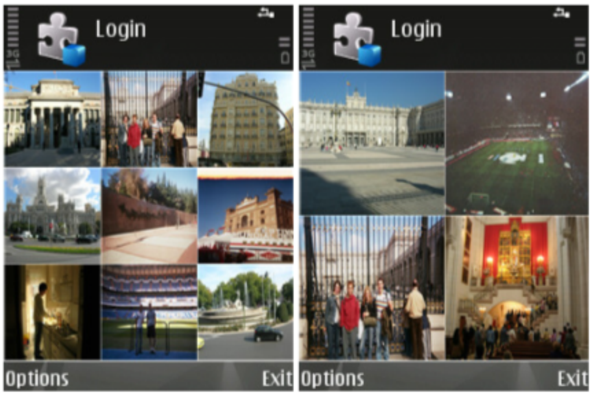
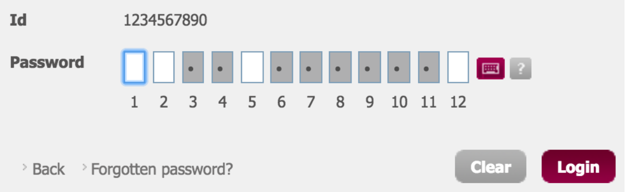


Figure 5.18: Interface of the *graphical portfolio authentication* scheme by Dunphy et al. as depicted in [62] on page 7 (“high entropy” version on the left, “low entropy” version on the right).



Id 1234567890

Password

Figure 5.19: Interface of the *partial passwords* scheme. Screenshot from `https://aliorbank.pl/hades/do/Login` (accessed on 2017-01-24).

directional control and is therefore anatomically-compatible ($R5$). However, as graphical scheme it is not compatible-with-text-passwords ($R3$). Yet, the portfolio approach has also been applied to text passwords (e.g. by banks [140]). Such *textual portfolio authentication* (also termed "masked passwords" [140] or "partial passwords" [11]) requires the user to enter only a subset of the characters of their password during each authentication attempt. Figure 5.19 depicts the interface of a partial password implementation as deployed on the Internet. In the gamepad context, such a scheme could be combined with an on-screen keyboard. The resulting scheme would fulfil all the requirements since the interaction is based on an on-screen keyboard, which leads to the fulfilment of requirements $R2 - R6$ and the portfolio approach leads to fulfilment of $R1$. Thus *textual portfolio authentication* seems to be suitable for usage with gamepads.

5.3 Discussion

This section discusses the results and the methodology of the assessment of the different schemes presented in this chapter. Thereafter, possible next steps to continue this line of work are outlined.

5.3.1 Results

In its first part, this chapter outlined the specific requirements of authentication in the gamepad context, which were grouped into three categories: security, technical, and usability. These requirements can guide and inform the design and development of new authentication schemes in the gamepad context. However, they only supplement the requirements put forward for authentication in general. Therefore, when designing new authentication schemes, it is of the essence to consider the full application scenario and the respective general requirements which might apply in the specific application scenario. Such general requirements are e.g. described by Bonneau et al. [26] in their framework for comparative evaluation of authentication schemes or by Mayer et al.'s [139] extension of Bonneau et al.'s framework. Using the terminology of [139], the requirements described in this chapter should be treated as hard-constraints, i.e. all authentication schemes in the gamepad context need to fulfil them.

In its second part, this chapter assessed authentication schemes currently deployed in the gamepad context. The de facto standard regarding deployed schemes are on-screen keyboards which are only somewhat adapted to the gamepad context, but implemented on all major platforms in a similar way. An alternative implementation geared more towards the usage with gamepads is the Daisy Wheel scheme. However, it is only available on the Steam platform with a special gamepad (i.e. most Steam users still use a traditional on-screen keyboard). None of the currently deployed schemes include specific counter-measures to shoulder-surfing. This is somewhat unsurprising, since the most common implementation is to use the default text entry method and no specific password entry scheme. Such default text entry schemes are mostly used to enter non-sensitive information (e.g. search terms in video streaming applications) and therefore give explicit feedback to users regarding the currently entered characters. This is unsuitable for the entry of secrets such as text passwords. Thus, it seems that none of the currently deployed schemes fulfil all six requirements outlined in this chapter, indicating a need for future designs to fill this gap.

One possibility to fill this gap is the adaptation of shoulder-surfing resistant authentication schemes proposed in the literature (potentially for different contexts). Therefore, in addition to the currently deployed schemes, a representative set of shoulder-surfing resistant authentication schemes proposed in the literature was assessed as well covering the five categories: usage of covert channels, obfuscation of the user's input through distractors, usage of indirect input, additional biometric layers, and portfolio authentication. It is found that only four meet all of the requirements identified in the first part and can be feasibly adapted to the gamepad context: the *grid-based scheme* [123], *PairPasswordChar* [166], the *cognitive trapdoor game* [173], and *textual portfolio authentication* [11, 140]. PairPasswordChar is however, a scheme requiring the memorisation and application of complex rules by the user to determine where to click in the grid of characters. It is therefore not easily understood, making it a subpar candidate. The cognitive trapdoor game was originally proposed as PIN replacement and when scaling it up to work with text passwords, users would have to complete 56 rounds of the trapdoor game. This highly repetitive task is likely to incur an unacceptable penalty to the scheme's usability and therefore renders it a subpar candidate as well. Textual portfolio authentication is feasible, but requires either increased lengths for the text passwords or will decrease the offered security level. The grid-based scheme seems to be the proposal which is most suitable for adaptation to the gamepad context, since the basic working principle and interface need to be changed only slightly in an adaptation. However, even for the grid-based scheme it becomes apparent that it does not leverage the true potential of the inputs available on gamepads. Therefore, novel proposals specifically tailored to the gamepad context are needed. One such proposal is the Colorwheels scheme which was developed as part of the work underlying this chapter, but is explicitly not a contribution in this thesis. It is therefore described as post scriptum to this chapter (section 5.5).

5.3.2 Limitations

The basis for the identified requirements is the current state of the art in terms of gamepad design, sensor availability, and typical platforms of gamepad-driven devices. If this basis changes, it will become necessary to update the requirements accordingly. Among the requirements which might easily change are $R2$ (i.e. the controls available on a typical gamepad) and respectively also $R6$ (i.e. anatomical constraints set by the available controls in conjunction with human anatomy), if new generations of gamepads are brought to the market and find wide-spread adoption. Likewise, $R3$ is directly linked to the human ability of handling high precision input with the controls available on a gamepad and might change depending on the available controls. To counteract the need to change these requirements over time, a more abstract wording could have been chosen for the requirements which might then just have referred to the current state of the art. This in turn, however, might have decreased the usefulness of the requirements for concrete designs, since their application would have necessitated a research in what the current state of the art is by the designer of the new authentication scheme. Also, changes to these requirements are expected to manifest only over longer periods of time, as the steady but slow evolution of gamepads has shown in the past [159]. In contrast, $R1$ will remain important as long as authentication takes place in shared spaces, i.e. as long as people authenticate on gamepad-driven devices while others are present. Due to the social nature of the activities gamepad-driven devices are used for (i.e. watching movies or playing games), one can argue that this is unlikely to change. Likewise, $R4$ depends on the perseverance of text passwords and is therefore also unlikely to change in the short term [98]. $R5$ depends on the platforms of gamepad-driven devices, which are usually relatively confined. However, more open platforms might still play a role in the future. A prediction with respect to $R5$ is therefore difficult.

Regarding the set of shoulder-surfing resistant authentication schemes from the literature, it must be acknowledged that this selection does not represent the result of a systematic literature research. Instead, the schemes were chosen to reflect representative authentication schemes of each of the five categories described in the literature: *covert channels*, *obfuscation of input*, *indirect input*, *behavioural biometrics*, and *portfolio authentication*. Therefore, while schemes that fulfil all six requirements could be identified, a systematic literature review might uncover additional schemes fulfilling all of the requirements.

5.3.3 Next Steps

From the assessment of the schemes, several open questions emerge, outlining the next steps in continuation of the work presented in this chapter:

- Since none of the currently deployed schemes include specific counter-measures to shoulder-surfing, it seems important to empirically assess a baseline for the on-screen keyboard as de-facto standard in the gamepad context. This would allow to more precisely gauge the shoulder-surfing threat in this context and determine how resistant to opportunistic shoulder-surfing the deployed schemes truly are.
- Empirically assessing alternatives to the on-screen keyboard in terms of both, usability and shoulder-surfing resistance, would allow to gauge potential trade-offs in shoulder-surfing resistance and usability as well as inform future development and evaluation of authentication schemes in the gamepad context.

In summary, a comparative empirical evaluation of the usability properties and shoulder-surfing resistance of the on-screen keyboard and alternative schemes is required. The execution of this empirical evaluation is presented in the next chapter.

5.4 Conclusion

This chapter outlined the requirements of authentication in the gamepad context. None of the currently deployed and only four of the assessed proposals from the literature fulfil all six requirements. Overall, more investigations into the suitability of authentication schemes in the gamepad context – in terms of both, security and usability – are needed. Only then can a true baseline for the performance of currently deployed schemes be established. Assessing the suitability of proposals – both, those adapted from non-gamepad contexts as well as those specifically tailored to the gamepad context – in comparison to this baseline is the most important next step to continue the work towards truly usable and secure user authentication in the gamepad context.

5.5 Chapter 5 Post Scriptum: Colorwheels

While some of the shoulder-surfing resistant schemes described in section section 5.2 can be adapted for use in the gamepad context, none of them were specifically developed for this context. As part of the work underlying this chapter (i.e. [138]) a novel password entry scheme called *Colorwheels* was developed based on the results described in this chapter. While *Colorwheels* explicitly is not a contribution of this thesis, an understanding of this scheme is required for the next chapter. Therefore, *Colorwheels* is described in the remainder of this section as post scriptum to this chapter.

Colorwheels is specifically designed for shoulder-surfing resistant input of text passwords using gamepads. The general design of the scheme is based on pie menu structures [42], similar to the Daisy Wheel scheme described in section 5.2. Its interface consists of two pie menu “flowers” with eight petals each. These overall 16 petals contain all possible characters (i.e. uppercase and lowercase letters, numbers, and special characters) for the password entry. This design is depicted in figure 5.20. *Colorwheels* is designed specifically for text entry with a gamepad, thus meeting requirements $R4$ and $R5$. Due to the two flowers, *Colorwheels*’s operation necessitates the availability of two analogue sticks on the gamepad. Each stick is used to select petals in one of the flowers: the left stick to select petals in the left flower and the right stick to select petals in the right flower. Each petal holds either 6 or 5 characters, allowing the placement of all 94 printable ASCII-characters on the petals of the two flowers.

The entry of each password character is performed with the following four-step procedure (see figure 5.21): (1) The characters appear randomly distributed on the petals of the two flowers and the user locates the petal with the desired character. (2) Once the user locates the petal with the desired character, they press the ×-button to confirm that they have found it. Upon pressing the button, all the characters vanish from the petals. (3) Then the user selects the petal. Each flower corresponds input-wise to one of the two analogue sticks. To select a petal in the left or right flower, the left or right analogue stick have to be used respectively. Since there are only eight different positions for each analogue stick, the scheme meets $R3$. The current input is reflected by a light change in colour of the selected petal (visible in figure 5.21 in step 3 where the upper petal on the right flower is selected). (4) The user confirms the selection of the petal using any of the shoulder buttons. Upon pressing the button to confirm the selection, a new random distribution of characters appears on the petals and the procedure begins anew for the next character.

Colorwheels belongs to the category of authentication schemes obfuscating the users’ inputs (see section 2.3.1). To obtain the password, a shoulder-surfing attacker would have to memorise the random distribution of all characters in the time the user locates the petal with the character they want to enter. The basic assumption underlying the shoulder-surfing resistance is that this task is cognitively demanding enough

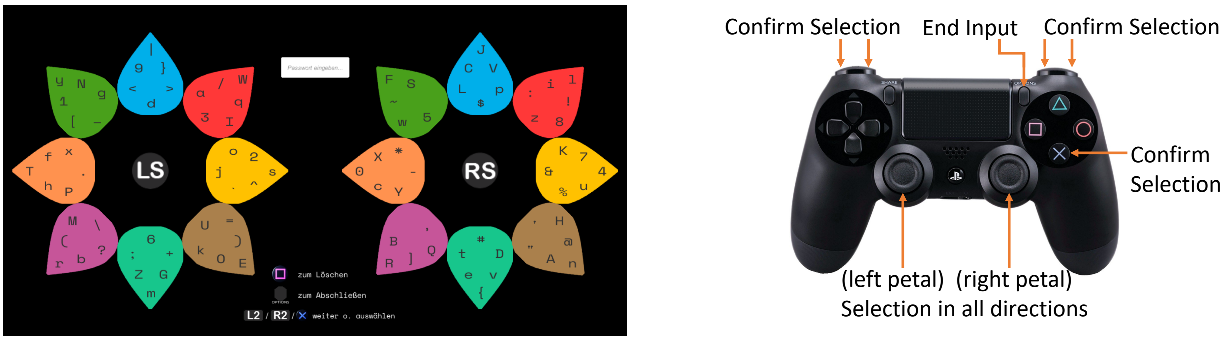
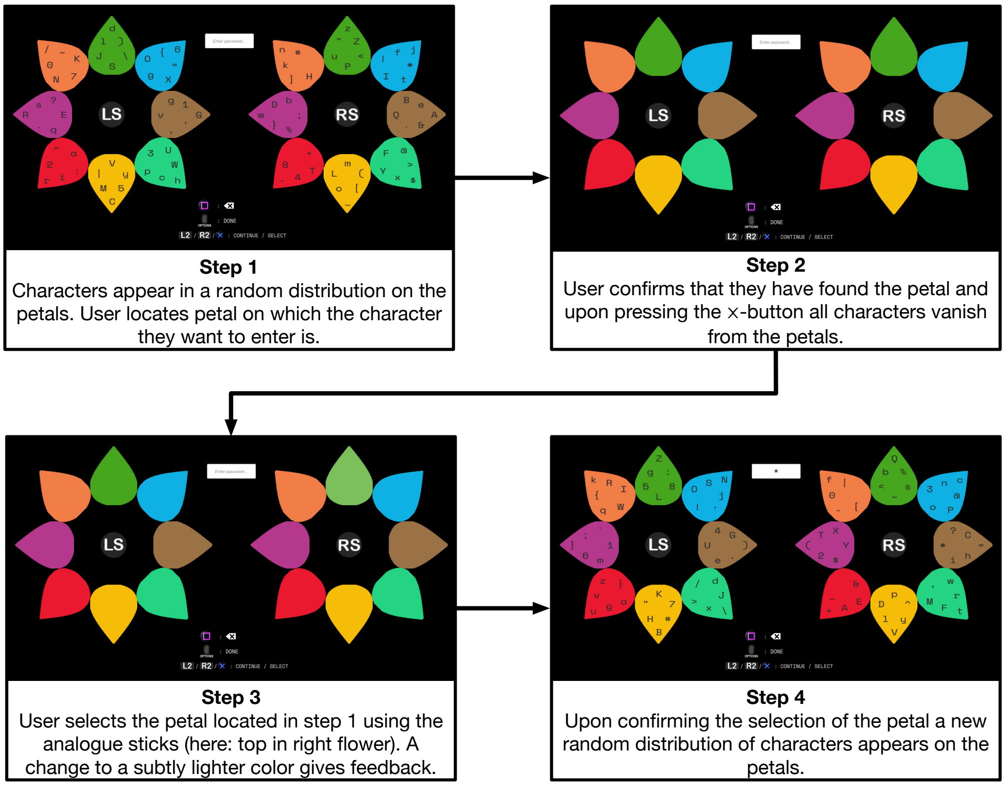


Figure 5.20: The interface of Colorwheels and the button layout to operate it.



Step 1
Characters appear in a random distribution on the petals. User locates petal on which the character they want to enter is.

Step 2
User confirms that they have found the petal and upon pressing the ×-button all characters vanish from the petals.

Step 3
User selects the petal located in step 1 using the analogue sticks (here: top in right flower). A change to a subtly lighter color gives feedback.

Step 4
Upon confirming the selection of the petal a new random distribution of characters appears on the petals.

Figure 5.21: The password entry procedure of Colorwheels.

to protect against opportunistic shoulder surfing, i.e. meeting requirement $R1$. However, this opportunistic shoulder-surfing property must still be validated in an empirical evaluation.

At any time during the procedure, the scheme requires less than two concurrent controls on the front and less than the number of overall available buttons, meeting requirements $R2$ and $R6$. Using the □-button, the last entered character can be deleted, allowing for easy recovery from errors.

6 Empirical Evaluation of Three Authentication Schemes in the Gamepad Context

From the last chapter, i.e. the assessment of the schemes currently deployed in the gamepad context (section 5.2.1) as well as the proposals in the literature for non-gamepad contexts (section 5.2.2), it became apparent that an empirical evaluation to determine a baseline in terms of usability as well as shoulder-surfing resistance is needed to gauge the real-world resistance to opportunistic shoulder-surfing of schemes already deployed in the gamepad context. Of most interest is the incumbent scheme, i.e. the on-screen keyboard. An empirical evaluation also allows to compare this baseline of the incumbent scheme to alternatives from the literature as well as schemes tailored-specifically to the gamepad context. Such a comparative empirical evaluation will indicate where trade-offs among the evaluated schemes exist, which possible alternatives to the incumbent scheme are better suited to be used in the gamepad context, and which design directions might be worth to follow.

Therefore, this chapter presents a comparative evaluation of three schemes discussed in the previous chapter: (a) the on-screen keyboard (section 5.2.1) as incumbent and de facto standard in the gamepad context in order to determine a baseline, (b) the grid-based scheme (section 5.2.2) as proposal from the literature, which was found in the discussion of the results in the last chapter (section 5.3.1) to be the scheme from the literature most suitable for an empirical evaluation, and (c) the Colorwheels scheme (section 5.5) as proposal designed specifically for password entry with gamepads. To that end, two user studies – one online study and one lab study – were conducted to comparatively evaluate the three schemes' resistance to opportunistic shoulder-surfing and their usability as determined by the metrics effectiveness, efficiency, and satisfaction of ISO 9241-11:2018-03 [108]. Conducting both, an online study and a lab study, allows harnessing the advantages of both study settings. Both studies are based on the same methodology (section 6.1). The procedures of the online study (section 6.2) and the lab study (section 6.3) differ slightly: while both studies empirically evaluate the opportunistic shoulder-surfing resistance, only the lab study evaluates the usability metrics. To assess the shoulder-surfing resistance, participants were asked to recover a password by observing video recordings of its entry. To assess the usability, participants used the three schemes themselves and their performance with respect to the metrics efficiency, effectiveness and satisfaction were measured.

From the discussion (section 6.4) it becomes apparent that the results of the two studies confirm how little protection even against opportunistic shoulder-surfing the commonly used *on-screen keyboard* provides: It is significantly more susceptible to shoulder-surfing than the other two schemes in both studies. Both other schemes fare better, but the *Colorwheels* scheme seems to exhibit a more robust shoulder-surfing resistance. Usability-wise, the *on-screen keyboard* fares best. It performs significantly better in terms of efficiency and satisfaction than the other two schemes as well as significantly better in terms of effectiveness than *Colorwheels*. *Colorwheels* scores significantly better in terms of efficiency and satisfaction than the *grid-based* scheme and is rated highest by the participants in terms of intention to use the scheme in the future. This indicates that there currently exists a clear trade-off between usability and shoulder-surfing resistance, but that users are willing make this trade-off in the context of text password entry on gamepads in order to better protect themselves from shoulder-surfing. Section 6.5 concludes this chapter.

Contributions described in this chapter:

- A baseline in terms of usability and shoulder-surfing resistance is established for the on-screen keyboard (as incumbent and de-facto standard in the gamepad context) by conducting two user studies – an online study and a lab study – using similar methodologies.
- The viability of alternatives to the on-screen keyboard is assessed, by evaluating two additional schemes in the same study settings:
  1. The grid-based scheme, which is a shoulder-surfing resistant authentication scheme proposed in the literature and identified as most viable candidate adaptable to the gamepad context.
  2. The novel Colowheels scheme (section 5.5) which was specifically designed for the gamepad context.

Parts of the results described in this chapter have been published in:

- P. Mayer, N. Gerber, B. Reinheimer, P. Rack, K. Braun, and M. Volkamer, "I (don't) see what you typed there! Shoulder-surfing resistant password entry on gamepads", Conference on Human Factors in Computing Systems (CHI), 2019.

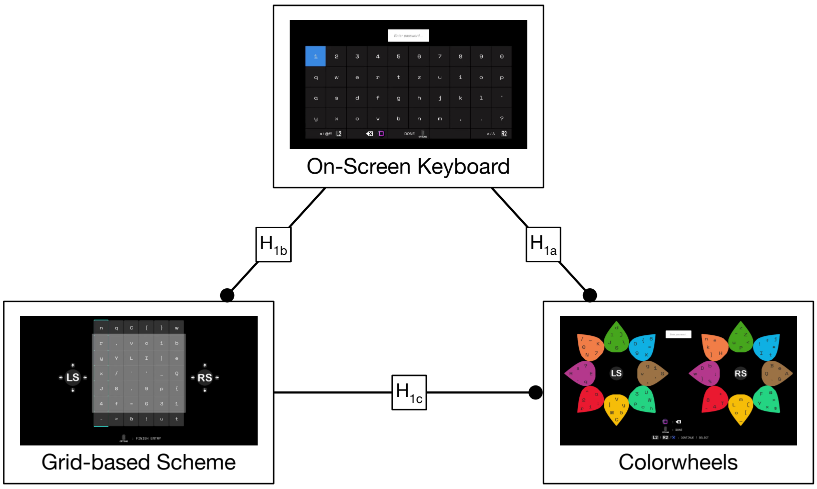


Figure 6.1: Overview of the hypotheses regarding the shoulder-surfing resistance of the three authentication schemes evaluated in the gamepad context. The ● indicates for each hypothesis which authentication scheme is expected to perform better (i.e. is more shoulder-surfing resistant).

6.1 Methodology of the User Studies

This section describes the hypotheses investigated in the two user studies, the design decisions regarding the study methodologies, and the implementations of the three schemes.

6.1.1 Hypotheses

Since the *on-screen keyboard* (as the baseline condition) does not employ any measures to counter shoulder-surfing attacks, it is expected that it is the least resistant scheme in this aspect. The respective hypotheses are:

H_{1a}: *The Colorwheels scheme is more resistant to opportunistic shoulder-surfing attacks than the on-screen keyboard.*

H_{1b}: *The grid-based scheme is more resistant to opportunistic shoulder-surfing attacks than the on-screen keyboard.*

While the *grid-based* scheme is geared towards being shoulder-surfing resistant, the full entry is gradually revealed during password entry and old input is visible until the complete password is entered. Therefore, a simple shoulder-surfing strategy is to memorise the characters in one column each time and then check during the next observation which one of the five characters appears again. In contrast, *Colorwheels* shows the randomly distributed characters only before the individual character selection is performed (cf. section 5.5). Therefore, the attacker would have to memorise the complete (randomised) character layout (i.e. the positions of all 94 characters) for each character in the password, when employing a similar strategy. Therefore, *Colorwheels* is expected to be more resistant:

H_{1c}: *The Colorwheels scheme is more resistant to opportunistic shoulder-surfing attacks than the grid-based scheme.*

Figure 6.1 gives an overview of these three hypotheses.

6.1.2 Design Decisions for the Methodology

Most studies regarding shoulder-surfing resistance are conducted as lab studies (e.g. [19, 36, 57, 59, 75, 121, 122, 175, 218, 229]). In contrast, only very few user studies have been conducted online [12, 219]. Yet, both types of studies have their advantages and disadvantages.

A lab study allows participants to use the schemes under test themselves to get familiar with the potentially unknown scheme. Thereby, a lab study also allows assessing the usability of the schemes. In contrast, an online study has to rely on explanatory videos and texts for the familiarisation with the schemes and assessing the usability is not possible. Also, users might be more motivated in a lab study since they are observed. Online studies, on the other hand, are unobserved and therefore controlling confounding variables might not be possible. Attention check questions can help mitigate this problem by asking whether the participant cheated [111], but such attention checks always rely on self-reported data and must therefore be complemented by other metrics. Yet, participants can engage the online study whenever they have time, without the need of supervision by an experimenter, facilitating the collection of large samples. Furthermore, while lab studies necessarily rely on local recruiting, online studies can reach a much wider population.

Due to these different advantages and disadvantages of online studies and lab studies, it was decided to conduct both: an online study and a lab study. Additionally, this allows the collection of more data to assess the shoulder-surfing resistance and allows the comparison of the two studies' methodologies with respect to the advantages and disadvantages outlined before.

The design of the two studies is based on the methodology of Aviv et al. [12], who conducted a blend of an online study and a lab study to gain a baseline for the shoulder-surfing resistance of smartphone PINs and the Android pattern lock. In the following, a brief overview of their methodology and the adaptations needed for its application in the gamepad context are given. The detailed procedures are described in section 6.2 for the online study and section 6.3 for the lab study.

The assessment of the shoulder-surfing resistance by Aviv et al. [12] is based on 10 attack trials using "video recordings of a single expert user being attacked by participants". This design decision is carried over into the gamepad context, since it minimises the introduction of unwanted variance in the collected data (as opposed to using recordings of different users with potentially different familiarity of the schemes). Following the methodology in [12], the videos for the grid-based and on-screen keyboard schemes were varied regarding the entry speed of the password and the interaction, i.e. different paths taken on the *on-screen keyboard* from one character to the next and different scrolling directions for the columns in the *grid-based* scheme. For *Colorwheels* only the speed was varied, since the interaction cannot be changed. Due to the mobile scenario, Aviv et al. [12] also varied the size of the device and the viewing angle in the videos and treated both as within-factors. However, it can be argued that this is not relevant in the gamepad scenario, since it can be assumed that the observer has a clear and unobstructed view of the full GUI on the screen where the text entry takes place (e.g. from the couch in front of the television, when the goal is to watch a movie or play a game together). Also, observing the gamepad in addition to the screen does not give an advantage, since all input is directly reflected in the authentication schemes' GUIs. In the *grid-based* scheme, the GUI shows which column is selected and the letters are shifted in the direction pressed on the analogue stick. Only which letters are visible on the screen at the end is important to an observing attacker, the input at the gamepad is not. For *Colorwheels*, the GUI highlights the petal which is selected using the analogue sticks, also directly reflecting the input. Consequently, the videos showed only the authentication schemes as displayed on the screen.

Each of the ten videos shows the entry of the same randomly generated password. This password was chosen to not introduce bias into the study. It requires usage of the different character sets in the *on-screen keyboard*, since otherwise the shoulder-surfing task might be rendered unnecessarily easy for this scheme and thereby favour the other two schemes in the study. Also, using a dictionary word might have put the *grid-based* scheme in a disadvantage and favoured the other two schemes. Therefore, the password in the videos was chosen at random to include uppercase letters, lowercase letters, numbers and symbols. The random password entered in all the videos for the *on-screen keyboard* and the *grid-based* scheme is *W8@b=L*. The length of six characters was chosen according to the NIST recommendation for random memorised secrets [85]. To align the guessing probability of *Colorwheels* with that of the *grid-based* scheme, the length of the password had to be increased by one character to *W8@b=Lx*.

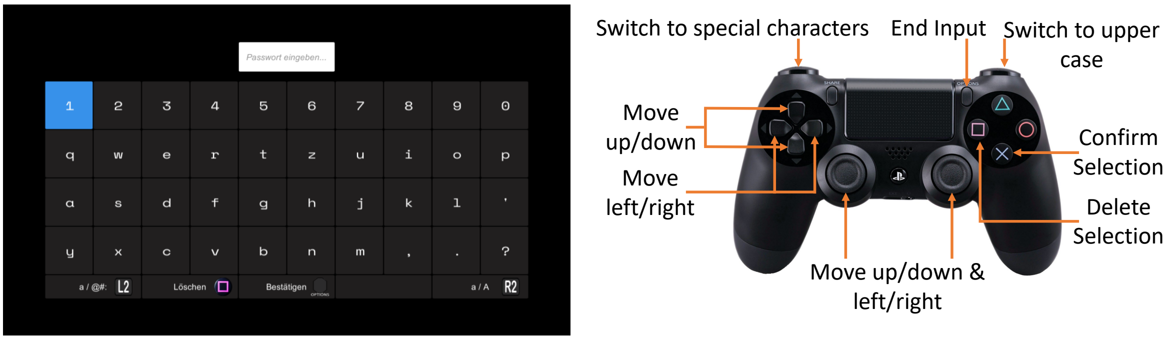
Participants saw each video only once and had to guess once after each video in both studies presented in this chapter. This design decision was made to be more consistent with the opportunistic nature of observations in the gamepad scenario, i.e. one video represents one opportunistic observation at another users home.

Analogously to Aviv et al. [12], the different schemes were treated as between-subjects factor and participants were not allowed to take notes. For our lab study, participants were recruited locally using flyers on campus, online gaming forums, and word of mouth. Potential participants of the lab study had to fill out a short sign-up questionnaire in which they stated their experience with game consoles.

6.1.3 Apparatus

All three schemes were re-implemented for this empirical evaluation using the Unity game engine[1] for usage with a Dualshock 4 gamepad. All three implementations share a common basis consisting of a menu structure to start the scheme and a backend which recorded the performance metrics for our usability evaluation, i.e. the time needed to authenticate (efficiency) and whether the authentication was successful (effectiveness). During development, all three schemes were tested in informal pre-test sessions with people recruited on campus. The user interaction of *Colorwheels* is already described in detail in section 5.5. In the following the user interaction of the other two schemes is described.

On-Screen Keyboard. The on-screen keyboard for our study was designed to resemble the layout and functionality commonly found on gamepad-driven devices. All keys are aligned in a grid showing numbers in the top row and lowercase letters and some commonly used symbols in the remaining rows. Figure 6.2 shows the displayed (default) keyboard and the gamepad button layout used to operate the keyboard. Additional

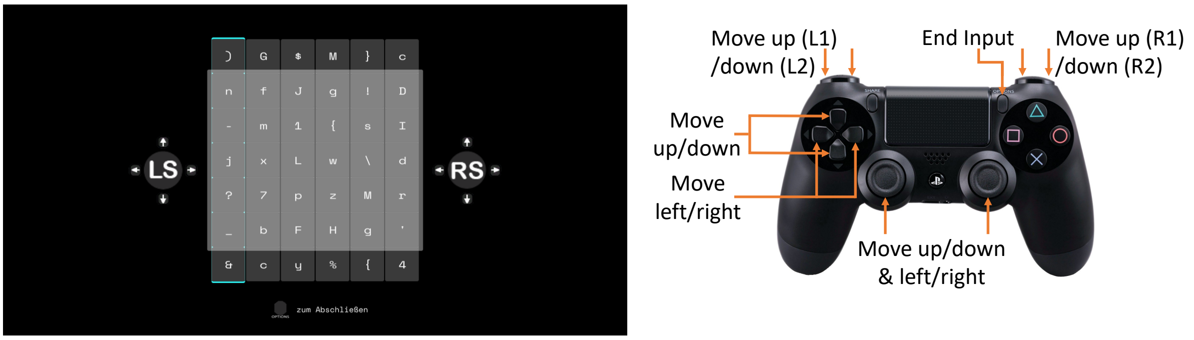


Passwort eingeben...

1	2	3	4	5	6	7	8	9	0
q	w	e	r	t	z	u	i	o	p
a	s	d	f	g	h	j	k	l	'
y	x	c	v	b	n	m	,	.	?

Figure 6.2: Our *on-screen keyboard* implementation. The currently selected character is highlighted in blue.

[1] https://unity3d.com/unity



)	G	$	M	}	c
n	f	J	g	!	D
-	m	1	{	s	I
j	x	L	w	\	d
?	7	p	z	M	r
_	b	F	H	g	'
&	c	y	%	{	4

Figure 6.3: Our implementation of the *grid-based* scheme with the original German interface as used in the study. The currently selected column is highlighted in turquoise. Each column corresponds to one character in the password (i.e. first column to first character, etc.). The characters in the light-grey cells are considered as input for the respective character of the password.

character sets (uppercase letters, symbols) can be accessed by pressing the shoulder-buttons of the gamepad. The desired character can be chosen by moving the selection (highlighted in blue) using either the directional control pad or the analogue sticks. Using the ×-button, the currently selected character is entered for the password. Pressing the □-button deletes the last entered character, so users can correct errors.

Grid-Based Scheme. The interface of this scheme was designed to resemble the original depictions in [123] as much as possible, while retaining a homogeneous look with the the on-screen keyboard and Colorwheels. The interface comprises a grid of 6 by 5 grey cells, as depicted in figure 6.3. Only the analogue sticks (or alternatively the directional control pad) are needed to operate the scheme. Pushing any stick to the left or right lets the user select the column. Pushing up or down scrolls the characters in that column. The password character has to be scrolled into any of the grey cells. The scroll buttons in the original interface (which were used for the mouse input) were not needed anymore and therefore removed.

From the results of the pre-tests, two changes were introduced into the user interaction for the grid-based scheme used in the study. Firstly, the grid was expanded to 6 by 7 cells with the two rows of black cells (one above and one below the grey cells as depicted in figure 6.3), since users felt that this change would help them to get an idea which characters are scrolled into view next. Secondly, the functionality to scroll the characters in the selected column up with the L1/R1-buttons and to scroll the characters in the column down with the L2/R2-buttons was added.

6.2 Online study

This section describes the procedure, the participant sample and the results of the online user study. The study methodology conforms to all requirements of the university's ethics commission.

6.2.1 Procedure

In the following the procedure of the online user study is described. It consists of five phases.

Introduction and Informed Consent. The participants first received a short briefing outlining the study scenario (shoulder-surfing in the gamepad context), the remainder of the study (including the shoulder-surfing task), and explanations in case they want to withdraw from the study. Further, participants were asked to not complete the study on a mobile device and to provide their consent for participation and processing of their data. Then, they were asked whether their eye sight was normal (with or without corrections). Participants who reported to have a bad eye sight (that was not corrected to normal) were told that they could unfortunately not participate in our study.

Familiarisation with the Scheme. Participants were randomly assigned to one of the three schemes and shown an explanatory text and illustrations describing the assigned scheme and how to operate it. They were also asked to watch a video similar to those used later in the study for the shoulder-surfing trials, but with a different password (they were told which password was entered in the video and could watch it as often as they liked).

Shoulder-surfing the Scheme. Participants had to play the role of an attacker performing an opportunistic shoulder-surfing attack. To that end, they watched the videos[2] as outlined in section 6.1. Participants were told that they are not allowed to pause or rewind the video, take pictures or videos of the online study and its contents, or use pen and paper to take notes. In order to prevent participants from pausing, rewinding, or replaying the video, the control elements of the video player were hidden. Additionally, the button for proceeding to the next page of the questionnaire was hidden until the video playback was finished. Videos could be started by the participant by clicking anywhere on it and were automatically played in full-screen mode. Upon completion of the playback, the HTML-node containing the video was deleted from the DOM of the survey page to prevent repeated playback. Participants watched the videos for the scheme they were assigned to one after the other until they either had successfully recovered the password from the input shown in the videos, or had watched all ten videos without successfully recovering the password. In between the videos, they could either enter a guess for the password in a free text field or indicate that they have no idea about the password at all. In case the participants successfully recovered the password they were complimented on their performance. In case they failed to recover the password they were told that they should not bother as the goal of the study was to evaluate the respective scheme's resistance against shoulder-surfing. Participants who were assigned *Colorwheels* or the *grid-based* scheme were further told that this scheme was specifically developed to be resistant against shoulder-surfing attacks.

Attention Checks. After the completion of the shoulder-surfing task, participants were asked (a) whether they had used any aids – such as pen and paper – to help them guess the password, (b) whether they had found and applied a possibility to pause or rewind the video, and (c) whether they had completed the study on a mobile device. These questions served both as attention check and to check whether participants had followed the instructions. If the answer to any of those questions indicated that a participant had not followed the instructions, they were excluded from the analysis.

Demographics and Debriefing. Finally, participants were asked to provide demographic information. On the last page, participants were thanked for their participation and were given the code they needed to receive their compensation as well as contact details in case any questions would arise.

[2] The videos used in the study are available along the other study materials for download as supplemental material to the paper underlying this chapter at `https://dl.acm.org/citation.cfm?id=3300779`.

Table 6.1: Overview of the demographics of the online study participants.

Authentication scheme	**Gender**		**Age**					
	Female	***Male***	*< 20*	*20–30*	*31–40*	*41–50*	*51–50*	*> 60*
On-screen keyboard	8	11	0	6	5	5	2	1
Grid-based scheme	6	13	0	4	8	4	3	0
Colorwheels scheme	14	12	1	9	7	6	3	0
Total	*28*	*36*	*1*	*19*	*20*	*15*	*8*	*1*

Table 6.2: Overview of the shoulder-surfing resistance results in the online study.

Authentication scheme	**Number of observations needed to obtain password**										**Failed to obtain password**
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
On-screen keyboard		4	1	1	2		1	3			7
Grid-based scheme							1		1		17
Colorwheels scheme											26

6.2.2 Participants

Participants were recruited using the German panel "Clickworker". Participants required on average 13 minutes and received a compensation of 3€. 93 participants completed our study. 10 had to be excluded from the analysis because they failed attention checks. In addition, 10 participants who stayed less than 14 seconds on the page introducing the scheme were excluded, since this is insufficient to familiarise themselves with the scheme[3]. Four participants had to be excluded because the completion times for the video pages were shorter than the length of the video and five participants had to be excluded due to technical problems. The final sample therefore includes 64 participants (see table 6.1 for an overview of the participants' demographics).

6.2.3 Results

To investigate H_{1a-c}, three Wilcoxon rank-sum tests were run to account for the ordinal scale level of the data (i.e. number of observations needed to obtain the password, whereas participants who failed to obtain the password after having watched all ten videos were coded with "11"), using a Bonferroni-Holm-adjusted alpha-level to correct for multiple testing. Table 6.2 lists the values for all three authentication schemes. The analysis shows that, in accordance to our assumptions, the *on-screen keyboard* is least resistant to shoulder-surfing attacks. Participants need less observations to obtain the password in comparison to *Colorwheels* ($Z = -4.61$, $p < .001$, $r = .687$). Therefore, H_{1a} is supported. Likewise, more observations are needed to obtain the password entered with the *grid-based* scheme than with the *on-screen keyboard* ($Z = -3.50$, $p = .002$, $r = .567$), thereby supporting H_{1b}. The analysis did not reveal significant differences between the number of observations needed to obtain the password entered with the *Colorwheels* scheme and with the *grid-based* scheme ($Z = -1.67$, $p = .094$). Consequentially, H_{1c} is not supported.

[3] 14 seconds was the shortest time spent on the introductory page by a participant who successfully guessed the password. In order to not favour any one scheme this same threshold for exclusion was applied to all schemes in the study.

6.3 Lab study

This section describes the procedure, the participant sample and the results of the lab user study. The study methodology conforms to all requirements of the university's ethics commission.

6.3.1 Procedure

In the following, the procedure of the lab user study is described. It consists of five phases.

Introduction and Informed Consent. The participants first received a short briefing outlining the study scenario, the remainder of the study, and explanations in case they want to withdraw from the study. Then they were asked to provide their consent for participation and processing of their data.

Familiarisation with the Scheme. Participants were randomly assigned to one of the three schemes and received an explanatory text and illustrations of the assigned scheme and its operation. To familiarise themselves with the assigned scheme they used it three times to enter a random password supplied to them on-screen. A new random password had to be entered during this familiarisation phase for each of the three runs. The password was displayed on-screen to decrease impact of the transcription.

Usability Assessment. All participants had to enter three different randomly generated passwords which were individually randomly generated, analogously to the familiarisation phase. However, during the three assessment runs in this phase, the ISO 9241-11:2018-03 [108] metrics effectiveness, efficiency, and satisfaction were recorded. Effectiveness was measured using the portion of successful password entries among the three (participants had only one attempt for each password). Efficiency was measured using the mean of the average time needed to enter the password across the three password entries. As in the familiarisation phase, the password was displayed on-screen. Satisfaction was measured using two metrics: (a) the SUS questionnaire and (b) the participants' usage intention of their assigned scheme, measured on a 5-point Likert scale as agreement to the statement "I would like to use this authentication scheme instead of the one I currently use in the future." (Likert scale labelled as "Disgree", "Disagree somewhat", "Undecided", "Agree somewhat", "Agree", and additional "Does not apply" option). The usability assessment (as well as the subsequent shoulder-surfing task) were performed on a 13" laptop screen instead of a television-sized screen, in order to remain comparable to the online study setting.

Shoulder-surfing the Scheme. The participant had to play the role of an opportunistic shoulder-surfing attacker. To that end, all participants watched the videos as outlined in section 6.1. They watched the videos for the scheme they had used during the usability assessment, one after the other until they either had successfully recovered the password from the input shown in the video, they had watched all ten videos without successfully recovering the password, or they asked to stop the experiment since they felt they would never be able to recover the password. In between the videos they noted their guess for the password on a paper provided to them. In case the participants successfully recovered the password they were complimented on their performance. In case they failed to recover the password they were told that they should not bother as the aim of the study was to evaluate the respective scheme's resistance against shoulder-surfing. Participants who were assigned *Colorwheels* or the *grid-based* scheme were further told that this scheme was specifically developed to be resistant against shoulder-surfing attacks.

Table 6.3: Overview of the demographics of the lab study participants.

Authentication scheme	**Gender**		**Age**					
	Female	***Male***	***< 20***	***20–30***	***31–40***	***41–50***	***51–50***	***> 60***
On-screen keyboard	14	15	3	16	3	5	1	1
Grid-based scheme	12	17	6	17	6	0	0	0
Colorwheels scheme	11	18	7	13	8	0	1	0
Total	*37*	*50*	*16*	*46*	*17*	*5*	*2*	*1*

Table 6.4: The participants' were distributed between the study groups to achieve homogeneous distribution of experience with game consoles.

Authentication scheme	**Experience with game consoles**		
	low	***medium***	***high***
On-screen keyboard	10	10	9
Grid-based scheme	10	9	10
Colorwheels scheme	10	10	9
Total	*30*	*29*	*28*

Demographics and Debriefing. In this last phase, participants were asked to fill out a questionnaire providing information about their demographics. Then, they received a short debriefing, were thanked for their participation and received their compensation.

6.3.2 Participants

A total of 87 individuals (37 female, 50 male) participated in the lab user study (see table 6.3 and 6.4 for participants' demographics). To achieve a diverse mix of participants having varying degrees of prior experience with gamepads and game consoles, all potential participants had to fill out a short online signup-questionnaire asking them about their experience with game consoles and gamepads (low, medium, or high) and their email address (to contact them in case they were selected for the study). Links to this signup-questionnaire were distributed on campus using flyers and mailinglists. Additionally, postings were made in several Facebook groups and online forums relating to console gaming to recruit participants outside the university. Participants received a compensation of 5€.

6.3.3 Shoulder-Surfing Resistance Results

Since the participants in the lab study had the opportunity to stop before having watched all ten videos (the reasons and implications of this aspect are further discussed in section 6.4.1), it is not possible to analyse how many observations participants needed to obtain the password. They might have been able to successfully guess the password if they had continued. Therefore it is analysed instead how many participants succeeded in obtaining the password, independently of how many videos they watched. Therefore, the shoulder-surfing resistance in the lab study is rather a measure of the perceived difficulty of shoulder-surfing attacks. This aspect is further discussed in section 6.4.

To account for the nominal scale level of the data, Fisher‘s exact test was used to investigate H_{1a-c}, using a Bonferroni-Holm-adjusted alpha-level to account for multiple testing. Similar to the results from the online study, the analysis showed that the *on-screen keyboard* provides little protection against shoulder-

Table 6.5: Overview of success rates for the shoulder-surfing attacks in the lab study and of of the online study (for easier comparison of portions of participants who guessed the password correctly).

Authentication scheme	Lab Study	Online Study
On-screen keyboard	93.1% (27/29)	63.2% (12/19)
Grid-based scheme	37.9% (11/29)	10.5% (2/19)
Colorwheels scheme	0.0% (0/29)	0.0% (0/26)

surfing, with 27 out of 29 participants successfully obtaining the password entered with this scheme, which is significantly more than the successfully obtained passwords for both, the *Colorwheels* scheme (0 out of 29, FET: $p < .001$) and the *grid-based* scheme (11 out of 29, FET: $p < .001$). H_{1a} and H_{1b} are thus supported by our results. Finally, a third Fisher's exact test revealed that significantly more participants succeeded in obtaining the password entered with the grid-based scheme than with the *Colorwheels* scheme (FET: $p < .001$), providing support for H_{1c}. Table 6.5 summarises the shoulder-surfing resistance results of both studies.

6.3.4 Usability Results

The usability of the three schemes was assessed in terms of effectiveness (i.e. portion of successful authentication attempts), efficiency (i.e. time needed to enter each of the three passwords), and the participants' satisfaction (i.e. SUS scores).

Effectiveness. Figure 6.4 shows an overview of the successful authentication attempts. The on-screen keyboard scores best in terms of effectiveness: out of the three password entry attempts each participant had, 82.1% of participants had three successful attempts, 17.9% had two successful attempts, and no participant had 1 or less successful attempts. The grid-based scheme fares slightly worse: 67.9% had three successful attempts, 28.6% had 2 successful attempts, and 3.5% had one successful attempt. Colorwheels fares worst: 48.3% had three successful attempts, 34.5% had two successful attempts, and 17.2% had one successful attempt. Non-parametric tests were used for the analysis since a Kolmogorov-Smirnov test indicated non-normally distributed scores for all three schemes ($p < .001$). A Kruskall-Wallis test revealed significant differences in effectiveness between the three schemes ($\chi^2(2) = 7.54$, $p = .023$, $\eta^2 = .066$). Pairwise comparisons using Wilcoxon rank-sum tests with a Bonferroni-Holm-adjusted alpha-level revealed a significantly higher rate of successful password entries for the *on-screen keyboard* compared to the *Colorwheels* scheme ($Z = -2.67$, $p = .008$, $r = .35$). There were no significant differences between the *Colorwheels* scheme and the *grid-based* scheme ($Z = -1.59$, $p = .113$) or the *grid-based* scheme and the *on-screen keyboard* ($Z = -1.22$, $p = .222$). A closer look at the errors participants made during the three authentication attempts reveals that the similarity of characters is a major problem of the *grid-based* scheme: 9 out of 11 failed attempts

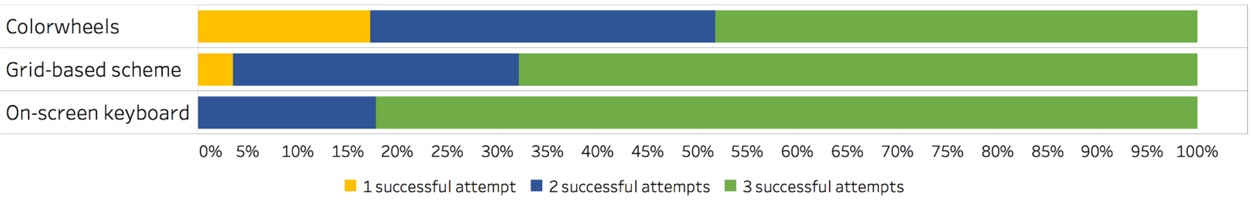


Figure 6.4: Effectiveness measured as the number of successful authentication attempts performed during the usability assessment phase of the lab study.

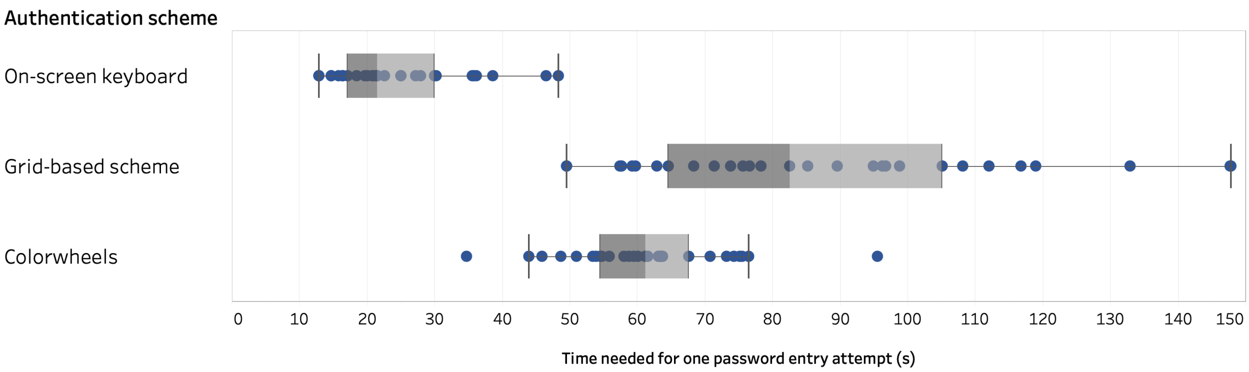


Figure 6.5: Boxplots of the efficiency data, measured as the mean of the overall time needed to enter the password across the three password entries.

to authenticate using the *grid-based* scheme can be attributed to participants confusing the target character with a similar character, whereas only 3 out of 6 failed authentication attempts arise from this problem for the *on-screen keyboard* and 9 out of 20 for the *Colorwheels* scheme. Other errors include participants mixing up uppercase and lowercase letters (1 for the *Colorwheels* scheme and *on-screen keyboard* each, 2 for the *grid-based* scheme), and forgetting to enter a character or entering an extra character (2 for the *on-screen keyboard*).

Efficiency. Figure 6.5 shows an overview of the times needed to enter the password. Outliers deviating more than 1.5-times the interquartile range from the mean were excluded from the analysis, resulting in an exclusion of four data points (two falling below the threshold for the *on-screen keyboard*, and one each falling below and exceeding the threshold for the *Colorwheels* scheme). It becomes clear that analogously to the effectiveness results, the on-screen keyboard fares best with 24.6 seconds as mean time needed for one password entry attempt. However, in contrast to the effectiveness results, the ranks of the other two schemes are reversed: Colorwheels takes less time for one password entry attempt than the grid-based scheme (61.3 seconds and 87.5 seconds respectively for mean times needed for one password entry attempt). All assumptions for conducting an ANOVA were met. Therefore, an ANOVA was run with the used scheme as the independent variable and the mean of the overall time needed to enter the password across the three password entries as the dependent variable. The analysis revealed significant differences in the mean time needed to enter the password ($F(2, 80) = 93.78$, $p < .001$, partial $\eta^2 = 0.701$). Pairwise comparisons with a Bonferroni-Holm-adjusted alpha-level showed that participants need significantly less time to enter the password using the *on-screen keyboard* compared to the *Colorwheels* scheme ($t(52) = -16.81$, $p < .001$, $r = .92$) and the *grid-based* scheme ($t(32.39) = -12.10$, $p < .001$, $r = .91$). However, authenticating themselves also took participants significantly less time using the *Colorwheels* scheme compared to the *grid-based* scheme ($t(34.36) = -4.88$, $p < .001$, $r = .62$).

Satisfaction. Figure 6.6 shows an overview of the SUS scores. Again, outliers deviating more than 1.5-times the interquartile range from the mean were excluded from the analysis, resulting in the exclusion of four data points (three falling below the threshold for the *on-screen keyboard* and one falling below the threshold for the *Colorwheels* scheme). The satisfaction results show the same ordering of the schemes as the efficiency results: the on-screen keyboard fares best with a mean SUS score of 88.0, then follows Colorwheels with a mean SUS score of 73.5, and last is the grid-based scheme with a mean SUS score of 64.0. All assumptions for conducting an ANOVA were met. Therefore, an ANOVA was run with the authentication scheme used to

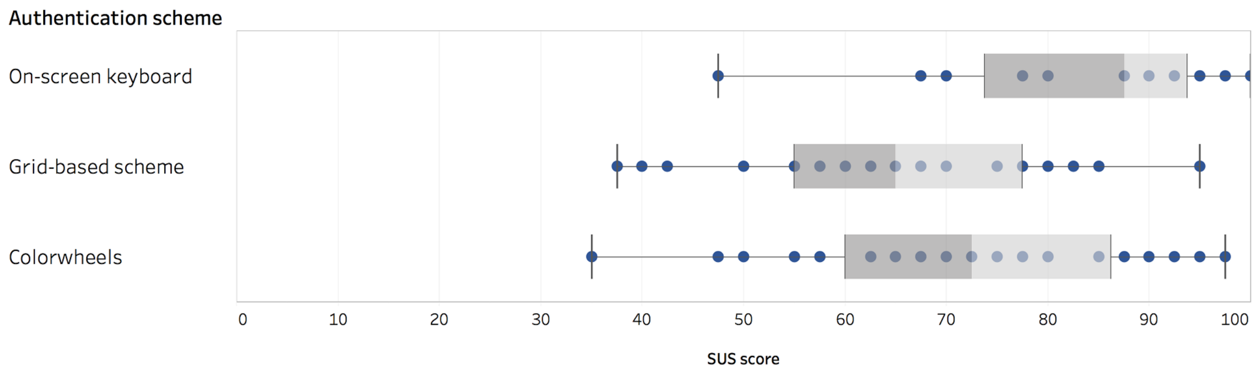


Figure 6.6: Boxplots of the satisfaction data measured with the SUS questionnaire.

enter the password as the independent variable and the SUS scores as the dependent variable. The analysis revealed significant differences in the SUS scores ($F(2, 80) = 33.40$, $p < .001$, partial $\eta^2 = 0.455$). Pairwise comparisons with a Bonferroni-Holm-adjusted alpha-level showed that the SUS scores were significantly higher for the *on-screen keyboard* compared to the *Colorwheels* scheme ($t(40.01) = 5.71$, $p < .001$, $r = .67$) and the *grid-based* scheme ($t(39.30) = 8.85$, $p < .001$, $r = .82$). However, the SUS scores also indicate that participants were significantly more satisfied with the *Colorwheels* scheme than with the *grid-based* scheme ($t(55) = 2.92$, $p < .005$, $r = .37$).

Figure 6.7 shows an overview of the participants' intention to use their assigned schemes in the future. Outliers deviating more than 1.5-times the interquartile range from the mean were excluded from the analysis, resulting in the exclusion of two data points (both falling below the threshold for the *Colorwheels* scheme). The "Does not apply" responses were excluded from the analysis. It becomes clear that *Colorwheels* fares best with respect to this metric (mean Likert score of 4.5, where higher values represent stronger usage intention), followed by the *grid-based scheme* and the *on-screen keyboard* (mean Likert scores of 3.0 and 2.8 respectively). However, it is important to note that only five of the participants assigned to the *on-screen keyboard* did not choose the "Does not apply" option and therefore data with respect to this scheme is scarce. Thus, due to the low number of responses with respect to the on-screen keyboard and the ordinal nature of the data, Fisher's Exact Test with Bonferroni-Holm-adjusted alpha-levels was used for the analysis. An overall Fisher's test found a significant difference in usage intention between the three schemes (FET: $p < .001$). Follow-up tests indicated the presence of significant differences between *Colorwheels* and the *on-screen keyboard* (FET: $p < .019$) as well as *Colorwheels* and the *grid-based scheme* (FET: $p < .001$). In

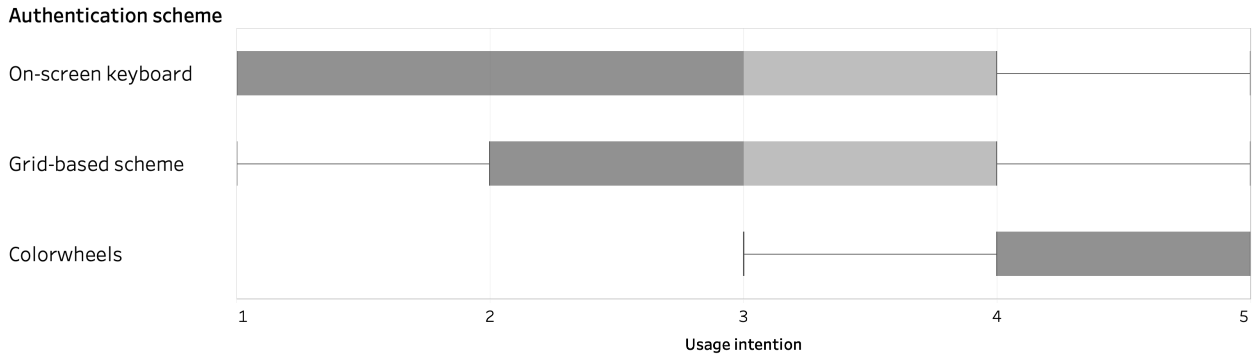


Figure 6.7: Boxplots of the usage intention data measured as 5-point Likert values (higher values represent stronger usage intention).

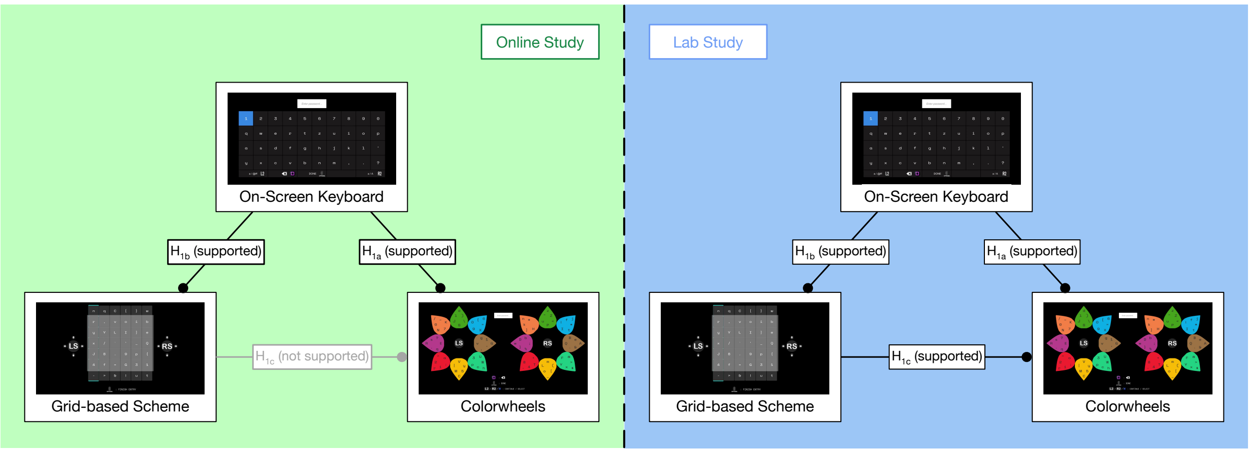


Figure 6.8: Overview of the results of the hypothesis tests of the two studies presented in this chapter. Note that the results of the lab study rather represent the perceived shoulder-surfing resistance with respect to Colorwheels. The ● indicates for each hypothesis which authentication scheme is expected to perform better (i.e. is more shoulder-surfing resistant). Note the difference in the statistical tests used for the analysis, as explained in section 6.2.3 and section 6.3.3.

contrast, the follow-up tests did not indicate a significance difference between the *on-screen keyboard* and the *gird-based scheme* (FET: $p = .062$).

6.4 Discussion

In this section, first the two studies' results and then their limitations are discussed.

6.4.1 Results

Figure 6.8 summarises the results of the two studies with respect to H_{1a-c}. Unsurprisingly, the *on-screen keyboard* does not fare well in terms of shoulder-surfing resistance. Hypotheses H_{1a} and H_{1b} are supported in both studies, indicating that the *grid-based* scheme and our own proposal *Colorwheels* are more shoulder-surfing resistant than the *on-screen keyboard*. The results regarding H_{1c}, i.e. the differences between the *grid-based* scheme and *Colorwheels*, are more ambiguous. While the online study does not indicate a difference, the lab study does. This discrepancy might be due to differences in the study setting. On the one hand, due to the substantial number of participants who had to be excluded from the online study, the sample sizes also vary considerably between the two studies (64 in the online study, 87 in the lab study). On the other hand, the attackers in the lab study were potentially stronger: they (a) had expertise with gamepads (cf. table 6.4), (b) used the schemes before the actual "attack" and (c) saw their previous guesses when writing down the next one (since all guesses were written down on the same sheet of paper). In contrast, the participants of the online study could familiarise themselves only with the explanatory video and the description of the schemes and did not see their previous guesses. Therefore, the participants in the lab study might have had a better understanding of the schemes. They actually used them, which might have facilitated the development of a shoulder-surfing strategy. Overall these results seem to suggest that *Colorwheels*'s shoulder-surfing resistance is more robust, even in the face of stronger attackers or is at least more effective in conveying the perception of a stronger resistance to opportunistic shoulder-surfing attacks. At the same time, these differences indicate that future work should investigate the influence of each of the three differences (usage of scheme before shoulder-surfing task, ability to review earlier guesses, and stopping to guess before having watched all ten videos) on the results in isolation.

Usability-wise, the *on-screen keyboard* fares overall best: it exhibits the highest success rates, the fastest password entry times, and the highest SUS scores. This is unsurprising, since the scheme is deployed in the wild and therefore participants are likely to know it and those having experience with gamepads and game consoles are likely to have used on-screen keyboards before. On the other hand, *Colorwheels* and the *grid-based* scheme were unknown to the participants. Regarding the differences between these two schemes, *Colorwheels* fares better. Efficiency-wise, the combination of visual search and interaction required to enter each character in the *grid-based* scheme seems to take longer than the alternating visual search and interaction tasks in *Colorwheels*. Shiffrin and Schneider [190] showed that people can learn to search in parallel for a particular set of targets, indicating that entry times with *Colorwheels* might decrease over time. Additionally, passwords are usually not entered very frequently in the gamepad context, rendering efficiency overall less important. Thus, the trade-off in terms of efficiency might not impact the user experience to a strong degree. Still, improving this aspect should be a goal of future work. Effectiveness-wise there does not seem to be a significant difference between the *grid-based scheme* and *Colorwheels*. Both schemes can be improved: despite the font in the schemes being explicitly chosen to be monospaced, with serfis, and so that usually similar looking characters could be distinguished (e.g. the letter “O” and the number “0”), it lead to a number of mistakes, where similar characters were confused with one another. Regarding user satisfaction, *Colorwheels* and the *grid-based* scheme performed worse in terms of SUS scores than the *on-screen keyboard*. However, *Colorwheels*’s SUS score (75) is still in the “good” range [16] and significantly exceeds the score of the *grid-based scheme* (64). In addition, *Colorwheels* scores best in terms of usage intention, indicating that in the context of text password entry on gamepads, users are willing to trade usability for increased shoulder-surfing resistance. Albeit, it must be noted that due to the high number of participants in the on-screen keyboard group who chose “Does not apply” as response to the usage intention statement, the data in this respect might be biased.

6.4.2 Limitations

The online study setting did not allow using a large screen (e.g. television) as monitor for the videos. Hence, to remain compatible, the lab study was performed on a smaller 13" screen as well. Participants might have performed differently in settings with large screens. However, it can be argued that this limitation does not negatively impact the two studies presented in this chapter. Participants had the explicit instruction to try and observe the password entry. Thus, the privacy-diminishing effects of large displays as outlined in section 5.1 lose importance. Yet, a comparative lab study using a large screen would allow a closer inspection of this issue. Furthermore, both studies used videos as a compromise due to the intention of conducting an online study. As Aviv et al. [13] report, depending on the scheme under test, a live observation study might lead to more successful observation attacks. Also, all videos were watched directly one after the other and there was no other interaction during password entry. Usually the victim would enter the password only once and (especially in case multiple people are present) the attacker might be distracted by e.g. a conversation. This might render the attackers in the two studies presented in this chapter stronger in this respect than they might be in a real setting. In the lab study the setting was easily controlled. It was ensured that participants followed all instructions. To compensate the lack of direct control over the participants, the online study used additional self-reported attention check questions (e.g. whether participants took notes) and technical measures (e.g. to prevent controlling the video playback and prevent skipping of videos by hiding the button to get to the next page until the video had finished playing). Yet, all measures ran in the participant's browser. Some participants managed to circumvent these measures and had to be excluded from our study. These exclusions were based on the participants' time spent on the pages showing the videos. Additionally, participants who spent very little time familiarising themselves with the authentication schemes in the online study were excluded from the analysis. Thereby, the same time limit was used for all schemes, despite participants being potentially already familiar with the *on-screen keyboard*, to prevent favouring the other schemes, by making sure people spend more time with their descriptions. Moreover, analogously to Aviv et al. [12] participants were not allowed to take any notes during the attack. This design decision was made to increase the consistency between the two studies and among the participants of the online study. Yet, in a real "attack", an attacker might use their smartphone to subtly take text notes, when recording a video is too obvious. Furthermore, the discrepancy in the shoulder-surfing resistance between *Colorwheels* and the *grid-based* scheme in the two studies might be due to the differences in the study setting leading to a stronger attacker in the lab study. Lastly, as all evaluations comparing established schemes (such as the on-screen keyboard) which participants might be familiar with to schemes which participants are unfamiliar with always introduces a bias. Even when the familiarity of the established scheme is considered in the study design as confounding variable and groups are balanced with respect to this aspect, it is not possible to know beforehand, whether familiarity with the established scheme is beneficial with respect to the participants' performance in the new scheme or even decreases it.

Using videos for the evaluation of the shoulder-surfing resistance poses the challenge of selecting representative videos. Therefore, recordings of one single expert user were used. These videos were, however, varied in terms of speed and interaction. The participants saw the videos in a randomised order to mitigate ordering bias. In order to increase the studies' replicability, the videos used for the familiarisation with the schemes and the shoulder-surfing trials are openly available. Another aspect potentially impacting replicability is the choice of the password used in the shoulder-surfing studies. As outlined in section 6.1.2, the password was chosen with great care as to not favour any scheme over the others. However, users seldom choose random passwords for their accounts. Consequentially, the shoulder-surfing results might differ, when another password is used. In particular, dictionary words might impair the resistance of the *grid-based* scheme. Furthermore, replicability among different participants could be increased by supplying a shoulder-surfing strategy (e.g. the optimal strategy for each scheme). However, this necessitates finding the optimal strategy first or might introduce bias if multiple equally optimal strategies exist straining different skills of the user and one is chosen for the study.

The implementations of the three schemes recorded the effectiveness and efficiency metrics automatically and therefore reliably without errors. The satisfaction was measured using the standardised SUS questionnaire. With respect to the recording of the shoulder-surfing metric (i.e. the password guesses) the online study recorded the password entry in a text field not obfuscating the input (i.e. not hiding the password behind symbols such as "*") and in the lab study the experimenter always clarified any legibility issues. Therefore, the results of the studies are as reliable as possible in this respect. However, contrary to expectations, participants of the lab study asked to be allowed to stop the shoulder-surfing task prematurely, if they believed it to be futile. Those participants were allowed to stop the task not only out of ethical considerations (participants were allowed to stop their participation in the study anytime), but also since this might reflect the sentiment of an opportunistic observer in the real world. However, this introduced the subjective perception of success probability into our shoulder-surfing metric for the lab study. Consequentially, a direct comparison of the results of the two studies is not possible.

6.5 Conclusion

This chapter presented two studies investigating the shoulder-surfing resistance and usability of three text-entry schemes in the gamepad context. Building on the results of the previous chapter, the following three schemes were investigated: the on-screen keyboard as incumbent in the gamepad context, the grid-based scheme as shoulder-surfing resistant proposal adapted for the usage with gamepads, and Colorwheels as scheme specifically designed for shoulder-surfing resistant text entry with gamepads.

The results of the two studies show that unfortunately there is a clear trade-off between shoulder-surfing resistance and usability among the evaluated schemes. The *on-screen keyboard* is highly susceptible to shoulder-surfing attacks, but scores highest in all usability metrics. The presented baseline values for the *on-screen keyboard* should serve as thresholds for future evaluations of alternative schemes. With respect to the alternatives tested in the two studies, *Colorwheels* seems to fare better than the *grid-based scheme*. It offers a more robust shoulder-surfing resistance and better usability than the grid-based scheme as well as the overall best scores in terms of usage intention. Therefore, *Colorwheels* seems to represent a design direction which is worth to follow. However, *Colorwheels* should still only be considered as a first step towards a truly deployable scheme, i.e. a scheme where the optimal trade-off between shoulder-surfing resistance and usability has been achieved.

Part III

Secure and Efficient Storage of Passwords in Portfolio Authentication

7 The (t,n)-threshold Verification Scheme for Portfolio Authentication

As becomes apparent from the description of attacks on user authentication in section 2.3, many of those attacks do not rely on guessing the correct password, but instead rely on capturing it in the clear, e.g. in a shoulder-surfing attack in shared spaces as described in chapter 5 and 6. Section 2.3.1 introduced several techniques to counter capture attacks such as shoulder-surfing: using covert channels, obfuscation of the user's input, using indirect input, additional biometric layers, and portfolio authentication. This chapter investigates an open problem of the latter of these techniques: While the results in usability studies of portfolio authentication schemes [62] are promising, one challenge currently faced in the implementation of such schemes is the lack of a verification scheme (section 2.2.4) which allows secure and efficient verification of the input provided by the user. The efficiency metric in the scope of this paper is *storage*, i.e. how much information must be stored to perform the verification of the password elements entered by the user.

Usually the secure verification of user input is implemented using secure key derivation functions (KDF), e.g. salted hashes. Consequently, a naive approach to storage could be to compute all authorised subsets of the password elements and store salted hashes of all the subsets for later comparison to the information entered by the user. However, the required storage of this naive approach grows factorially in the difference of the sizes of the portfolio and the authorised subsets. Therefore, it is inefficient in terms of the required storage.

To address this challenge of secure and efficient storage of passwords in portfolio authentication schemes, this chapter proposes the *(t,n)-threshold verification scheme*. First, the general operations and properties required of verification schemes in the domain of portfolio authentication are summarised (section 7.1). Then, the general concept of the (t,n)-threshold verification scheme is described (section 7.2). It is based on two building blocks – cryptographic secret sharing and key derivation functions – to provide a secure way to derive a common secret for all authorised subsets of the password. Two variants of the (t,n)-threshold verification scheme are presented, based on two different secret sharing schemes, namely Blakley secret sharing [23] (section 7.2.1) and Shamir secret sharing [187] (section 7.2.2). These two (t,n)-threshold verification variants are compared against the naive approach and each other. Section 7.3 provides a discussion of the security properties of the three different approaches (i.e. the two (t,n)-threshold verification variants as well as the naive approach) and section 7.4 presents a storage efficiency comparison of the three schemes. Then, section 7.5 summarises and discusses the findings, describes further applications of the *(t,n)-threshold verification scheme*, and points out areas of future work. Section 7.6 concludes this chapter.

Contributions described in this chapter:

- Compilation of the general operations and properties required of verification schemes in the domain of portfolio authentication.
- The (t, n)-threshold verification scheme as solution to the problem of secure and efficient password storage in portfolio authentication with two concrete variants using two different secret sharing schemes, i.e. Blakley secret sharing and Shamir secret sharing.
- A comparison of the two proposed (t, n)-threshold verification variants regarding security and efficiency against each other and a naive approach.

Parts of the results described in this chapter have been published in:

- P. Mayer and M. Volkamer, "Secure and Efficient Key Derivation in Portfolio Authentication Schemes Using Blakley Secret Sharing", Annual Computer Security Applications Conference (ACSAC), 2015, pp. 431–440.

7.1 Requirements

This section describes the required operations and properties that any verification scheme (i.e. any scheme that operates in the verification phase; cf. section 2.2) used for portfolio authentication should provide. The requirements presented here are mainly derived from findings and properties described by Pieprzyk et al. [165].

7.1.1 Required Operations

There are two primary operations every password verification scheme needs to provide, independently of its use in a portfolio setting. Firstly, the scheme must be able to *create the verification information* by translating the password into verifiable information during enrolment. Secondly, the scheme must offer an operation for the actual *verification.* In the case of traditional text passwords these two operations would be applying an appropriate KDF to the password (creation of the verification information) and the comparison of the value derived from the user input to the value stored during enrolment (verification). Additionally, it must be possible to *delete the verification information* during the termination phase. Since this is a trivial operation, it is excluded in the following.

7.1.2 Optional Operations

While not strictly necessary, additional operations might be desirable when a verification scheme is used with portfolio authentication to allow for a higher usability in some re-enrolment scenarios.

Adjusting the Portfolio Overhead. A change in the security policy of the authentication scheme might require the user to enter a smaller or larger subset of their password for each authentication attempt. This corresponds to adjusting the portfolio overhead. If the portfolio overhead can be adjusted without changing the password, this can prevent users from having to learn completely new passwords in such cases. Instead, only additional challenge-response pairs need to be learned.

Adding and Removing Challenges. In case a challenge-response pair is known to be compromised, it is desirable to allow removal of the respective pair and addition of a replacement pair. Ideally this operation involves only changing the respective pairs, while leaving the remaining portfolio untouched. That way, the user does not have to learn a completely new password but only the respective challenges.

7.1.3 Security Properties

As explained in section 2.1, the primary goal of any authentication scheme is to protect access-restricted resources. Therefore, it is imperative that a verification scheme never impairs an authentication scheme's security properties. In terms of security, the focus of verification schemes lies on two aspects, namely *secure storage* and *guessing resistance.*

Secure Storage. It has long been best practice to not store passwords in the clear [28]. Therefore, every verification scheme is required to not rely on the availability of the password in the clear, i.e. the verification scheme is required to not prevent the password's storage in a cryptographically hashed form. In particular, the verification scheme should offer storage comparable to cryptographically secure salted hashes[1]

Guessing Resistance of Password Elements. The strength of every authentication scheme against guessing attacks depends on its password space. Therefore, a verification scheme should not decrease the space of possible passwords, below the desired security threshold[2]. It needs to be adaptable to the desired strength of the authentication scheme.

Guessing Resistance of Verification Scheme Values[3]. All verification schemes which rely on storing multiple values as verification information instead of just one need to ensure that the guessing resistance of these values of the verification scheme is at least as high as the guessing resistance of the authentication secrets. Otherwise, it might be easier for an attacker to guess the verification scheme values instead of the actual authentication secret. In such a case the overall guessing resistance might be impaired.

7.1.4 Efficiency Properties

Any proposed verification scheme should be more efficient than naive approaches such as the one outlined in the introduction to this chapter. Efficiency in the scope of this chapter refers to *storage*, where the influence of the portfolio overhead should be minimised in order to keep the cost of the shoulder-surfing resistance as low as possible. In contrast, computation time (another typical efficiency metric) is not considered, since it is best practice to artificially prolong the verification to thwart guessing attacks [73].

7.2 The (t,n)-threshold Verification Scheme

The key difference between portfolio authentication schemes (such as those studied by Dunphy et al. [62]) and traditional authentication schemes (such as text passwords) lies in the variability of the allowed user input during the authentication procedure. In portfolio authentication the password is regarded as being composed of elements and every authorised subset of password elements is sufficient to authenticate a user. To address the open problem of secure and efficient password storage in this setting, a novel verification scheme designed for usage with portfolio authentication schemes is introduced. It utilises secret sharing and key derivation functions (KDF) to derive the same secret from all authorised subsets of the password elements. It is denoted (t, n)-threshold verification scheme (in resemblance of (t, n)-threshold secret sharing), where n is the total number of elements in the password and t the threshold size of authorised subsets of password elements. Figure 7.1 depicts an overview of the phases of the user access control procedure as introduced in section 2.2, setting the (t, n)-threshold verification scheme in its generic form into the context of the procedure's different phases. It is important to note that the password elements e are not used directly for input into the (t, n)-threshold verification scheme, but only the derivatives $\overline{e}$ of these elements after application of an appropriate (i.e., cryptographically and computationally secure) key derivation function (KDF).

[1] An explicit recommendation of a specific hash function is not given here, since the security of hash functions changes over time. However, the reader is advised that there are hash functions specifically created for password storage, such as argon2 (`https://github.com/p-h-c/phc-winner-argon2`).

[2] According to Florêncio et al. [73] at the time of writing resisting 10^6 guesses is enough to thwart online guessing attacks and 10^{14} guesses is enough to thwart offline guessing attacks.

[3] This requirement is derived from the work presented in the remainder in this chapter, but already explained here, since it is very relevant as requirement to any other verification scheme that stores multiple vales as verification information.

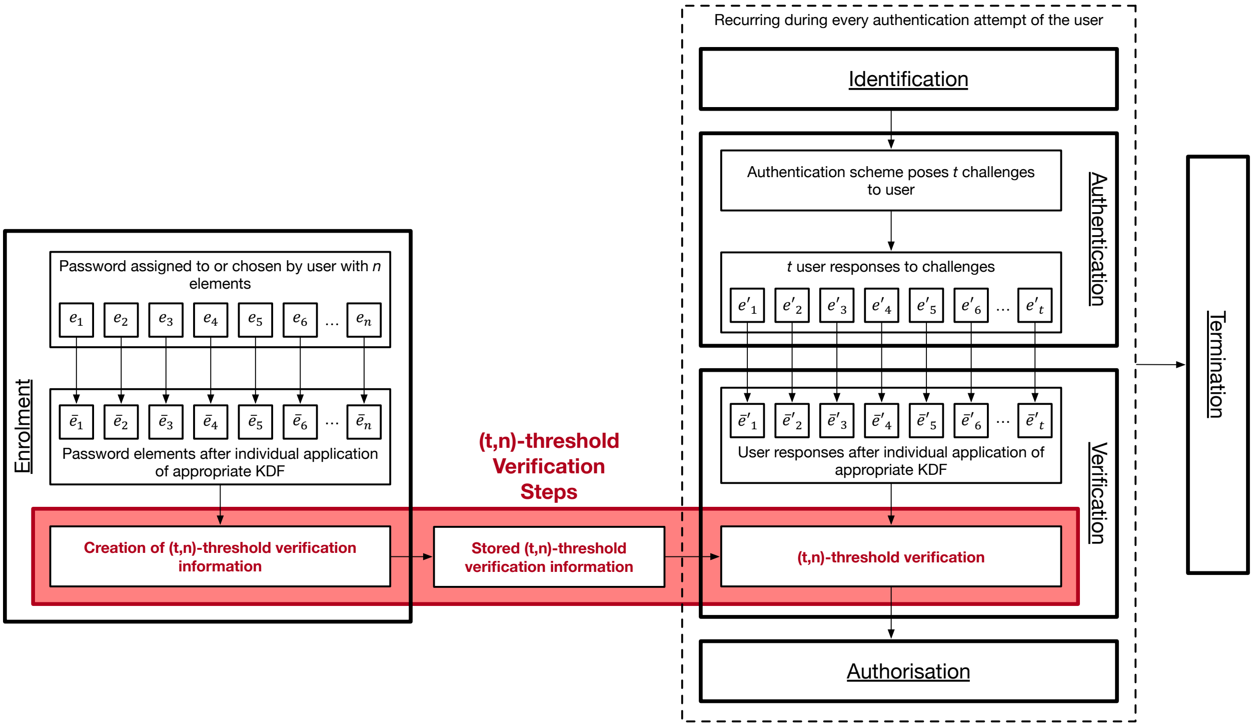

Recurring during every authentication attempt of the user
Identification
Authentication scheme poses t challenges to user
t user responses to challenges
Authentication
Termination
Enrolment
Password assigned to or chosen by user with n elements
Password elements after individual application of appropriate KDF
(t,n)-threshold Verification Steps
User responses after individual application of appropriate KDF
Verification
Creation of (t,n)-threshold verification information
Stored (t,n)-threshold verification information
(t,n)-threshold verification
Authorisation

Figure 7.1: Overview of all phases of the user access control procedure (cf. section 2.2): Enrolment, Identification, Authentication, Verification, Authorisation, and Termination. The steps in which the (t, n)-threshold verification scheme operates are marked in red.

In the remainder of this section, two variants of (t, n)-threshold verification are being described, one based on Blakley secret sharing and one based on Shamir secret sharing.

7.2.1 Variant 1: Based on Blakley Secret Sharing

The first variant of the (t, n)-threshold verification scheme is based on Blakley secret sharing [23]. The reasoning behind choosing Blakley secret sharing, is the observation that Blakley secret sharing – being based on hyperplane geometry – allows to predetermine the individual shares with only slight modifications. Thereby, the derivative pseudo-random values $\bar{e}$ can be used directly as shares.

In the following first the basic working principles of Blakley secret sharing are described. Then, the Blakley secret sharing variant of the (t, n)-threshold verification scheme is described. Figure 7.3 at the end of this section gives an overview of this variant in the context of the phases of the user access control procedure.

Blakley Secret Sharing

Blakley secret sharing is a form of cryptographic (t, n)-threshold secret sharing (cf. section 2.6). Blakley proposed to use hyperplane geometry to solve the cryptographic (t, n)-threshold secret sharing problem [23]. The shared secret is canonically defined as the first coordinate of a randomly chosen point x in a t-dimensional vector space over a Galois field $GF(p)$, where p is a prime (see figure 7.2). For the remainder of this chapter, $i \in \{1, \ldots, n\}$ denotes the index of the party and $j \in \{1, \ldots, t\}$ denotes the coordinate in the t-dimensional vector space.

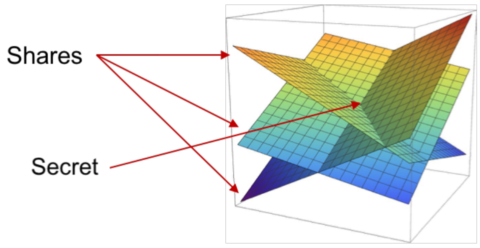


Figure 7.2: Conceptual depiction of the Blakley secret sharing scheme, which is based on hyperplane geometry. The shared secret is a random point in a t-dimensional vector space over a Galois Field $GF(p)$, where p is a prime. The shares are hyperplanes intersecting in the secret point.

Dealing Phase. To distribute the shares, the dealer chooses a sufficiently large prime p and a t-dimensional point

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_t \end{pmatrix}$$

at random. In canonical implementations, its first coordinate x_1 serves as the secret. Then, the shares for each of the n parties are generated. The dealer chooses t values m_{ij} at random for each party and calculates the shares y_i using the equation:

$$m_{i1}x_1 + m_{i2}x_2 + \cdots + m_{it}x_t = y_i \tag{7.1}$$

The resulting $n \times t$ matrix (the entirety of all coefficients m_{ij}) is denoted M. M is public information and does not need to be kept secret [31]. It can be stored in the clear. The dealer distributes only the values y_i to the n parties.

Combination Phase. To reconstruct the secret, t parties need to combine their shares and the respective coefficients from M to form the system of equations

$$M'x = y', \tag{7.2}$$

where y' is the vector of shares provided by the parties and M' is the $t \times t$ matrix of the respective coefficients. This linear system of equations is then solved for x. The reconstructed secret is x_1.

Working Principle of the Variant Based on Blakley Secret Sharing

Enrolment. The basic working principles of Blakley's secret sharing scheme are unchanged for the (t, n)-threshold verification scheme. Each challenge-response pair in the authentication scheme corresponds to one party in the secret sharing scheme. The first step remains choosing a suitable p and a t-dimensional point

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_t \end{pmatrix}.$$

This point x is the common secret and corresponds to the actual password in a traditional text password scheme. Consequently, it should only be stored after the application of an appropriate KDF. Using the complete point x and not only its first coordinate x_1 is a change from Blakley's original procedure introduced after consultation of an expert in the field of cryptography. In the following, the value derived from x is denoted $s = KDF(x)$. Care should be taken when choosing the value p for $GF(p)$. The larger p is, the more resilient the scheme is to guessing attacks (see section 7.3.1 for details).

The rest of the creation procedure deviates slightly from Blakley's secret sharing scheme. Only $t - 1$ coefficients m_{ij} are chosen at random. The remaining coefficient m_{it} is calculated using equation (7.1), by using the values m_{ij} chosen before and as shares y_i the values $\overline{e}_i = KDF(e_1)$. This is a notable deviation from Blakley's original procedure insofar as the distributed share is predetermined by the system and not chosen at random. However, the use of cryptographic hash functions as KDF ensures that all values are indistinguishable from randomly chosen values. Therefore, the security properties of Blakley's scheme remain unchanged (assuming the usage of a cryptographically strong hash function; see section 7.3 for details). The depiction of the enrolment phase in figure 7.3 illustrates the involved operations.

All coefficients of M and the value s are retained and stored for the verification procedure. Together they represent the verification information and correspond to the password hash in a traditional text password setting.

Note that when using Blakley secret sharing in affine geometry as we do here, it is important to ensure during the creation of the verification information that none of the hyperplanes are parallel (i.e. the determinants of all $t \times t$ submatrices M' of M are unequal to zero). Otherwise, there exist M' which are not uniquely solvable, i.e. the secret point x is not unique.

Verification. Whenever a user wants to authenticate to the system, first the user's inputs to the authentication scheme e'_i need to be collected during the authentication phase. The collected e'_i are used in the verification phase to derive the values $\overline{e}'_i$ in the same fashion as during the creation procedure. Then, the $\overline{e}'_i$ and the stored coefficients m_{ij} are used to form the linear system of equations (7.2) which is solved for x. The value s' obtained by application of the respective KDF to x is then compared to the stored common secret s. During verification there is no deviation from Blakley's original procedure. Thus, all authorised subsets of password elements can recover the common secret. The depiction of the verification phase in figure 7.3 illustrates the involved operations.

Adjusting the Portfolio Overhead. Adjusting the portfolio overhead corresponds in the (t, n)-threshold verification scheme to changing the threshold t. In order to adjust the threshold (i.e. transforming the (t, n)-threshold verification scheme into a $(t \pm k, n)$-threshold verification scheme), the matrix M needs to be recreated. This requires the point x or all shares y_i. While an authorised subset is not sufficient for the actual adjustment, any authorised subset of the password can recover x. Using the recovered x, the procedure is then similar to the original creation procedure as outlined in the explanations of the enrolment phase at the beginning of this section. Only the number of columns in M needs to be adjusted according to the desired change (i.e. the index in each row of M is adjusted to $j \in \{1, \ldots, t \pm k\}$).

Adding and Removing Challenges. To add challenge-response pairs (i.e. transforming the (t, n)-threshold verification scheme into a $(t, n + k)$-threshold verification scheme), k rows have to be added to M. To perform the necessary calculations, x has to be reconstructed (which is possible, given any authorised subset of the password). With x the creation procedure using equation (7.1) as outlined in the explanations of the enrolment phase at the beginning of this section can be used to add rows to M.

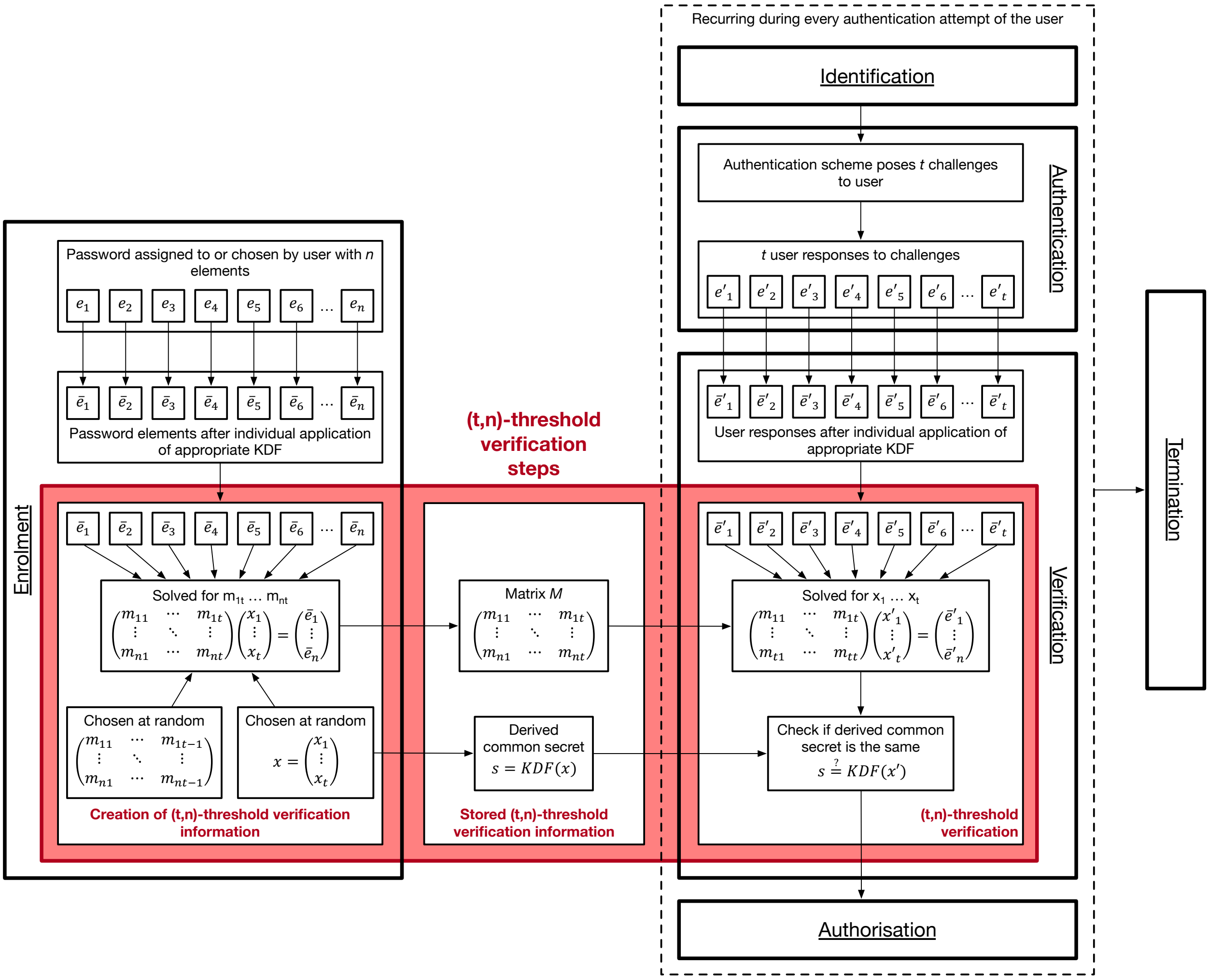

Recurring during every authentication attempt of the user
Identification
Authentication
Authentication scheme poses t challenges to user
t user responses to challenges
Password assigned to or chosen by user with n elements
Password elements after individual application of appropriate KDF
User responses after individual application of appropriate KDF
(t,n)-threshold verification steps
Enrolment
Verification
Termination
Solved for m1t … mnt
Matrix M
Solved for x1 … xt
Chosen at random
Chosen at random
Derived common secret
s = KDF(x)
Check if derived common secret is the same
s ≟ KDF(x′)
Creation of (t,n)-threshold verification information
Stored (t,n)-threshold verification information
(t,n)-threshold verification
Authorisation

Figure 7.3: Overview of the Blakley secret sharing variant of the (t, n)-threshold verification scheme in the context of the phases of the authentication procedure.

Removing challenges from the scheme (i.e. transforming the (t, n)-threshold scheme into a $(t, n-k)$-threshold scheme) is trivial. The respective k rows (i.e. shares and coefficients) are simply removed from M. Of course this is only viable if $n - 1 \geq t$.

7.2.2 Variant 2: Based on Shamir Secret Sharing

The second variant of the (t, n)-threshold verification scheme is based on Shamir secret sharing [187]. The reasoning behind choosing Shamir secret sharing is that it is one of the most widely used secret sharing schemes and therefore implementations are widely available. In this variant, the pseudo-random values $\bar{e}_i$ are not used directly as shares, but rather as keys to encrypt the random shares.

In the following, first the basic working principles of Shamir secret sharing are described. Then, the Shamir secret sharing variant of the (t, n)-threshold verification scheme is described. Figure 7.5 at the end of this section gives an overview of this variant in the context of the phases of the user access control procedure.

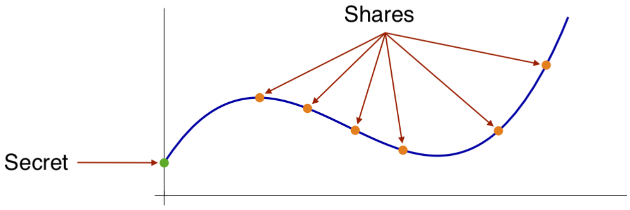


Figure 7.4: Conceptual depiction of the Shamir secret sharing scheme, which is based on polynomial interpolation over $GF(p)$, where p is prime. The shares are points on that polynomial.

Shamir Secret Sharing

Like Blakley secret sharing, Shamir secret sharing is a form of cryptographic (t, n)-threshold secret sharing and consists of two phases: (1) the dealing phase in which each of the n parties is assigned a secret share and (2) the combination phase in which t or more parties can collaborate to reconstruct the secret using their shares.

Shamir proposed to use polynomial interpolation to solve the cryptographic (t, n)-threshold secret sharing problem [187]. The shared secret is canonically defined as the constant term a_0 of a randomly chosen polynomial $f(x) = a_0 + a_1 * x + a_2 * x^2 + \cdots + a_{t-1} * x^{t-1}$ in $GF(p)$, where p is a prime (see figure 7.4). For the remainder of this chapter, $i \in \{1, \ldots, n\}$ denotes the index of the party.

Dealing Phase. To distribute the shares, the dealer chooses a sufficiently large prime p, the shared secret $a_0 \in GF(p)$, and a set of coefficients $a_i \in GF(p)$ at random. Then, the shares for each of the n parties are generated. The dealer chooses n mutually distinct non-zero values $x_i \in GF(p)$ (one for each of the i parties) and calculates the shares y_i using the equation:

$$y_i = f(x_i) = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 + \cdots + a_{t-1} \cdot x_i^{t-1} \tag{7.3}$$

The dealer distributes the values (x_i, y_i) to the n parties.

Combination Phase. To reconstruct the secret, t parties need to combine their shares and use Lagrange interpolation to reconstruct the polynomial f:

$$f'(\bar{x}) = \sum_{i=1}^{t} y'_i \prod_{j=1, j \neq i}^{t} \frac{x'_j - \bar{x}}{x'_j - x'_i} \tag{7.4}$$

where (x'_i, y'_i) are the share values provided by the parties. Using the reconstructed polynomial f', the shared secret $f'(0) = a'_0$ can be calculated.

Working Principle of the Variant Based on Shamir Secret Sharing

Enrolment. The basic working principles of the Shamir secret sharing scheme are unchanged for this variant of the (t, n)-threshold verification scheme. Each challenge-response pair in the authentication scheme corresponds to one party in the secret sharing scheme. The first step remains choosing a suitable p, the secret $a_0 \in GF(p)$, a set of coefficients $a_i \in GF(p)$ for the polynomial f, and n mutually distinct non-zero values $x_i \in GF(p)$. Then, the shares $(x_i, y_i = f(x_i))$ are calculated using equation (7.3).

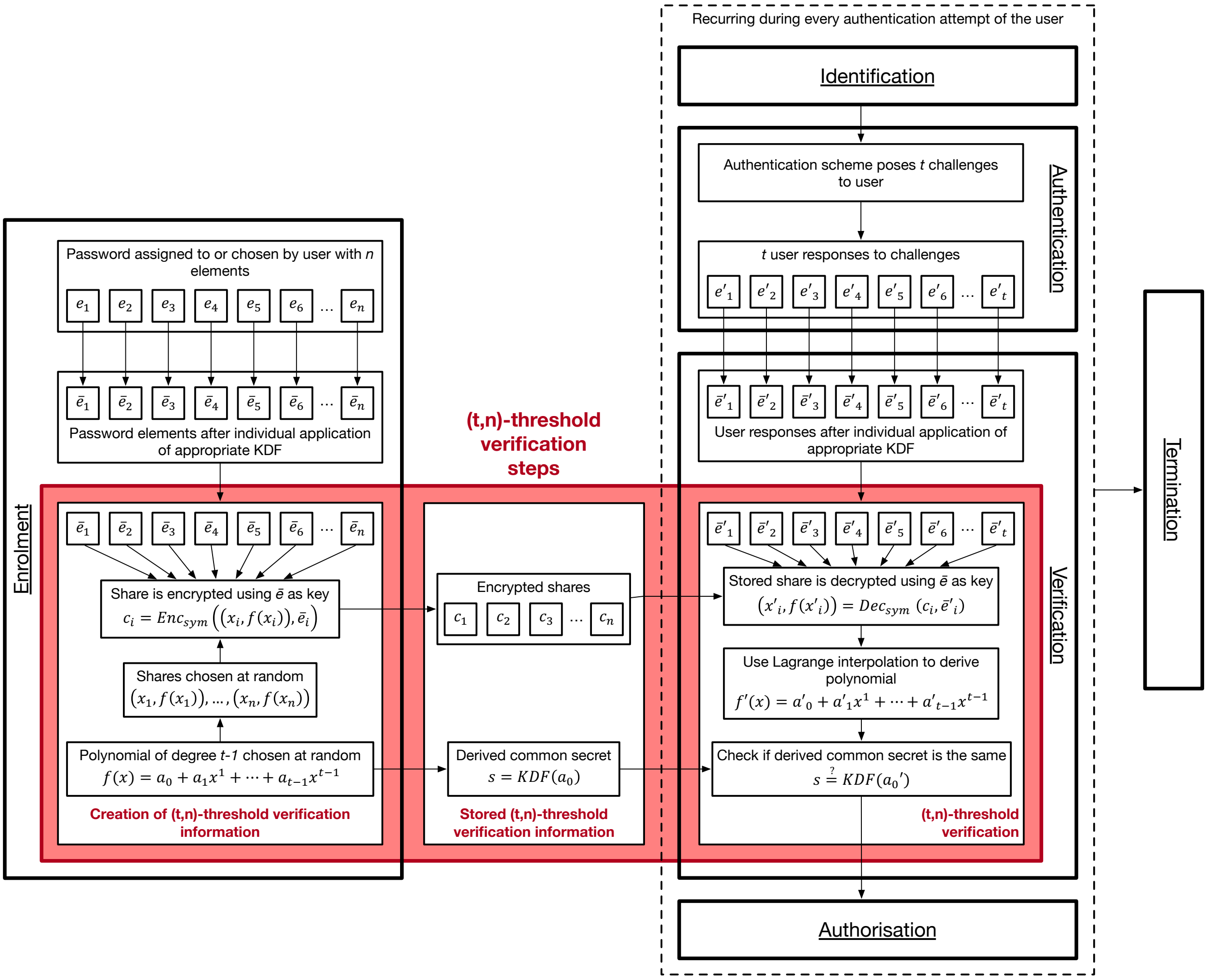

Recurring during every authentication attempt of the user
Identification
Authentication scheme poses t challenges to user
t user responses to challenges
e'1 e'2 e'3 e'4 e'5 e'6 … e't
Authentication
Password assigned to or chosen by user with n elements
e1 e2 e3 e4 e5 e6 … en
ē1 ē2 ē3 ē4 ē5 ē6 … ēn
Password elements after individual application of appropriate KDF
(t,n)-threshold verification steps
ē'1 ē'2 ē'3 ē'4 ē'5 ē'6 … ē't
User responses after individual application of appropriate KDF
Termination
Enrolment
ē1 ē2 ē3 ē4 ē5 ē6 … ēn
Share is encrypted using ē as key
c_i = Enc_sym((x_i, f(x_i)), ē_i)
Shares chosen at random
(x_1, f(x_1)), …, (x_n, f(x_n))
Polynomial of degree t-1 chosen at random
f(x) = a_0 + a_1x^1 + ⋯ + a_{t−1}x^{t−1}
Creation of (t,n)-threshold verification information
Encrypted shares
c1 c2 c3 … cn
Derived common secret
s = KDF(a_0)
Stored (t,n)-threshold verification information
ē'1 ē'2 ē'3 ē'4 ē'5 ē'6 … ē't
Stored share is decrypted using ē as key
(x'_i, f(x'_i)) = Dec_sym(c_i, ē'_i)
Use Lagrange interpolation to derive polynomial
f'(x) = a'_0 + a'_1x^1 + ⋯ + a'_{t−1}x^{t−1}
Check if derived common secret is the same
s ≟ KDF(a_0')
(t,n)-threshold verification
Verification
Authorisation

Figure 7.5: Overview of the Shamir secret sharing variant of the (t, n)-threshold verification scheme in the context of the phases of the authentication procedure.

Thereafter, the values $\bar{e}_i = KDF(e_i)$ are derived from the password elements e_i and used as (symmetric) keys to encrypt the shares for storage with a secure symmetric encryption algorithm:

$$c_i = Enc_{\text{sym}}\Big(\left(x_i, f(x_i)\right), \bar{e}_i \Big)$$

All encrypted shares c_i and the value $s = KDF(a_0)$ are retained and stored for the verification procedure. Together they represent the verification information and correspond to the password hash in a traditional text password setting.

Verification. Whenever a user wants to authenticate to the system, first the user's inputs to the authentication scheme e'_i need to be collected during the authentication phase. The collected e'_i are used in the verification phase to derive the values $\bar{e}'_i$ in the same fashion as during the creation procedure. Then, the $\bar{e}'_i$ are used to decrypt the stored shares

$$\left(x'_i, f(x'_i)\right) = Dec_{\text{sym}}\left(c_i, \bar{e}'_i\right).$$

Using the decrypted shares $\left(x_i', f(x_i')\right)$ and equation (7.4), the polynomial f is then reconstructed. The value $s' = KDF(f(0))$ obtained by application of the respective KDF to $f(0)$ is then compared to the stored common secret s. All authorised subsets of password elements can be used to reconstruct the correct polynomial. The depiction of the verification phase in figure 7.5 illustrates the involved operations.

Adjusting the Portfolio Overhead. Adjusting the portfolio overhead corresponds to changing the threshold t. In order to adjust the threshold, i.e. transforming this second variant of the (t, n)-threshold verification scheme from a (t, n)-threshold verification scheme into a $(t \pm k, n)$-threshold verification scheme, a new polynomial of degree $t \pm k - 1$ needs to be chosen. Therefore, the complete enrolment has to be redone, encrypting the new shares with the derived values $\overline{e}_i$ derived from the password elements e_i.

Adding and Removing Challenges. To add challenge-response pairs, i.e. transforming this second variant from (t, n)-threshold verification scheme into a $(t, n+k)$-threshold verification scheme, additional shares have to be created and encrypted using the new password elements. To perform this operation, the polynomial $f(x)$ has to be reconstructed (which is possible, given any authorised subset of the password). With $f(x)$, additional shares can be created as outlined in the explanations of the enrolment phase at the beginning of this section.

Removing challenges from the scheme (i.e. transforming the (t, n)-threshold scheme into a $(t, n-k)$-threshold scheme) is easily possible by deleting the respective k encrypted shares c_i. Of course this is only viable if $n - 1 \geq t$.

7.3 Security Evaluation

This section deliberates the proposed (t, n)-threshold verification scheme variants regarding the security aspects identified in section 7.1.3. First the guessing resistance is discussed, then the secure storage.

7.3.1 Guessing Resistance

The guessing resistance is determined by the two values p and t. Thereby, p determines the guessing resistance when guessing one value of the (t, n)-threshold verification scheme directly, e.g. any share y_i or the shared secret x or a_0[4]. On the other hand, t determines the guessing resistance when guessing the password elements e_i and subsequently applying the normal (t, n)-threshold verification procedures. In the following, first the security with respect to guessing values of the (t, n)-threshold verification scheme directly is discussed. Thereafter, the security regarding the guessing of password elements is discussed. Note that the former is not relevant for the naive approach outlined in the introduction to this chapter, since it only uses KDF and no other underlying mathematical structure only direct guessing of the password elements is possible. Therefore, the naive approach is only discussed in the context of guessing password elements.

Guessing Resistance of (t,n)-threshold Verification Scheme Values

Both, Blakley and Shamir secret sharing, are perfect secret sharing schemes. Therefore, the guessing resistance of the variables in the (t, n)-threshold verification scheme is directly related to the size of $GF(p)$, i.e.

[4] In perfect secret sharing, an attacker would have to guess at least t shares.

the number of possible values for each variable. There are only p distinct values any variable in the system can represent. Therefore, any attacker needs on average $\frac{p}{2}$ attempts to guess the correct value. This holds obviously if the attacker tries to directly guess x or a_0 respectively.

Additionally, guessing the secret, i.e. x or a_0, does not get easier if a share is known to the attacker. If the attacker tries to guess the correct shares y_i and was (in the worst case) able to obtain all but one share they have again to try on average $\frac{p}{2}$ values for the remaining share. This holds for both variants of the (t, n)-threshold verification scheme.

- **Variant 1 - Blakley secret sharing:** By definition, the linear system of equations (7.2) has one solution, since the determinant of the $t \times t$ matrix M is unequal to zero for all possible vectors of shares y. Consequently, all shares $y_i \in \{1, 2, \ldots, p\}$ are equiprobable.
- **Variant 2 - Shamir secret sharing:** By definition, the Lagrange interpolation polynomials are unique for each set of $t < n$ points. Therefore, each of the p different shares y_i yields a valid polynomial of degree $t - 1$ in equation (7.4). Consequently, all shares $y_i \in \{1, 2, \ldots, p\}$ are equiprobable.

Also, for both variants guessing the share is not easier than a standard brute force attack, even when the number of available shares $p' < p$ is constrained by the authentication scheme. However, the guessing resistance then decreases to $\frac{p'}{2}$.

It is important to choose p for both variants as explained below, to ensure that the space of actually chosen passwords is not shrunk unintentionally. Note that in the following it is assumed that the passwords (i.e. the sets $P = \{e_1, \ldots, e_n\}$) are randomly chosen (i.e. user choice is not modelled). Following the classical information theoretic argumentation in [116] it is of the essence to ensure that

$$H \leq -\sum_{i=1}^{p} \frac{1}{p} \log_2 \left(\frac{1}{p} \right).$$

As stated before, the attacker has to test on average $\frac{p}{2}$ values to to find x_1. Consequently, p should ideally be chosen such that

$$p \geq 2^{H+1}, \tag{7.5}$$

where H is the desired strength against guessing attacks of the authentication scheme in bit. Otherwise, guessing one share is easier than guessing an authorised subset of the password.

Guessing Resistance of Password Elements

When guessing password elements, knowledge-based authentication exists mostly at two security levels: the *PIN-level* and the *password-level*. While the main focus of this thesis is the *password-level*, both of these these two levels are specifically considered in the following deliberations. For both of these levels a sample configuration will be provided below, which does not impair the guessing resistance in comparison to non-portfolio authentication schemes. The portfolio overhead of $o = \frac{3}{2}$ as used in [62] is applied in all configurations.

PIN-level. The PIN-level spans a password space of 10^4 entries. It is used widely, from unlocking smartphones to banking applications. To achieve this security level in a portfolio setting, t is chosen as

$$t_{PIN} = 4.$$

Applying the portfolio overhead $o = \frac{3}{2}$, n is therefore set to

$$n_{PIN} = 6.$$

In accordance with the deliberations regarding the size of p for the two (t, n)-threshold verification variants, it was chosen as $p = 2^{16} - 15$, so that guessing the shared secret, i.e. x or a_0 respectively, is harder than guessing an actual authorised subset of password elements. Using these values for the parameters n, t, and p, the setting of the PIN-level is equivalent to random PINs of length 6, where 4 elements of the PIN have to be entered during the authentication phase over the usual 10 digit alphabet $A = \{0, 1, \ldots, 9\}$. This results in an unchanged overall effort for the attacker of $|A|^t = 10^4 = 2^{\log_2(10) \cdot 4}$. While this effort is below the threshold of 10^6 for online guessing attacks proposed by Florêncio and Herley [73], it offers the same guessing resistance as normal PINs. This holds for the two (t, n)-threshold verification variants as well as the naive approach.

Password-level. The password-level is more ambiguously defined, but can generally be regarded as the security level needed to withstand guessing attacks on traditional text passwords. As Florêncio and Herley [73] note in their review of password research literature and best practices, online guessing is a prevalent and easily performed attack, while offline guessing is only relevant when a specific set of prerequisites are met. They propose 10^6 guesses as threshold to resist online guessing attacks. Since any attacker needs on average to exhaust half the available password space to guess a password, the password space of the password-level of security must be larger than $2 \cdot 10^6$. As Florêncio and Herley [73] note, exceeding this threshold generally does not hold benefits, unless the necessary guesses also exceed the threshold for offline guessing (i.e. 10^{14}). Consequently, the following example of the password-level stays as closely as possible above the online guessing threshold. To that end, t is chosen as

$$t_{Password} = 5.$$

Again, n is determined by applying the portfolio overhead $o = \frac{3}{2}$. It is therefore set to

$$n_{Password} = 8.$$

Using these values for n and t, the setting of the password-level is equivalent to random passwords of length 9, where 6 elements of the password have to be entered during each authentication attempt. In order to reach the desired password space of $2 \cdot 10^6$, an alphabet of size 19 is chosen ($19^5 \approx 2^{21.24} > 10^6$). This is also in line with the findings of Florêncio and Herley [72], who found that password policies used by large Internet companies result in minimum strengths of about 20 to 27 bits. In order to achieve that guessing the shared secret, i.e. x or a_0 respectively, is harder than guessing any authorised subset of the password, p was chosen as $p = 2^{32} - 5$ for the two (t, n)-threshold verification variants.

7.3.2 Secure Storage

Both (t, n)-threshold verification scheme variants provide secure storage of the verification information. Analogously to the procedure common in traditional text password settings, the shared secret, i.e. x or a_0 respectively, is only stored for later verification of the user input after the application of an appropriate KDF. As long as the KDF is secure, the shared secret is secure. In particular, it is recommended to follow best practice and include salting into the KDF. These considerations hold analogously for the naive approach outlined in the introduction to this chapter. Additionally, the shares in the Shamir secret sharing variant have to be encrypted with a secure symmetric encryption algorithm.

7.4 Efficiency Evaluation

Efficiency in the domain of verification is determined by the required storage[5]. Therefore, this section compares the two (t, n)-threshold verification scheme variants against each other and the naive approach outlined in the introduction to this chapter (a short description of the naive scheme is also provided below).

7.4.1 Storage Calculation for the Naive Approach

To the author's knowledge, there is no other verification scheme for portfolio authentication described in published literature. Therefore, in the absence of viable alternatives, the following naive verification scheme already outlined in the introduction to this chapter is used as a baseline for the efficiency comparison.

The naive scheme creates salted hashes for all authorised subsets during the enrolment phase and stores all these hashes. During the verification phase the user's input is hashed and compared to all possible hashes. The verification is successful if the hashed user input is equal to one of the stored hashes.

The required storage in bytes of this naive approach grows factorially in the portfolio overhead and can be determined using the equation

$$b_{\text{naive}} = \left(\begin{array}{c} n \\ \\ t \end{array} \right) \cdot b_{\text{hash}}, \tag{7.6}$$

where n denotes the total number of elements in the password, t denotes the size of the authorised subsets (i.e. the number of elements required during authentication) and b_{hash} denotes the size of one hash in bytes (including the salt).

7.4.2 Storage Calculation for Blakley (t,n)-threshold Verification Variant

The first variant of the (t, n)-threshold verification scheme, i.e. the variant based on Blakley secret sharing, needs to store the shared secret s and the coefficients m_{ij}. The values m_{ij} are in $GF(p)$. As outlined above in section 7.3.1, using a sufficiently large prime number p is essential in order to not decrease the security of the used scheme against guessing attacks. The number of coefficients needed depends on the overall password size (number of elements) and the size t of the authorized subsets. The storage requirement $b_{Blakley}$ in bytes is given by: (a) the $n \cdot t$ coefficients of M, whose size is determined by the number of bytes necessary to store one integer smaller or equal to p (rounded up), plus (b) the stored secret s, whose size is determined

[5] In contrast, it is generally undesirable if password verification is too fast, since this renders guessing attacks easier.

by the number of bytes necessary to store the value (rounded up). Formally this can be expressed using the equation

$$b_{Blakley} = n \cdot t \cdot \left\lceil \frac{\log_2(p)}{8} \right\rceil + b_{\text{hash}}, \tag{7.7}$$

where b_{hash} denotes the size of the hash s of x in bytes (rounded up, including the salt). From this equation it becomes apparent that the required storage grows polynomially in n and t.

7.4.3 Storage Calculation for Shamir (t,n)-threshold Verification Variant

The second variant of the (t, n)-threshold verification scheme, namely the variant based on Shamir secret sharing, needs to store the shared secret s and the encrypted shares c_i. Both parts in the encrypted share (i.e. x_i and $f(x_i)$) are in $GF(p)$. Again a sufficiently large prime number p is essential in order to not decrease the security of the used scheme against guessing attacks. The number of encrypted shares is determined by n. The storage requirement b_{Shamir} in bytes is therefore determined by: (a) the n encrypted shares c_i and the encryption initialisation vectors whose size is determined by the length of the cipher blocks necessary to store two integers smaller or equal to p (rounded up), plus (b) the stored secret s, whose size is determined by the number of bytes necessary to store the value (rounded up). Formally this can be expressed using the equation

$$b_{Shamir} = \left(n \cdot \left\lceil \frac{2 \cdot \log_2(p)}{b_{\text{block}} \cdot 8} \right\rceil + 1 \right) \cdot b_{block} + b_{\text{hash}}, \tag{7.8}$$

where b_{hash} denotes the size of the hash s of a_0 in bytes (rounded up, including the salt) and b_{block} denotes the size of the cipher block in bytes (rounded up). From this equation it becomes apparent that the required storage grows polynomially in n.

7.4.4 Comparison Methodology

To compare the storage requirements, the naive approach as well as both (t, n)-threshold verification variants are evaluated at PIN-level and password-level security. The example configurations described in section 7.3 are used as basis for this storage evaluation.

To explore the properties of the (t, n)-threshold verification variants beyond these two security levels, results for six additional configurations are provided. These additional configurations are based on random strings over the broadly considered alphabet of the 95 characters on a standard US keyboard (see e.g. [100]). The recommended key lengths for high security cryptographic keys given by Eastlake et al. [63] serve as upper bound in terms of password strength. At the time of writing appropriate key lengths are in the 86 to 106 bit range. The parameter p for these configurations is chosen as the largest prime representable with the same number of bytes as the minimum value for p as determined by equation (7.5).

Barker et al. provide recommendations in terms of key length and respective hash functions [17]. At the time of writing SHA-256 and AES-128 are adequate choices. Thus, salted SHA-256 (hash size 32 bytes) is considered as hashing algorithm and AES-128 as symmetric encryption throughout the whole evaluation. Salted SHA-256 hashing is used as KDF for both, the naive and the (t, n)-threshold verification scheme. Following the recommendations of Moriarty et al. [147] the salt was chosen to be 64 bit. Therefore, the size of the hashes and salts is $b_{\text{hash}} = 40$ bytes for both (t, n)-threshold verification variants as well as the naive approach.

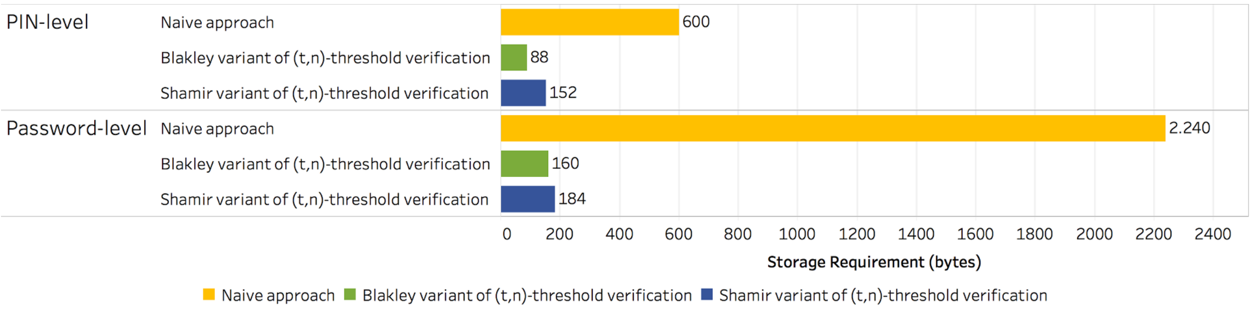


Figure 7.6: The storage requirements of the naive approach and the two (t, n)-threshold verification scheme variants in bytes. Lower values are better.

7.4.5 Comparison Results

In the following, the storage at the aforementioned security levels is described. The PIN-level and the password-level are described in detail (for a summary of the results see figure 7.6). The levels beyond these two are summarised.

PIN-level

Using equation (7.6) with settings for the PIN-level, the overall storage requirement for the naive approach is

$$b_{\text{naive}} = \binom{6}{4} \cdot 40 \text{ bytes} = 600 \text{ bytes}.$$

For the two (t, n)-threshold verification variants, the storage requirements in the PIN-level setting are:

$$b_{\text{Blakley}} = 6 \cdot 4 \cdot 2 \text{ bytes} + 40 \text{ bytes} = 88 \text{ bytes}$$
$$b_{\text{Shamir}} = (6 + 1) \cdot 16 \text{ bytes} + 40 \text{ bytes} = 152 \text{ bytes}$$

It becomes apparent that the variant based on Blakley secret sharing is the most efficient option at the PIN-level.

Password-level

On the password level, the storage requirements of the naive approach as well as the two (t, n)-threshold verification variants are:

$$b_{\text{naive}} = \binom{8}{5} \cdot 40 \text{ bytes} = 2240 \text{ bytes}$$
$$b_{\text{Blakley}} = 8 \cdot 5 \cdot 4 \text{ bytes} + 40 \text{ bytes} = 160 \text{ bytes}$$
$$b_{\text{Shamir}} = (8 + 1) \cdot 16 + 40 \text{ bytes} = 184 \text{ bytes}$$

Table 7.1: The storage requirements in bytes b of the proposed (t, n)-threshold verification scheme for the additional configurations. The values for the naive approach are also given for reference. H is the desired strength of the authentication scheme. n and t are the portfolio parameters.

H	n	t	b_{naive}	b_{Blakley}	b_{Shamir}
39,42	9	6	3360	364	200
52,56	12	8	19800	712	248
65,70	15	10	120120	1390	536
78,84	18	12	742560	2200	632
91,98	21	14	4651200	3568	728
105,12	24	16	29418840	5416	824

Analogously to the PIN-level, at the password-level the variant based on Blakley secret sharing is the most efficient option at the password-level.

Beyond password-level security

Table 7.1 shows the storage requirements of the six additional configurations based on the settings outlined in section 7.4.4. From the data it becomes apparent that the more authorised subsets there exist for one password, the more storage is required by both, the naive approach and the (t, n)-threshold verification variants. However, due to the different growths of the required storage in both approaches (polynomial vs. factorial) the difference between the naive approach and the (t, n)-threshold verification variants steadily increases. The (t, n)-threshold verification variants are substantially more efficient in terms of storage than the naive approach for large numbers of authorised subsets. However, the differences between the two variants become also more pronounced. While for the PIN-level and password-level the variant based on Blakley secret sharing is the more efficient choice, this reverses for larger configurations (i.e. larger values of n, t, and p).

7.5 Discussion

This section discusses the (t, n)-threshold verification scheme presented in this chapter along the results of the security and efficiency evaluation as well as the limitations of the two evaluations. Thereafter, it highlights possible next steps for the continuation of this work.

7.5.1 Results

This chapter introduces the (t, n)-threshold verification scheme, a novel verification scheme to facilitate secure and efficient verification in portfolio authentication schemes. Two (t, n)-threshold verification variants are proposed. The first variant is based on Blakley secret sharing and therefore uses hyperplane geometry to derive a shared secret from all authorised subsets of password elements. The second variant is based on Shamir secret sharing and therefore uses Lagrange interpolation to derive a shared secret. The storage efficiency of both variants was evaluated against a naive approach and their security properties discussed.

In terms of storage efficiency, the comparison revealed that both of the proposed (t, n)-threshold verification variants require substantially less storage space for the verification information than the naive approach. The required storage of the (t, n)-threshold verification scheme grows only polynomially, while the storage of the naive approach grows factorially in the length of the password and the size of the authorised subsets. The efficiency in terms of storage space can be regarded as the most important trait of the (t, n)-threshold verification scheme. The storage requirement of the (t, n)-threshold verification scheme is four times (variant 2) to six times (variant 1) smaller in the PIN-level security setting and twelve times (variant 2) to fourteen times (variant 1) smaller in the password-level security setting than for the naive approach. The additional configurations let this difference become even more apparent: the longer the password is (assuming the same portfolio overhead), the larger the difference becomes.

Furthermore, differences between the two (t, n)-threshold verification variants became apparent. While the first variant (based on Blakley secret sharing) is more efficient for small n, t, and p, the second variant (based on Shamir secret sharing) is more efficient for larger configurations. This is due to the usage of block ciphers in the second variant. The cipher's block size leads to storage inefficiencies in small configurations, since smaller amounts of data are then simply padded with zeros and take the same space as larger amounts. However, depending on the specific application scenario, it might also be possible to further increase the storage efficiency of the second variant by (a) decreasing the block size of the cipher, (b) using a simple counter instead of a random IV, or (c) using a secure stream cipher instead of the block cipher. Yet, a full investigation of such variants constitutes future work. Overall, the differences between the two variants show that it is of the essence to be aware of one's exact requirements in terms of the parameters n, t, and p when deciding for one of the two variants in terms of efficiency.

It is also important to acknowledge that using portfolio authentication with either approach imposes a penalty to storage efficiency. The storage requirement of both (t, n)-threshold verification variants is much larger than in the non-portfolio scenario, where only one hash of 40 bytes would have to be stored for a traditional password in this comparison. Yet, this additional storage requirement is not unreasonable: assuming the password-level security and an organisation with 30.000 user accounts, traditional text passwords would require 1.2 megabytes of storage capacity ($40 \cdot 30'000 = 1'200'000$ bytes) and the use of (t, n)-threshold verification would require 4.8 megabytes of storage ($160 \cdot 30'000 = 4'800'000$ bytes).

In terms of security, the properties regarding secure storage and the guessing resistance of the (t, n)-threshold verification variants were investigated. Regarding the secure storage of the authentication information it could be shown that the choice of a secure KDF allows secure storage for both (t, n)-threshold verification variants. Regarding the guessing resistance, the verification scheme can be adapted to the desired strength of the authentication scheme by choosing p large enough according to equation (7.5). However, it becomes apparent from equation (7.7) and equation (7.8) that balancing the storage requirements and the resistance to guessing attacks by carefully choosing the parameter p can be of the essence. For PIN-level or password-level secrets and in scenarios where storage efficiency is not of utmost importance, this is not critical. However, when longer secrets are stored and storage efficiency is of the essence, the advantage in terms of storage can decrease.

Additionally, the (t, n)-threshold verification variant based on Shamir secret sharing relies not only on an appropriate KDF, but also on a symmetric cipher, rendering implementations potentially more complex. Therefore, implementers should mind the different mathematical structures (hyperplane geometry versus Lagrange interpolation) underlying the two variants and base their decision of which variant to implement not only on efficiency considerations.

7.5.2 Limitations

While this chapter discusses the security properties of the (t, n)-threshold verification scheme, it is conceivable that detailed cryptanalyses using techniques such as lattice-based algorithms or a modelling using integer linear programming might potentially necessitate to re-evaluate the choice of the parameters n, t, and p for either of the two (t, n)-threshold verification variants. Thus, it might be a worthwhile direction of future work to apply such techniques.

With respect to the efficiency analysis, several configurations based on the usual portfolio overhead of $o = \frac{3}{2}$ were presented. Depending on the specific use case, evaluations including other portfolio overhead values and different sizes for the alphabet might be required for practitioners to get an overview of possible configurations more easily. Also, salt values longer than the 64 bits chosen from the recommendations in the literature might increase security, but also impact storage efficiency. The 64 bits represent the lower bound of the recommendations in the literature and this value was chosen, since it minimises the impact of the salt on the overall storage comparison.

7.5.3 Next Steps

From the work presented in this chapter, two lines of future work emerge:

- To illustrate the application of the (t, n)-threshold verification scheme to different authentication schemes, concrete use cases should be developed which can serve as blueprints for researchers and practitioners seeking to implement portfolio authentication schemes based on (t, n)-threshold verification.
- To strengthen the evaluation of the security properties, additional cryptanalyses using lattice-based algorithms or integer linear programming modelling should be conducted by experts in the respective fields.

To facilitate the usage of the (t, n)-threshold verification scheme in portfolio authentication schemes, the next chapter presents three use cases of the scheme's application as blueprints for future implementations. The additional cryptanalyses are left explicitly as future work for experts in the respective fields.

7.6 Conclusion

This chapter introduced the (t, n)-threshold verification scheme. It serves as an important enabler for new and existing portfolio authentication schemes, offering secure and efficient password verification. The two (t, n)-threshold verification variants offer several points of differentiation.

The first variant is based on Blakley secret sharing and therefore uses as underlying mathematical structure hyperplane geometry. It seems to be the most efficient option for small n, t, and p. Also, it easily offers all optional operations. In terms of security it only relies on the security of the used KDF.

The second variant is based on Shamir secret sharing and therefore uses Lagrange interpolation as underlying mathematical construct. It seems to be the most efficient option for higher security configurations (i.e. beyond PIN-level security). Analogously to variant 1, it offers all optional operations. However, in terms of security it not only relies on the security of the used KDF, but also on the security of the used symmetric block cipher.

8 Use Cases of the (t,n)-threshold Verification Scheme

As described in the previous chapter, the (t, n)-threshold verification scheme is an enabler of secure and efficient storage for a wide variety of authentication schemes. In this chapter three use cases of applying the (t, n)-threshold verification scheme are presented. Two of these use cases represent the original knowledge-based authentication application scenario: graphical recognition-based passwords (section 8.1) and partial passwords (section 8.2). In addition, the third use case will discuss the application of the (t, n)-threshold verification scheme in a different scenario: the ZeTA authentication scheme which is based on the human ability to connect semantically related concepts (section 8.3). In each use case the mappings of the elements in the respective authentication scheme onto the elements of the (t, n)-threshold verification scheme are described along the procedures necessary for enrolment as well as authentication and verification.

Last but not least, the use cases as well as further areas of application spanning different types of authentication schemes and enabling redundancy for authentication factors in case of loss, damage or theft are discussed (section 8.4). Section 8.5 concludes this chapter.

Contributions described in this chapter:

- Description of three use cases for the (t, n)-threshold verification scheme: graphical recognition-based passwords, partial passwords, and the ZeTA authentication scheme.
- Discussion of further areas of application beyond the three use cases described in detail.

Parts of the results described in this chapter have been published in:

- P. Mayer and M. Volkamer, "Secure and Efficient Key Derivation in Portfolio Authentication Schemes Using Blakley Secret Sharing", Annual Computer Securit Applications Conference (ACSAC), 2015, pp. 431–440.
- P. Mayer and M. Volkamer, "Poster: Secure Storage of Masked Passwords", European Symposium on Security and Privacy Posters (Euro S&P Posters), 2017
- A. Gutmann, K. Renaud, J. Maguire, P. Mayer, M. Volkamer, K. Matsuura, and J. Muller-Quade, "ZeTA-Zero-Trust Authentication: Relying on Innate Human Ability, Not Technology", European Symposium on Security and Privacy (Euro S&P), 2016, pp. 357–371.

8.1 Use Case 1: Graphical Recognition-Based Passwords

This section presents the first use case for the (t, n)-threshold verification scheme: *graphical recognition-based passwords*. Graphical recognition-based passwords were already introduced in section 2.1.1: the password in these schemes is composed of graphical elements and instead of freely recalling the password, users have to decide whether presented information is familiar or not (i.e. recognise if the information is part of the password).

As outlined in section 2.3, it was proposed to use portfolio authentication as a measure to mitigate shoulder-surfing risks of graphical recognition-based passwords [62]. However, the secure storage was left as an open problem. In the following it is outlined, how (t, n)-threshold verification can be used to securely store graphical recognition-based passwords. First the enrolment and then the authentication and verification procedures are explained. Figure 8.1 depicts an example of the full procedure.

Enrolment. The graphical recognition-based password P is split up in its elements, i.e. the single images $\hat{e}_i$. For each image its index in the grid j and an identifier of the respective grid k are concatenated to create the elements $e_i = j.k$. This step ensures that the elements e_i do not only depend on the size of the grid, but also on the number of grids. Following the procedure of the (t, n)-threshold verification scheme, these elements are then hashed to generate the derivatives $\overline{e}_i$. Then, the shared secret x or a_0 (depending on the used variant) is chosen and its derivative $s = KDF(x)$ or $s = KDF(a_0)$ (depending on the used variant) stored for later verification. Using the values $\overline{e}_i$ and either x or a_0 (depending on the used variant),

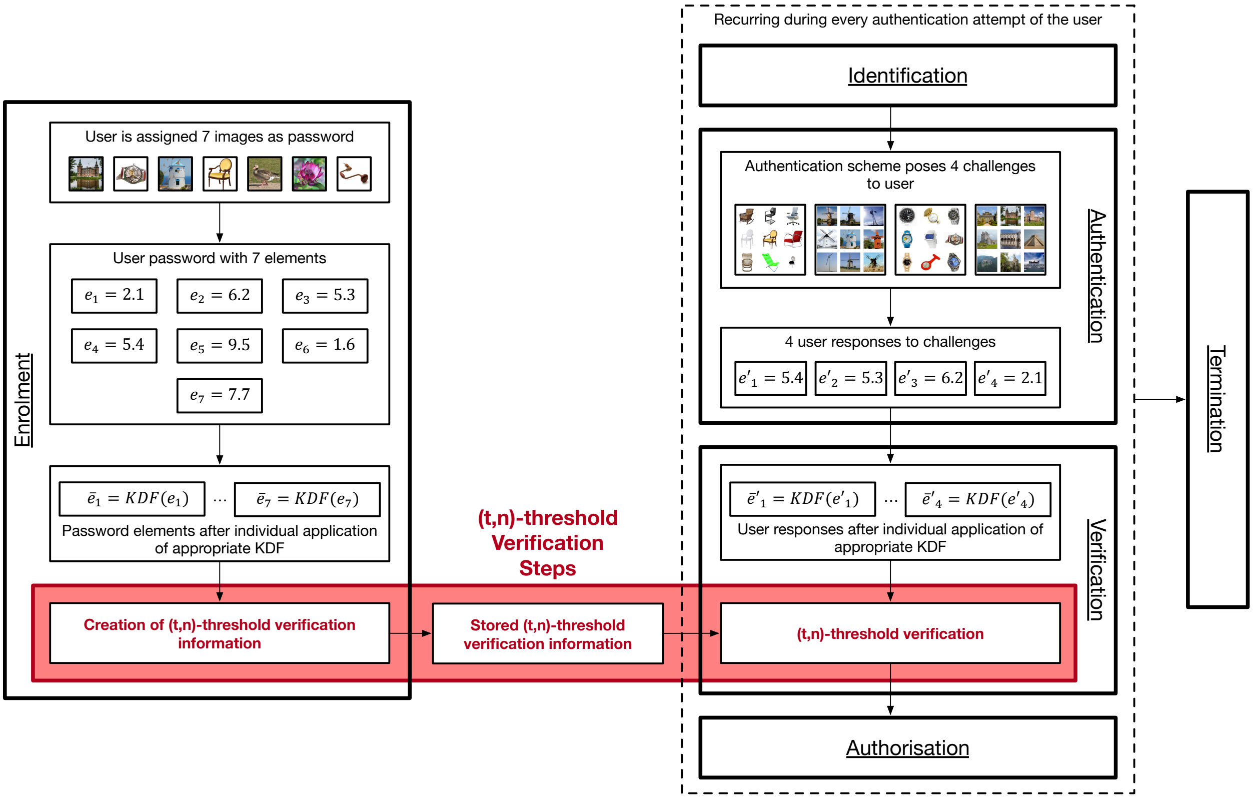

Recurring during every authentication attempt of the user
Identification
Authentication scheme poses 4 challenges to user
Authentication
4 user responses to challenges
$e'_1 = 5.4$
$e'_2 = 5.3$
$e'_3 = 6.2$
$e'_4 = 2.1$
Termination
Enrolment
User is assigned 7 images as password
User password with 7 elements
$e_1 = 2.1$
$e_2 = 6.2$
$e_3 = 5.3$
$e_4 = 5.4$
$e_5 = 9.5$
$e_6 = 1.6$
$e_7 = 7.7$
$\bar{e}_1 = KDF(e_1)$ … $\bar{e}_7 = KDF(e_7)$
Password elements after individual application of appropriate KDF
(t,n)-threshold Verification Steps
Creation of (t,n)-threshold verification information
Stored (t,n)-threshold verification information
(t,n)-threshold verification
$\bar{e}'_1 = KDF(e'_1)$ … $\bar{e}'_4 = KDF(e'_4)$
User responses after individual application of appropriate KDF
Verification
Authorisation

Figure 8.1: Overview the different phases of the authentication procedure when applying the (t, n)-threshold verification scheme to graphical recognition-based passwords. Here the KDF is applied directly on the images instead of their indices in the grid. This deviation from the textual description is intended to increase the clarity of the illustration. The index of the respective grid is used as normal in this step and denoted Gx for the xth grid.

the enrolment procedure for the respective variant of the (t, n)-threshold verification scheme is applied (i.e. either the matrix M is generated using the method described in section 7.2.1 or the encrypted shares c_i are created using the method described in section 7.2.2) and the result stored alongside the value s.

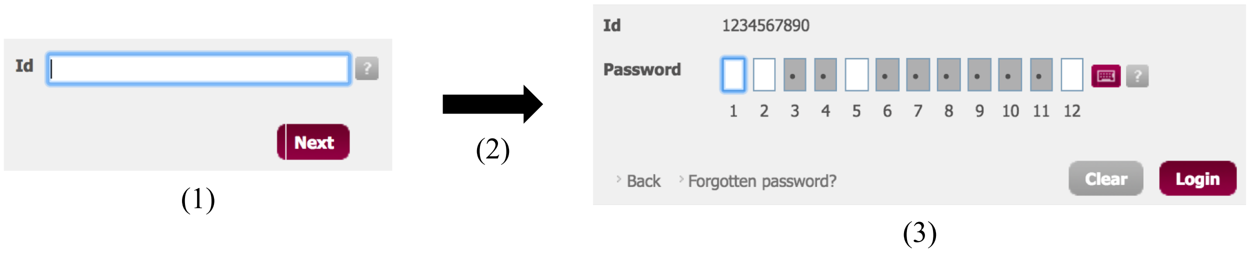
Authentication and Verification. For the authentication, the system generates a set of t challenges. Each challenge is represented by one grid (including the respective password element it contains). The t challenges are displayed to the user who clicks the image $\hat{e}'_i$ belonging to the password. Then, analogously to the enrolment, the indices j and k are concatenated to form the elements $e'_i = j.k$ and then hashed to derive the values $\overline{e}'_i$. The $\overline{e}'_i$ are then used as input for the verification procedure of the respective (t, n)-threshold verification variant (i.e. to solve the linear system of equations $Mx' = y'$ for x' or use Lagrange interpolation to restore a'_0). As last step, it is verified that the value $s' = KDF(x')$ or $s' = KDF(a'_0)$ (depending on the used variant) matches the previously stored s. If the two values match, the user has entered the correct authorised subset as response to the challenges and the authentication attempt is successful.

8.2 Use Case 2: Partial Passwords

This section presents the second use case for the (t, n)-threshold verification scheme: *partial passwords*. Partial passwords are a special form of authentication based on text passwords. When using partial passwords, the users are required to enter only a randomly chosen subset of the password's characters instead of the complete password. In that, partial passwords are the result of a straight forward application of portfolio authentication to text passwords. Figure 8.2 depicts the procedure as seen by the user on a banking website. Sometimes this technique is also used in two-factor schemes [88]. This section outlines how to apply the (t, n)-threshold verification scheme to partial passwords.

In the following, first the enrolment and then the authentication and verification procedures are described, when applying (t, n)-threshold verification to partial passwords. Figure 8.3 depicts an example of the full procedure for the first variant of (t, n)-threshold verification.

Enrolment. First, the textual password P is split up into its characters and each character $\hat{e}_i$ is concatenated with its index i in the password to create the elements $e_i = \hat{e}_i.i$. This step ensures, that in order to guess a share, not only the right character, but also its correct position in the password is required. Following the procedure of (t, n)-threshold verification, these elements are hashed to generate the derivatives $\overline{e}_i$. Thereafter, the shared secret x or a_0 (depending on the used variant) is chosen and its derivative $s = KDF(x)$ or



Id 1234567890

Figure 8.2: A typical login procedure using partial passwords: (1) the user specifies her/his user name (Id in this example), (2) the user clicks "Next" to proceed to the password entry, (3) randomly selected characters of the password have to be entered (in this example, the first, second, fifth, and twelfth character). Screenshots from `https://aliorbank.pl/hades/do/Login` (accessed on 2017-01-24).

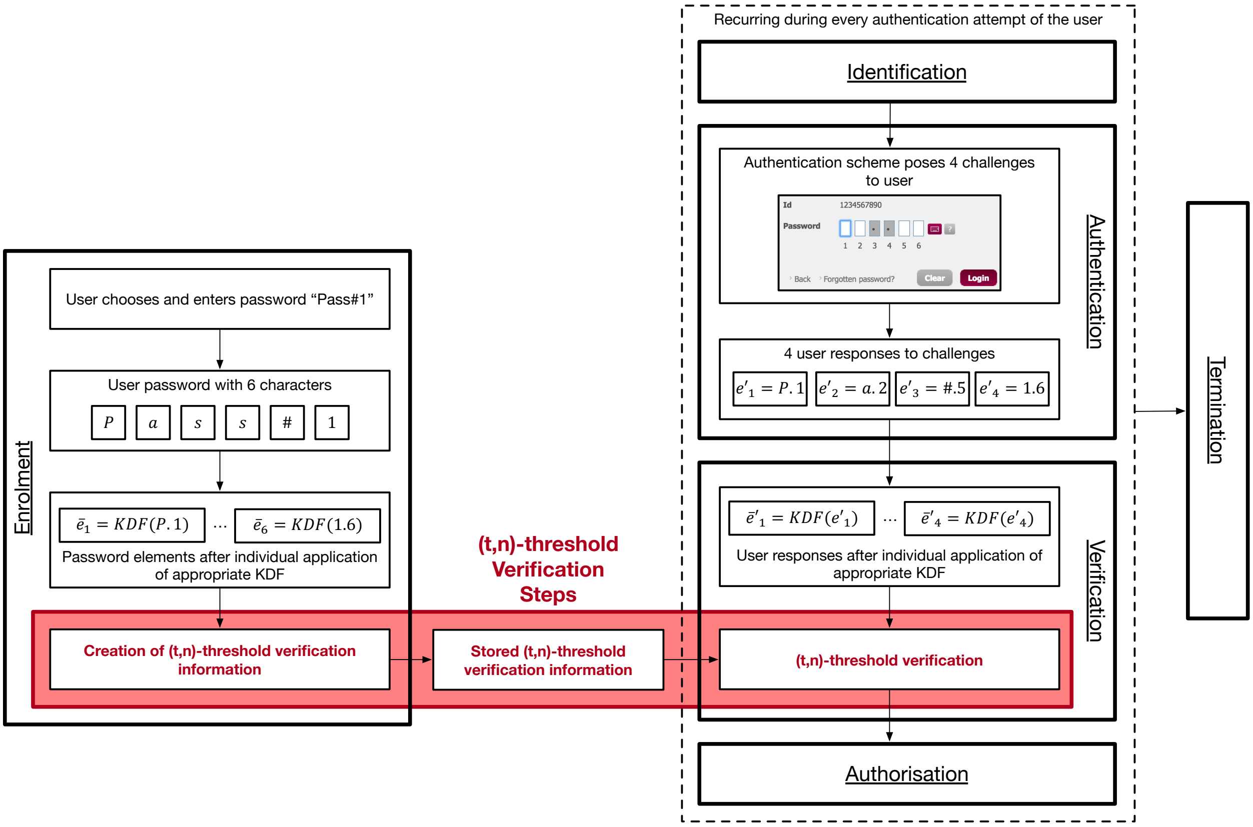

Recurring during every authentication attempt of the user
Identification
Authentication scheme poses 4 challenges to user
Id
1234567890
Password
1 2 3 4 5 6
Back
Forgotten password?
Clear
Login
4 user responses to challenges
$e'_1 = P.1$
$e'_2 = a.2$
$e'_3 = \#.5$
$e'_4 = 1.6$
Authentication
Termination
Enrolment
User chooses and enters password “Pass#1”
User password with 6 characters
P
a
s
s
#
1
$\bar{e}_1 = KDF(P.1)$
…
$\bar{e}_6 = KDF(1.6)$
Password elements after individual application of appropriate KDF
(t,n)-threshold Verification Steps
$\bar{e}'_1 = KDF(e'_1)$
…
$\bar{e}'_4 = KDF(e'_4)$
User responses after individual application of appropriate KDF
Verification
Creation of (t,n)-threshold verification information
Stored (t,n)-threshold verification information
(t,n)-threshold verification
Authorisation

Figure 8.3: Overview the different phases of the authentication procedure when applying the (t, n)-threshold verification scheme to partial passwords.

$s = KDF(a_0)$ (depending on the used variant) stored for later verification. Using these values $\overline{e}_i$ and either x or a_0 (depending on the used variant), the enrolment procedure for the respective variant of the (t, n)-threshold verification scheme is applied (i.e. either the matrix M is generated using the method described in section 7.2.1 or the encrypted shares c_i are created using the method described in section 7.2.2) and the result stored alongside the value s. Note that the storage verification information allows determining the length of the password through the number of rows in M or number of encrypted shares c_i. However, as becomes apparent from figure 8.2 this does not reveal information about the password which is not visible on the login interface anyway and therefore does not impair the security properties of partial passwords.

Authentication and Verification. To authenticate a user, the system generates a challenge by randomly selecting t positions $i \in \{1, 2, \ldots, n\}$ in the password, where $n = |P|$. This challenge is displayed to the user who has to enter the respective characters $\hat{e}'_i$ of the password. Then, analogously to the enrolment, the pair of $\hat{e}'_i$ (supplied by the user) and the respective i (supplied by the server) are concatenated and hashed on the server to derive the values $\overline{e}'_i$. These values are then used as input for the verification procedure of the respective (t, n)-threshold verification variant (i.e. to solve the linear system of equations $Mx' = y'$ for x' or use Lagrange interpolation to restore a'_0). In the last step, it is verified that the value $s' = KDF(x')$ or $s' = KDF(a'_0)$ (depending on the used variant) matches the previously stored s. If the two hashes match, the user has entered the correct authorised subset and the authentication attempt is successful.

8.3 Use Case 3: ZeTA

This section presents the third use case for the (t, n)-threshold verification scheme: *ZeTA*. ZeTA (zero trust authentication on untrusted channels) is an authentication scheme designed to be secure in spite of untrusted devices and communications or shoulder-surfing attacks [89][1]. This authentication scheme harnesses the human capability to build up semantic networks of related concepts and is thus based on innate human-based computation.

A secret in ZeTA is built up of two or more concepts (e.g. words) and logical connections between them, such as *yellow OR wheel*. During authentication, the user is then asked in a challenge-response fashion, whether another concept is related. Suppose the secret is *yellow OR wheel* and the challenge was *sunflower*, then the user would have to answer *yes* (or *TRUE*). If the challenge for the same secret was *water*, the user would have to answer *no* (or *FALSE*). Consequently, one important prerequisite of ZeTA is the availability of a sufficiently large knowledge base of concepts. From this knowledge base, the verifier must be able to create challenges and check the user's response against the secret. The user has to respond to multiple challenges until a previously set certainty threshold is reached. This certainty threshold determines the security level of the ZeTA scheme by defining the probability with which an attacker is able to guess the correct answers to the aforementioned yes/no-challenges. Depending on the implementation it is also possible to introduce noise by requiring the user to respond to a certain portion of the challenges incorrectly.

This section outlines how (t, n)-threshold verification can be used to prevent needing to store ZeTAs authentication secret in the clear to perform the necessary checks of the responses. In the following, first the enrolment and then the authentication and verification procedures are described, when applying (t, n)-threshold verification to the ZeTA scheme. Figure 8.4 depicts the full procedure.

Enrolment. Analogously to the previous two use cases a shared secret x or a_0 (depending of the used variant) is chosen and its derived value $s = KDF(x)$ or $s = KDF(a_0)$ (depending on the used variant) stored for later verification. However, unlike the previous two use cases of the (t, n)-threshold verification scheme, this third use case does not utilise the authentication secret directly to derive the verification information. Instead all concepts in the knowledge base are used as values $\hat{e}_i$ and form a tuple with the respective correct answer with regard to the users secret. Revisiting the example from before and assuming the secret to be *yellow OR wheel*, the concept $\hat{e}_1 = \text{“}sunflower\text{”}$ would then form the tuple $e_1 = (\text{“}sunflower\text{”}, TRUE)$ and the concept $\hat{e}_2 = \text{“}water\text{”}$ would then form the tuple $e_2 = (\text{“}water\text{”}, FALSE)$. From these tuples are then derived the values $\overline{e}_i = KDF(e_i)$. Using the values $\overline{e}_i$ and either x or a_0 (depending on the used variant), the enrolment procedure for the respective variant of the (t, n)-threshold verification scheme is applied (i.e. either the matrix M is generated using the method described in section 7.2.1 or the encrypted shares c_i are created using the method described in section 7.2.2) and the result stored alongside the value s.

Authentication and Verification. To authenticate the user, the system then chooses any concept $\hat{e}_i$ from the knowledge base as challenge (potentially marking the chosen concept as used to prevent replay attacks). The user's response r_i is then used to form the tuple $e'_i = (\hat{e}_i, r_i)$ and its derivative value $\overline{e}'_i = KDF(e_i)$. These values are then used as input for the verification procedure of the respective (t, n)-threshold verification variant (i.e. to solve the linear system of equations $Mx' = y'$ for x' or use Lagrange interpolation to restore a'_0). Last but not least, it is verified that the value $s' = KDF(x')$ or $s' = KDF(a'_0)$ (depending on the

[1] Only section 4.3 of [89] represents a contribution of this thesis's author.

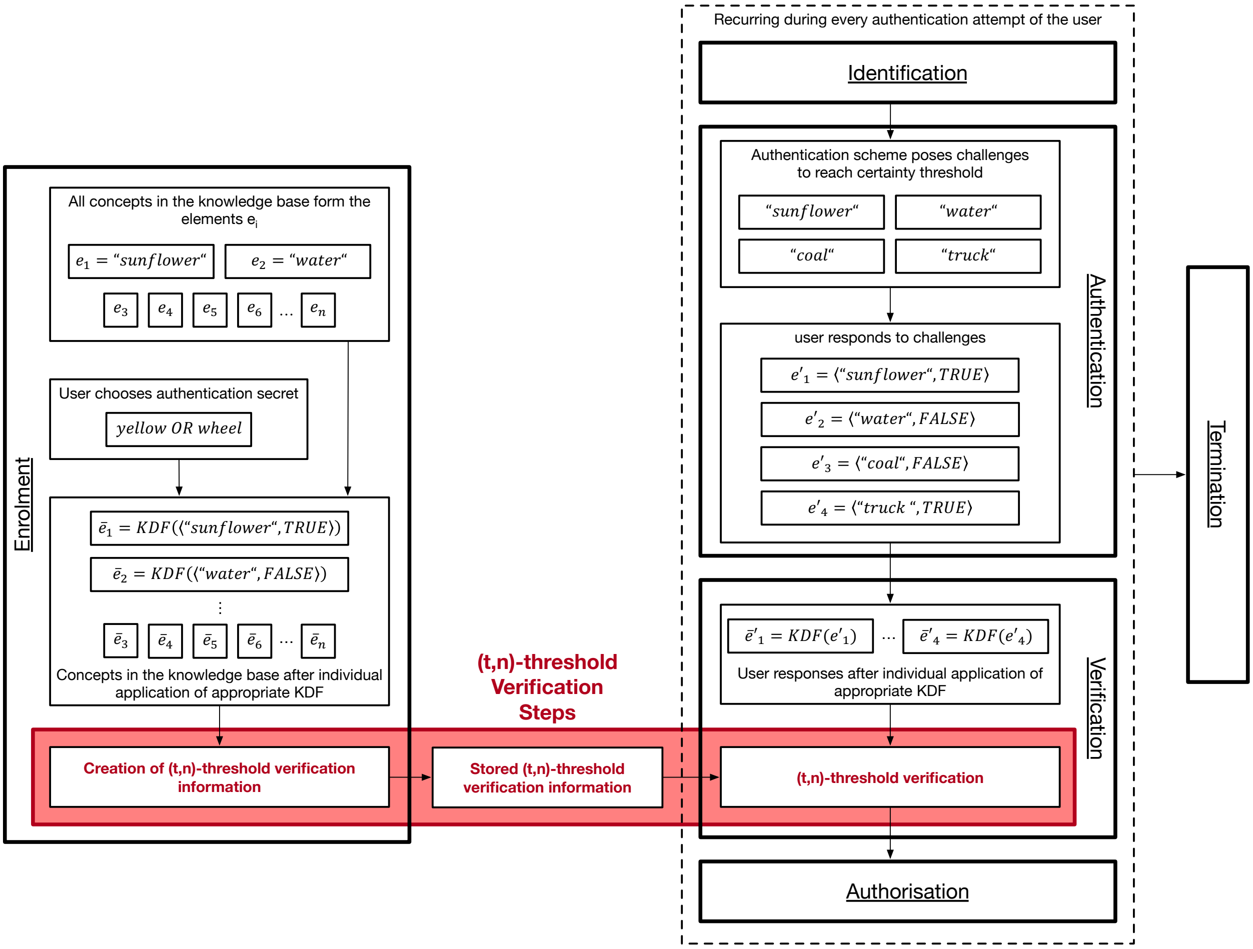

Enrolment
All concepts in the knowledge base form the elements e_i
e_1 = "sunflower"
e_2 = "water"
e_3 e_4 e_5 e_6 … e_n
User chooses authentication secret
yellow OR wheel
$\bar{e}_1 = KDF(\langle$"sunflower", TRUE$\rangle)$
$\bar{e}_2 = KDF(\langle$"water", FALSE$\rangle)$
$\bar{e}_3$ $\bar{e}_4$ $\bar{e}_5$ $\bar{e}_6$ … $\bar{e}_n$
Concepts in the knowledge base after individual application of appropriate KDF
Creation of (t,n)-threshold verification information
(t,n)-threshold Verification Steps
Stored (t,n)-threshold verification information
Recurring during every authentication attempt of the user
Identification
Authentication
Authentication scheme poses challenges to reach certainty threshold
"sunflower"
"water"
"coal"
"truck"
user responds to challenges
$e'_1 = \langle$"sunflower", TRUE$\rangle$
$e'_2 = \langle$"water", FALSE$\rangle$
$e'_3 = \langle$"coal", FALSE$\rangle$
$e'_4 = \langle$"truck", TRUE$\rangle$
Verification
$\bar{e}'_1 = KDF(e'_1)$ … $\bar{e}'_4 = KDF(e'_4)$
User responses after individual application of appropriate KDF
(t,n)-threshold verification
Authorisation
Termination

Figure 8.4: Overview the different phases of the authentication procedure when applying the (t, n)-threshold verification scheme to the ZeTA authentication scheme

used variant) matches the previously stored s. If the two hashes match, the user has entered the correct authorised subset and the authentication attempt is successful.

8.4 Discussion

This section discusses the three use cases for the (t, n)-threshold verification scheme presented in this chapter and the limitations that arise from these use cases.

8.4.1 Results

This chapter presented three use cases of the (t, n)-threshold verification scheme described in chapter 7. The first two use cases show how easily (t, n)-threshold verification can provide secure and efficient storage of the passwords for portfolio authentication schemes such as graphical recognition-based passwords and partial passwords. The third use case shows the versatility of the (t, n)-threshold verification scheme beyond portfolio authentication: instead of applying (t, n)-threshold verification directly to the secret, it is applied to the structure on which the ZeTA authentication scheme is based.

While the three use cases outlined in this chapter only cover knowledge-based authentication, the (t, n)-threshold verification scheme can be applied to far more general contexts. All devices and authentication schemes come with design spaces that are specific to the respective device and authentication scheme (e.g. [180]). Using the proposed (t, n)-threshold verification scheme, each device can offer multi-factor authentication tailored to its specific requirements and features (screen size, presence of keyboard or touchscreen, biometric sensors etc.) while still being able to verify all the information securely and efficiently. The (t, n)-threshold verification scheme is an important enabler in this regard. For example, it is a natural extension of the (t, n)-threshold verification scheme to support passwords that comprise elements of different types (i.e. originate from different types of authentication schemes). Such an extension requires only an adequate KDF to derive the respective value $\overline{e}_i$ from the user input. Therefore, it is conceivable to allow multi-factor passwords in systems in which not all factors have to be provided for each authentication attempt, e.g. one element of the password could be a text password, a second element a certificate on a smart card, a third element a biometric such as a fingerprint, another element a USB-token and so on, but only two of these are required at login. This can be beneficial in scenarios where not all factors are available at all times, where factors are prone to error, or when factors become unavailable (e.g. due to theft or failure).

One simple illustrative example for the combination of different authentication schemes by using the (t, n)-threshold verification scheme can be directly derived from the use cases outlined in this chapter. Imagine an application scenario, where authentication to the same account takes place on different devices, e.g. a laptop or desktop computer and a mobile device with only a touchscreen as input device (e.g. tablet or smartphone). While the laptop offers a keyboard for text password entry, the touchscreen device does not. Using on-screen keyboards has been shown to decrease the guessing resistance of password which have to be entered on them [143]. Therefore, a possible solution enabled by the (t, n)-threshold verification scheme would be to combine a text password with a graphical recognition-based password. During the enrolment, the user would have to choose a text password and would be assigned a graphical recognition-based password in the usual manner for each of the schemes. The (t, n)-threshold verification scheme would then be configured in such a way that either of these secrets are sufficient to login. During authentication the interface would either ask for the text password or the graphical password depending on the device the user is using at the time. Of course this scheme requires the user to memorise two secrets instead of just one and therefore serves only to illustrate the type of solutions (t, n)-threshold verification enables. The design of practical solutions of this kind remains future work.

8.4.2 Limitations

From the three use cases a limitation already discussed in section 7.5 becomes clearly visible: while the storage requirements of the (t, n)-threshold verification scheme are considerably lower than for naive approaches, they also represent a substantial uptake in comparison to non-portfolio schemes. The magnitude of this uptake profoundly depends on the respective authentication scheme. Among the use cases presented in this chapter, it becomes most pronounced in the third use case: for ZeTA the verification information is based on the entire knowledge base underlying the scheme. Storing this information for each user of a system might amount to overall sizeable storage requirements. In contrast, the uptake in the other two use cases seems negligible in most application scenarios. Consequently, storage must still be a consideration when evaluating the usage of the (t, n)-threshold verification scheme in conjunction with any authentication scheme.

8.5 Conclusion

This chapter presented three use cases of the (t, n)*-threshold verification scheme* from the knowledge-based authentication domain. While these three use cases already outline the wide range of possible applications of the (t, n)-threshold verification scheme, possible areas of application go beyond these presented use cases and can span multiple types of authentication, thereby enabling novel authentication schemes adapting to the users environment by utilising available authentication factors and compensating for factors that might be unavailable in certain situations (e.g. fingerprint recognition shortly after taking a shower or bathing).

Conclusion

9 Conclusion

This chapter highlights the most important aspects of this thesis in a brief summary of the results (section 9.1) and the main contributions (section 9.2). Thereafter, the directions of future research (section 9.3) outline the continuation of the research presented in this thesis. A few final remarks conclude this thesis (section 9.4).

9.1 Thesis Summary

The research presented in this thesis focuses on enabling secure and usable user authentication. To this end, its three parts addressed three distinct challenges.

9.1.1 Part I: Effective Password Security Awareness Materials

The first part of this thesis (chapter 3 & 4) addresses the challenge of how to create password security awareness materials which are correct, complete, understandable, and effective. To achieve this, a systematic process for the development of awareness materials is developed and applied to the password security context. The first step focused on the aggregation of relevant content, i.e. descriptions of relevant attacks and defences as well as interventions to clear up common misconceptions about password security identified in a systematic literature review. Thereafter, in the second step structured feedback of independent experts was used to ensure the material's correctness and completeness. Then, in the third step feedback of lay-users was used to ensure the material's visual appeal and understandability. Finally, in the fourth step an evaluation in the field with employees of three German SMEs ensured the material's effectiveness and provided additional feedback for improvements to create the final version of the material. From the results of the evaluation it became apparent that the awareness material was received very positively by the participating employees. Additionally, it significantly improved their ability to identify secure behaviour with respect to passwords and significantly decreased the prevalence of the identified misconceptions.

9.1.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

The second part of this thesis (chapter 5 & 6) addresses the development of shoulder-surfing resistant text password entry, when using constrained input devices such as gamepads in shared spaces. It first describes the requirements of text password entry using gamepads. Then, it presents an assessment along these requirements of (a) schemes currently deployed in the gamepad context and (b) shoulder-surfing resistant authentication schemes proposed in non-gamepad contexts. The assessment shows that none of the currently deployed and only three of the proposals in the literature fulfil all requirements. Thereafter, the shoulder-surfing resistance and the usability of three authentication schemes in the gamepad context were evaluated in two user studies. The results of this evaluation show that *on-screen keyboards* (the de facto standard in this context) are highly susceptible to shoulder-surfing attacks, but fare best in terms of usability. In contrast, the *Colorwheels* scheme (a novel shoulder-surfing resistant authentication scheme specifically geared towards the gamepad authentication context) seems to offer the most robust shoulder-surfing resistance, while still exhibiting more favourable usability results than the adapted *grid-based scheme* (a shoulder-surfing resistant scheme from the literature identified as adaptable during the assessment). In particular, the *Colorwheels* scheme shows the highest future usage intention, which indicates that users are willing to trade usability for increased shoulder-surfing resistance in the context of text password entry on gamepads and that this scheme's design direction is worth pursuing further in future work.

9.1.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

The third part of this thesis (chapter 7 & 8) addresses the challenge of providing a solution for secure and efficient storage of passwords in portfolio authentication schemes. In portfolio authentication the password is regarded as being composed of elements and every authorised subset of password elements is sufficient to authenticate a user (cf. section 2.5). Such schemes are used to counter capture attacks such as shoulder-surfing, but a verification scheme which allows secure and efficient verification of the input provided by the user had been missing until now. Therefore, the novel (t, n)-threshold verification scheme is presented. It allows the derivation of the same secret from all authorised subsets of the password elements through the use of secret sharing and key derivation functions. The security as well as efficiency properties of two variants of the scheme are evaluated against a naive implementation and against each other. These evaluations show the (t, n)-threshold verification scheme always exhibits more favourable properties than the naive approach. The application of the (t, n)-threshold verification scheme is illustrated based on three use cases. Two of the use cases illustrate the application of the (t, n)-threshold verification scheme in its originally intended application scenario, i.e. knowledge-based portfolio authentication. The third use case shows the applicability of the scheme beyond its originally intended application scenario.

9.2 Summary of Contributions

Several contributions to the field of usable and secure user authentication are presented throughout the three parts of this thesis:

9.2.1 Part I: Effective Password Security Awareness Materials

- The results of the systematic literature review give an overview of the misconceptions about password security which were identified in the literature. Overall, 23 misconceptions were identified, covering a wide range of aspects of password security. The misconceptions can be grouped into four categories: composition, handling, attacks, and miscellaneous.
- A systematic development process is of the essence when trying to create awareness materials which are truly correct, complete, understandable, and effective. The process presented in this thesis for the creation of information security awareness materials combines the state of the art described in the current scientific literature, the expertise of independent experts from academia and industry, feedback from lay-users, as well as a formal evaluation in three organisations in order to ensure all

four of these properties. Thereby, it enables the creation of awareness materials which are correct, complete, understandable, and effective.

- Using the process outlined before, a password security awareness material was created. Its evaluation showed that users exhibit a significantly improved ability to assess password-related behaviour, a significantly improved ability to assess the security of passwords, and a significantly decreased overall prevalence of misconceptions about password security.

9.2.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

- The investigation of authentication in the gamepad context identified six requirements across three categories (i.e. security, technical, and usability) that all authentication schemes intended to be used in this context need to fulfil. These requirements can inform the design of new authentication schemes in the gamepad context.
- Based on the identified requirements, an assessment of the authentication schemes currently deployed in the gamepad context as well as a representative set of shoulder-surfing resistant authentication schemes proposed in the literature was performed. From the results of this assessment it becomes apparent that none of the currently deployed authentication schemes and only four of the proposals from the literature fulfil all requirements.
- From the results of two studies – one online study and one lab study using similar methodologies – a baseline performance regarding the shoulder-surfing resistance and usability for the on-screen keyboard (the incumbent and de-facto standard of authentication in the gamepad context) was established. While being rated favourably in terms of usability, the on-screen keyboard proved highly susceptible to shoulder-surfing attacks.
- Using the same methodology employed in the evaluation of the on-screen keyboard, the shoulder-surfing resistance and usability of two alternative schemes was assessed:

  (a) The grid-based scheme by Kim et al. [123], which is a shoulder-surfing resistant authentication scheme proposed in the literature and identified as the most viable candidate adaptable for the gamepad context.

  (b) The novel Colorwheels scheme (section 5.5) which was specifically designed for the gamepad context.

  The assessment of the schemes showed that the Colorwheels scheme seems to exhibit a more robust shoulder-surfing resistance and future usage intention than both, the on-screen keyboard and the grid-based scheme, while also showing better usability properties than the grid-based scheme.

9.2.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

- To address the challenge of secure and efficient password storage in portfolio authentication schemes, the (t, n)-threshold verification scheme was proposed. Two variants of the verification scheme – one based on Blakley secret sharing and one based on Shamir secret sharing – were presented.

- The security and efficiency of the two variants of the (t, n)-threshold verification scheme were evaluated and compared to a naive approach and against each other. The results of this evaluation showed that the two (t, n)-threshold verification variants exhibit more favourable properties.
- To illustrate the application of the (t, n)-threshold verification scheme to different knowledge-based authentication schemes, three use cases were presented: graphical recognition-based passwords, partial passwords, and the ZeTA authentication scheme.
- To expand upon the three use cases, further potential applications of the (t, n)-threshold verification scheme beyond the scope of knowledge-based authentication are described. These further areas of application span far more general contexts including different types of authentication (i.e. knowledge-based, token-based, biometric). Thereby, the usage of (t, n)-threshold verification allows to develop authentication schemes which can adapt to the user's environment and provide authentication using factors available in that environment (e.g. specific biometric sensors) as well as fault tolerance in the face of failing authentication factors.

9.3 Future Research

Each of the three parts of this thesis raises and discusses individual open questions for future research. In the following the most important of these questions are described for each of the three parts.

9.3.1 Part I: Effective Password Security Awareness Materials

Regarding the creation of password security awareness materials, it is important to base the materials on a solid foundation, i.e. the content which is included in the material must be carefully selected and curated. One of the initial sources for the content of the awareness material whose creation and evaluation is described in chapter 4 are the misconceptions identified in the systematic literature review in chapter 3. From the evaluation presented in section 4.4, it becomes apparent that most of these misconceptions are prevalent in the sample of SME employees in the presented study. Yet, it is important to acknowledge that the relevant set of misconceptions will change over time as evidenced by the material's effectiveness in reducing the prevalence of the misconceptions. The same holds for the other content of the password security awareness material as well. Therefore, the material should be updated regularly to include the relevant contents. With respect to the misconceptions in particular, future research should focus on identifying the patterns behind the changes to the individual misconceptions induced by the material over time, in order to identify the most effective designs for the interventions.

Additionally, the investigation of interactions between the prevalence of different misconceptions might prove to be a valuable direction of future research: Identifying clusters of misconceptions which are usually prevalent in the same user would allow clearing up these misconceptions in the awareness material together and therefore might allow to form additional synergies in the material's content.

Last but not least, certain trade-offs had to be made in the design of the evaluation of the awareness material. Conducting the study in the real work environment of the participating employees left them unsupervised during the entire study procedure. Therefore, the results should be validated in future research using a more regulated setting (e.g. a lab study) and different participant samples.

9.3.2 Part II: Shoulder-surfing Resistant Text Password Entry on Gamepads

Part II includes two studies, one online study and one lab study. Since the results of these two studies differ with respect to one hypothesis, it is important to note the three aspects in which they differ. Firstly, the participants in the lab study had a deeper understanding of the scheme they had to shoulder-surf, due to the fact that they actually operated the scheme during the usability assessment before the shoulder-surfing attack. Secondly, the participants in the lab study wrote their guesses on one sheet of paper. Therefore, when making a guess, they saw their previous guesses. Thirdly, the participants in the lab study asked whether they could stop guessing the password during the evaluation of the shoulder-surfing resistance and due to ethical reasons they were of course allowed to stop. All three of these aspects warrant further investigations. The impact of each of these differences should be investigated individually in future research, in order to identify which influence they have on the tested hypotheses.

Additionally, the studies suggest that users are willing to trade usability for increased shoulder-surfing resistance in the context of text password entry on gamepads. Gaining a deeper understanding of the factors involved in the users' willingness to accept this trade-off might prove valuable for future authentication scheme designs. Also, the question arises whether this effect translates to other contexts in public, shared, and private spaces.

9.3.3 Part III: Secure and Efficient Storage of Passwords in Portfolio Authentication

For part III two aspects stand out as particularly valuable directions for future work. Firstly, a more in-depth cryptanalysis using lattice-based algorithms or a modelling using integer linear programming might uncover situations in which a re-adjustment of the choices for the parameters n, t, and p for either of the two variants of the (t, n)-threshold verification scheme becomes necessary.

Secondly, the (t, n)-threshold verification scheme can serve as an enabler for authentication schemes that adjust to the environment (e.g. available inputs on the user's device) in which the authentication takes place. The use cases presented in chapter 8, already outline how to apply the (t, n)-threshold verification scheme to individual authentication schemes. As outlined in section 8.4, one natural extension of the presented use cases would be the combination of a graphical recognition-based scheme and a traditional text password. Such a combination would allow to e.g. use the text password on devices that have a hardware keyboard and the graphical scheme on devices with a touchscreen. However, it is unclear how users would react to such hybrid schemes and their investigation might prove to be a valuable direction of future work.

9.4 Final Remarks

Despite several prophecies to the contrary, text passwords (and their derivatives such as PINs) are the dominant type of authentication today and are unlikely to be replaced any time soon. In contrast, the ever growing number of devices and accounts users will use on these devices will only further aggravate the problems underlying the challenges addressed by the research in this thesis. The ever expanding range of attacks employed by cyber criminals necessitates the availability of correct, complete, understandable, and effective awareness materials explaining the relevant attacks and respective defences to users. The advent of new form factors of devices and the usage of text passwords on constrained input devices such as gamepads require new authentication schemes specifically adapted to these contexts. The increased use of mobile devices in public or shared spaces increasingly shapes a new threat landscape, where capture attacks such as shoulder-surfing are more relevant than ever.

By addressing three distinct challenges encompassing the three entities involved in user authentication (i.e. the user, the authentication scheme, and the verifier), this thesis demonstrates the breadth of the field of usable and secure user authentication ranging from awareness materials, to the development and assessment of authentication schemes, to applying cryptography to craft secure password storage solutions. Therefore, I argue that the research processes, results, and insights described in this thesis represent important and meaningful contributions to the state of the art in the research on usable and secure user authentication, offering benefits for users, organisations, and researchers alike.

Appendix

A Appendices to Part I

A.1 Examples for Attack Automation

Table A.1: Examples of how the attacks from the framework by Bonneau et al. [26] could be automated.

Attack	**Automation**
Attacks targeted directly at the user and the interaction with their devices	
Phishing	Using markov models and neural networks allows the automated creation of phishing messages [113].
Theft of an Insecurely Stored Physical of the Password	While no automation for stealing objects using drones is described in the research literature, it is easily conceivable that through advances in object recognition and autonomous navigation drones might be able to identify and steal authenticators such as tokens or smart cards or even physical copies of passwords.
Shoulder-surfing	The automated analysis of password entry is easily possible [135]. Such techniques could be combined with strategically placed high resolution surveillance cameras to capture usernames and password while using image recognition to identify the context the password was entered in to fully automate this type of attacks.
Attacks targeting the user's device as well as remote services	
Compromising the User's Device	Automated attacks based on internal observation facilitated through compromising a user's device can be mounted using any self-replicating malware such as viruses or worms [201].
Targeted Guessing	While this attack is based on intimate knowledge of the user, automated attacks combining information available from different social networks have been shown to be feasible [21].
Untargeted Guessing	Among the most important cases of untargeted guessing are online attacks using remote servers as oracle. Such attacks can be automated with software such as Hydra-THC[2] which allows attacks over a wide variety of protocols commonly used by Internet services.
Guessing after System Break-In	Software to mount unthrottled automated guessing attacks after passwords have been stolen from servers is widely available. Among the most popular software are Hashcat[3] and John the Ripper[4].
Theft of an Digital Copy of the Password	Automated theft of password vaults through malware has been well documented [87].
Attacks targeted at the communication between the user's devices and remote services	
Compromising the network traffic between the users' devices and the remote services	The use of unencrypted communication can be especially dangerous and facilitate easily automated interception of passwords or secondary authentication information (such as session cookies) [41]. However, even encrypted traffic can be targeted using man-in-the-middle attacks [102,104]. For example the compromise of the Dutch CA DigiNotar lead to illicitly issued certificates for man-in-the-middle attacks and the CA was therefore subsequently removed as root CA from all major browsers [3].

[2] `https://github.com/vanhauser-thc/thc-hydra` (visited 2019-07-06)

[3] `https://hashcat.net` (visited 2019-07-06)

[4] `https://https://www.openwall.com/john/` (visited 2019-07-06)

A.2 Intervention Texts

Table A.2: The wording of the intervention texts after the round of expert feedback. Note that the wording represents a translation into English from the German original versions created for the study. Creating German texts was necessary to allow the study presented in chapter 4.

ID	Wording of the Intervention after Expert Feedback
M1	Using specialized software, attackers try to mimick human behaviour when guessing passwords. Thereby, they use a long list of words (from dictionaries, but also passwords from past breaches) and apply different common modifications to these words to generate additional words they will also use as guesses for the password: • Appending or prepending numbers and symbols (e.g. adding an "!" to the end is a popular choice) • Substituting letters with numbers (e.g. E → 3) or with symbols (e.g. a → @) • Substituting lowercase letters with uppercase letters (in particular at the beginning of words)
M2	
M3	
M4	
M5	Attackers will adjust their list of passwords to try according to the circumstance. In particular, they can use words from many languages to try and guess your password.
M6	Unfortunately, it happens time and again that even large web services handle the passwords of their users carelessly which then leak to attackers. Therefore, even passwords which are very hard to guess should not be reused. If a password is leaked by a service in the clear, it does not matter how hard it is to guess. This also holds for passwords you enter frequently. Do not reuse them. The more often you reuse a password at different services, the bigger is the chance of it getting into the hands of an attacker.
M7	
M8	If you store a written down password in a safe and secure location, making such notes can actually be beneficial (e.g. after changing a password until you have memorized it). However, if you do not need a note of a password anymore, you should dispose it (e.g. burn it). If you keep your written down passwords securely stored, having such notes is more secure than reusing passwords: Guessing a unique password is more difficult to guess than a password that might have leaked from another service which was affected by a leak, even if that password is a variation of a different password.
M9	Written down passwords must always be stored in a secure location (i.e. a location that can only be accessed by yourself). If you use a password manager you should in most cases set a strong master password. Only in cases where all of the four criteria below are met, it is not mandatory to set a password: • You are the only user of your devices • The hard drive of your device is encrypted • You do not synchronize your passwords across your devices • You always lock your devices, when you are not using them Otherwise you have to set a master password which can withstand guessing after a system break-in.
M10	Scientists have found that changing passwords you need to remember proactively (i.e. without occurrence of an incident) is unhelpful in protecting your accounts. The additional effort required of the users is unproportionally larger than the achieved security benefits. Even governmental bodies, such as the US NIST or the British NCSC are already adapting their recommendations. They recommend to change passwords only when the old one has fallen into the hands of an attacker instead of changing it proactively. These governmental bodies believe that the web services have to implement a rigorous monitoring of their own systems and that they should use so-called lock-out mechanisms (e.g. limiting the number of possible login attempts). Changing passwords that have to be remembered is therefore obsolete advice which should not be followed anymore.
M11	Internet browsers often have an integrated password manager, which allows saving passwords entered on websites. Saving passwords in a browser is the same as saving them in a dedicated password manager. If the hard drive of your device is not encrypted and your passwords are saved in a password manager, which is not protected by a master password (no matter if in a browser or as dedicated program), attackers can easily copy your passwords off your hard drive, if they have physical access to it or your device is infected with malware.
M12	Using walks or patterns on your keyboard as password (e.g. "1QAY2WSX") is no good practice to generate secure passwords. As a matter of fact, such patterns will be present in the dictionary of every attacker trying to guess passwords.
M13	Attackers can easily use specialized software to guess all combinations of days, months, and years. Therefore, using your dog's birthday or the birthday of your favorite actor instead of your own will not render your password harder to guess for a professional attacker. This holds in all situations, where attackers can make large numbers of guesses (e.g. when a web service does not limit the number of possible login attempts, before the account is locked).
M14	Attackers can easily automate their attacks using readily available specialized software. Using such software, attackers can easily e.g. test many different passwords or snoop on unencrypted network traffic.
M15	Some people think that only cyber-criminals from the other side of the planet will typically attack them. Others believe that only those close to them will try to get access to their devices. However, attackers can come from both of these groups.
M16	
M17	Even if you use your email account only to send messages to others, it is still a valuable target for attackers. Exploiting the possibility to reset passwords of other accounts is particularly relevant. Your email account holds email from all the services you use. Therefore, attackers can easily look up all the web services you use and reset their passwords to access them.

Continued on next page

Table A.2 – ***continued from previous page***

ID	Wording of the Intervention after Expert Feedback
M18	The SIM-PIN is the PIN you have to enter to unlock the telephony functions of your mobile phone. This PIN is not sufficient to protect the data which is stored on a smartphone. An attacker with physical access to the phone can easily remove the SIM from the phone to bypass this check. Moreover, many phones simply allow bypassing the entry of the SIM-PIN, resulting in a phone that is not able to make calls, but is unlocked to access all data on it.
M19	You should set a password lock and use it whenever you leave a device unattended – even if a co-worker or friend is in the vicinity (e.g. same office). If during your absence these people leave too (e.g. are called away, make telephone calls, or go to the toilet) your device is completely unprotected.
M20	You are responsible for the security of your devices. Even at work when there is a dedicated IT department, it is your job to keep the device secure.
M21	The value of your accounts is not influenced by how regularly you use it. Instead, the only determining factors are the data which can be accessed in the account and which actions can be performed in the account.
M22	Money is not the only valuable associated with your bank account. When accessing your online banking account, attackers can easily see where you are shopping, when you book your holidays and where you are paying abroad, your address, your phone number . . . All they need is your password, no TANs are needed. *(Note for readers: TANs are transaction numbers. All German banks require these to authorize transactions. Since all participants in our study were German, all knew this term and its semantics in the context of online banking.)*
M23	Some people believe that they are not important enough to even be targeted by attackers. This is a misjudgment which can have potentially severe consequences. Many attacks on the Internet (e.g. guessing passwords of users at a web service or sending phishing emails) can be easily automated and are then performed in an untargeted fashion. Do not fall for the illusion of not being affected by attacks. Learn how to defend yourself, for when attackers just try to get access to as many accounts as possible. Otherwise you will fall victim to an attacker.

A.3 Scenario Texts

Table A.3: The scenarios used in the study. Originally developed and used in German they are presented here translated in English. For each attack there are two scenarios, one representing secure and one representing insecure behaviour.

Attack		#	Scenario
Fraudulent messages	Secure	1	Mr. Schmidt works together with his colleague Mr. Müller on the same project. Mr. Schmidt is the vacation substitution for his colleague. He receives an email in which his colleague asks him to send the project plan to his private email address, because he wants to work on it in his rainy vacation. Mr. Schmidt does not send the information to the private email address.
	Insecure	2	Mr. Schmidt's boss is on a business trip to visit a client. Mr. Schmidt receives an email in which his boss informs him that a person from the help desk of the client will contact Mr. Schmidt to get access to the web-interface of the project management software. Shortly after, Mr. Schmidt's phone rings: it is the person from the help desk. Since he received the announcement of the call from the email-address of his boss, Mr. Schmidt gives the person from the help desk the required password.
Theft of an insecurely stored physical copy of the password	Secure	3	Mr. Schmidt finds it difficult to remember his passwords. Therefore, he keeps a note of his private passwords at home in a locked drawer of his desk, which only he can open.
	Insecure	4	Mr. Schmidt has to change the password for one system in the company every 90 days. He already had to call the help desk of his company multiple times to have them reset a password he could not remember after a mandatory change. When changing the password for the next time, he makes a note of it and stores the note under his mouse pad on his desk.
Shoulder-surfing	Secure	5	Mr. Schmidt sits in the train on his way to a client. The train is fully booked, the seat next to him taken. Mr. Schmidt checks emails using his smartphone. Due to an urgent request from his boss, he has to access the web-interface of the project management software used in this company to list a cost report. He notices that the person in the seat next to him tries to look at the screen of his smartphone inconspicuously. Therefore, he leaves his seat and moves to an area in the train where he is undisturbed, so that no one can spy on the sensitive data he is accessing.
	Insecure	6	Mr. Schmidt is sitting in a café and waits on his colleague to have lunch together. Since his colleague sent him a text message saying that he will be 30 minutes late, Mr. Schmidt wants to use the time to work on his laptop. While he is working a couple approaches and asks whether they can join him at the table. Since all other tables in the café are fully occupied, Mr. Schmidt agrees. One of them sits down on the opposing side of the table, one sits down next to Mr. Schmidt who continues his work and logs in multiple times to the web-interface of the project management software of his company.
Compromising the users' devices	Secure	7	Mr. Schmidt sits in his office. He is printing presentation slides for a meeting. The printer is located at the other end of the corridor. Before Mr. Schmidt leaves his desk to fetch the print-out, he locks his laptop.
	Insecure	8	Mr. Schmidt has to share a file with this colleague Mr. Müller. The file is too large to attach it to an email. Since he has no USB stick at hand, he uses the one he found last week in the parking lot of his company.
Eavesdropping on unencrypted communication	Secure	9	Mr. Schmidt is on his way to a client. Unfortunately, the train is delayed. Therefore, he sits down in a café at the train station. There he uses an open wifi. He uses his laptop as usual, but pays attention that he visits all websites using an encrypted connection.
	Insecure	10	Mr. Schmidt is on a business trip visiting a client in a different city. There he stays in a hotel and uses its charged premium unencrypted wifi to work in his room. To login to the wifi, he has to enter a user name and a password.
Eavesdropping on encrypted communication	Secure	11	Mr. Schmidt has to access the web-interface of a client's system. He receives a warning that no encrypted connection is possible although this has worked in the past. Therefore, Mr. Schmidt calls the client using a phone number known to him, describing the problem. He does not access the web-interface until the problem is solved.
	Insecure	12	Mr. Schmidt is at a client in a different city to prepare a new project. He has to stay several nights and books a room in a hotel. Once he is in his room, he tries to access the web-interface of the project management software of his company. He receives a warning stating there is problem with the security of the connection, although the connection is encrypted. The problem does not occur with the web-interface of his email account. He infers that the web-interface of the project management software is misconfigured and enters his credentials.

Continued on next page

Table A.3 – ***continued from previous page***

Attack		**#**	**Scenario**
Targeted guessing	Secure	13	Mr. Schmidt takes his private smartphone to work (but does not use it for business purposes). He does not want that friends or colleagues can access the phone by guessing his PIN. Since he shares the phone with his wife in their free-time, she should be able to easily remember the PIN. Therefore Mr. Schmidt uses as PIN the birthday of the family dog, which is only known to him and his wife.
	Insecure	14	Mr. Schmidt finds it difficult to remember passwords. Therefore, he uses as password for his laptop at work *Alexander1997*, the first name and the year of birth of his child.
Untargeted guessing	Secure	15	Mr. Schmidt has to perform small design tasks. For this purpose, he has to open an account with Adobe to purchase and download software such as Photoshop and InDesign. For the user account, he chooses a long password (substantially longer than 8 characters), which is neither in the lists of frequently chosen passwords nor derived from the company name Adobe.
	Insecure	16	Mr. Schmidt has problems remembering all the passwords he needs for his job and privately. Therefore, he uses walks on the keyboard, such as *1q2w3e4r%*, to create secure passwords.
Guessing after system break-in	Secure	17	Mr. Schmidt is frequently on business trips and once he had his laptop almost stolen at the airport. Therefore, he chooses to encrypt the hard drive using a password with more than 20 characters which he creates concatenating multiple words to one another.
	Insecure	18	The company in which Mr. Schmidt is employed uses an external web service to store important client data. Mr. Schmidt learns that this web service was the target of a successful hacker attack and that password data was stolen. His password is *Al3xand3r!*, derived from the name of his son. Since the password is longer than 8 characters and contains multiple numbers and a symbol he does not change it.
Theft of a digital copy of the password	Secure	19	Since Mr. Schmidt has problems remembering the many passwords he needs for his job, he asks the IT-department whether they can install a password manager on his work laptop. Since he wants to synchronise the passwords to his business smartphone, he chooses a master password with more than 20 characters, which he creates by concatenating multiple words.
	Insecure	20	Mr. Schmidt has to use many passwords in his daily job to log on to all the different systems he needs to access. Since he also works on his business smartphone, Mr. Schmidt saves all the passwords in a Word document and synchronises this document through a public third-party cloud storage provider between his laptop and his smartphone.
Exploiting a weak reset-mechanism	Secure	21	Mr. Schmidt uses different external web services, as is usual in his company. For one of the web services, the password can be reset using personal security questions. Instead of answering the questions truthfully, Mr. Schmidt chooses a random character sequence as answers, writes this sequence down, and stores it where only he can access it.
	Insecure	22	Mr. Schmidt uses different web services in his private life. For one of the web services the password can be reset using a link in an email sent to him. As password for the respective email account he chooses *@lex@nder1997* (derived from the first name and year of birth of his son), since it is more than 8 characters long and contains multiple special characters.

A.4 Misconception Items

Table A.4: The items for each of the misconceptions, translated from the German originals used in the study. The IDs relate to the misconceptions as presented in chapter 3. The items marked with an '*' are inversely phrased.

ID	Item of the Intervention
M1	Adding numbers makes passwords automatically more difficult to guess.
M2	Adding symbols makes passwords automatically more difficult to guess.
M3	Adding uppercase letters makes passwords automatically more difficult to guess.
M4	Replacing lowercase letters in the password with numbers, symbols, or uppercase letters makes the password automatically more difficult to guess.
M5	A word from another language than your own mother tongue is a secure password.
M6	Reusing passwords is OK for secure passwords, but should be avoided for weak passwords.
M7	It is OK to reuse passwords from user accounts that you log in frequently for different user accounts.
M8*	Security-wise, it is better to write passwords down and keep them in a secure location than to reuse passwords for different user accounts.
M9*	Notes of passwords must always be stored in a secure location.
M10	Passwords should be changed frequently.
M11*	Storing passwords in the browser is the same as storing passwords in a password manager.
M12*	Walks or patterns on the keyboard (e.g., 1qay2wsx) represent insecure passwords.
M13	Using dates of birth that are not your own is a good way to choose secure passwords.
M14*	Attacks on user accounts can be automated.
M15	All attackers are strangers from the other end of the world.
M16	All attackers are only people you know.
M17*	Email accounts have particularly high security requirements.
M18	A SIM PIN is sufficient to protect data on a smartphone.
M19*	It is necessary to lock your devices (PC, laptop, smartphone, etc.), even if you leave them unattended only for a short time.
M20	User accounts in an organization have lower security requirements than private user accounts, because the IT department watches over them.
M21	The security requirements of a user account depend on how often it is used.
M22	Only bank accounts with high account balance are a rewarding target for attackers.
M23	One can be too unimportant to be attacked.

B Appendices to Part II

B.1 Overview of the Assessment of Authentication Schemes

Table B.1: Assessment of whether the existing shoulder-surfing resistant authentication schemes fulfil requirements R2-R6 as defined in section 5.1. For the sake of completeness this table also includes the Colorwheels scheme which is specifically tailored to the gamepad context and described in more detail in section 5.5.

Type	Proposal	R2	R3	R4	R5	R6
Covert Channels	Secure Haptic Keypad [19]	yes	yes	yes	somewhat, haptic feedback not for individual buttons	yes
	Glass Unlock [229]	yes	yes	yes	no, requires external private display	yes
	force-PIN [127]	no, requires more force sensitive controls	yes	somewhat, requires extension of character grid	yes	yes
	Undercover [176]	no, requires concealed placement of trackball	yes	somewhat, requires extension of the input grid	yes	yes
Obfuscation of Input	Tetrad [169]	yes	yes	no, graphical	yes	yes
	Draw-A-Secret variants [233]	yes	yes	no, graphical	yes	yes
	Convex Hull Click [226]	yes	somewhat, designed for mouse input	no, graphical	yes	yes
	S3PAS [237]	yes	somewhat, designed for mouse input	yes	yes	yes
	Grid-based scheme [123]	yes	yes	yes	yes	yes
	PairPasswordChar [166]	yes	yes	yes	yes	yes
	Colorwheels (section 5.5)	yes	yes	yes	yes	yes
Indirect Input	Cognitive Trapdoor [173]	yes	yes	yes	yes	yes
	SwiPIN [218]	no, requires too many directional controls when scaled up from PIN to text password	yes	yes	yes	yes
	Xside [57]	no, requires touch interface on back of device	yes	somewhat, requires extension of character grid	yes	yes
	MagiSign [175]	yes	yes	yes	no, requires magnetic sensor	yes
Behavioural biometrics [56]		yes	yes	no, needs additional verification of biometric information	yes	yes
Portfolio Authentication	Graphical Portfolio Authentication [166]	yes	yes	no, graphical	yes	yes
	Textual Portfolio Authentication (section 5.5)	yes	yes	yes	yes	yes

References

List of Figures

List of Tables

Bibliography

[1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *International Conference on Human Factors in Computing Systems*, pages 3751–3763, 2017.

[2] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[3] Heather Adkins. An update on attempted man-in-the-middle attacks, August 2011. `https://security.googleblog.com/2011/08/update-on-attempted-man-in-middle.html` (accessed: 2019-07-07).

[4] Maarten Aertsen, Maciej Korczyński, Giovane C M Moura, Samaneh Tajalizadehkhoob, and Jan van den Berg. No domain left behind: is Let's Encrypt democratizing encryption? In *Applied Networking Research Workshop*, pages 48–54, 2017.

[5] Parag Agrawal. Keeping your account secure, May 2018. `https://blog.twitter.com/official/en_us/topics/company/2018/keeping-your-account-secure.html` (accessed: 2019-07-11).

[6] Anonymous. Graphical Passwords. *International Technology Disclosures*, 10(1), 1992.

[7] Anti-Phishing Working Group. Phishing Activity Trends Report - 1st Quarter 2019. Technical report, 2019.

[8] Apple Inc. Face ID Security. Technical report, November 2017.

[9] Apple Inc. iOS Security. Technical report, 2019.

[10] Lynn Y Arnaut and Joel S Greenstein. Human Factors Considerations in the Design and Selection of Computer Input Devices. In *Input Devices*, pages 71–122. 1988.

[11] David Aspinall and Mike Just. "Give Me Letters 2, 3 and 6!": Partial Password Implementations and Attacks. In *International Conference on Financial Cryptography and Data Security*, pages 126–143, 2013.

[12] Adam J Aviv, John T Davin, Flynn Wolf, and Ravi Kuber. Towards Baselines for Shoulder Surfing on Mobile Authentication. In *Annual Computer Security Applications Conference*, pages 486–498, 2017.

[13] Adam J Aviv, Flynn Wolf, and Ravi Kuber. Comparing Video Based Shoulder Surfing with Live Simulation. In *Annual Computer Security Applications Conference*, pages 453–466, 2018.

[14] Mohammed Awad, Zakaria Al-Qudah, Sahar Idwan, and Abdul Halim Jallad. Password security: Password behavior analysis at a small university. In *International Conference on Electronic Devices, Systems and Applications*, 2016.

[15] Maria Bada and Angela Sasse. Cyber Security Awareness Campaigns - Why do they fail to change behaviour? Technical report, 2014.

[16] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[17] Elaine Barker, William Barker, William Burr, William Polk, and Miles Smid. NIST Special Publication 800-57: Recommendation for Key Management Part 1: General (Revision 3). Technical Report 800-57, 2012.

[18] Achim Berg and Michael Niemeier. Wirtschaftsschutz in der digitalen Welt. Technical report, 2019.

[19] Andrea Bianchi, Ian Oakley, and Dong Soo Kwon. The secure haptic keypad: a tactile password system. In *Conference on Human Factors in Computing Systems*, pages 1089–1092, 2010.

[20] Robert Biddle, Sonia Chiasson, and P C van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 44(4), 2012.

[21] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *International Conference on World Wide Web*, pages 551–560, 2009.

[22] Matt Bishop. *Computer Security: Art and Science*. Addison-Wesley, 1st edition, 2002.

[23] George R Blakley. Safeguarding cryptographic keys. *National Computer Conference*, 48:313–317, 1979.

[24] Christian Bock. Fujitsu and Microsoft focused on advancing security in the modern workplace, 2018. `https://blogs.windows.com/business/2018/02/08/fujitsu-microsoft-focused-advancing-security-modern-workplace/` (accessed: 2018-06-15).

[25] Joseph Bonneau, Elie Bursztein, Ilan Caron, Rob Jackson, and Mike Williamson. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In *International Conference on World Wide Web*, pages 141–150, 2015.

[26] Joseph Bonneau, Cormac Herley, Paul C van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *IEEE Symposium on Security and Privacy*, pages 553–567, 2012.

[27] Joseph Bonneau, Mike Just, and Greg Matthews. What's in a Name? In *International Conference on Financial Cryptography and Data Security*, pages 98–113, 2010.

[28] Joseph Bonneau and Sören Preibusch. The password thicket: technical and market failures in human authentication on the web. In *Workshop on the Economics of Information Security*, 2010.

[29] Nathan Bos, Darren Gergle, Judith S Olson, and Gary M Olson. Being there versus seeing there: trust via video. In *CHI Conference Extended Abstracts*, pages 291–292, 2001.

[30] Jeffrey E Boyd and James J Little. Biometric Gait Recognition. In *Advanced Studies in Biometrics*, pages 19–42. 2005.

[31] lker Nadi Bozkurt, Kamer Kaya, Ali Aydin Selçuk, and Ahmet M Güloglu. Threshold Cryptography Based on Blakley Secret Sharing. *Information Sciences*, 2008.

[32] Sacha Brostoff and M Angela Sasse. Are Passfaces more usable than passwords? A field trial investigation. In *People and Computers XIV - Usability or Else!*, pages 405–424. 2000.

[33] Sacha Brostoff and M Angela Sasse. Safe and sound: a safety-critical approach to security. In *New Security Paradigms Workshop*, pages 41–50, 2001.

[34] Alan S Brown, Elisabeth Bracken, Sandy Zoccoli, and King Douglas. Generating and remembering passwords. *Applied Cognitive Psychology*, 18(6):641–651, 2004.

[35] Burcu Bulgurcu, Hasan Cavusoglu, and Izak Benbasat. Information Security Policy Compliance: An Empirical Study of Rationality-based Beliefs and Information Security awareness. *MIS quarterly*, 2010.

[36] Andreas Bulling, Florian Alt, and Albrecht Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Conference on Human Factors in Computing Systems*, pages 3011–3020, 2012.

[37] Xavier Bultel, Jannik Dreier, Matthieu Giraud, Marie Izaute, Timothée Kheyrkhah, Pascal Lafourcade, Dounia Lakhzoum, Vincent Marlin, and Ladislav Motá. Security analysis and psychological study of authentication methods with PIN codes. In *International Conference on Research Challenges in Information Science*, 2018.

[38] Bundesamt für Sicherheit in der Informationstechnik. IT-Grundschutz-Kompendium. Technical report, 2020.

[39] Mark Burnett. Today I Am Releasing Ten Million Passwords , February 2015. `https://xato.net/today-i-am-releasing-ten-million-passwords-b6278bbe7495` (accessed: 2017-10-17).

[40] William E Burr, Donna F Dodson, Elaine M Newton, Ray A Perlner, W Timothy Polk, Sarbari Gupta, and Emad A Nabbus. NIST Special Publication 800-63-2: Electronic Authentication Guideline. Technical Report 800-63-2, 2013.

[41] Eric Butler. Firesheep, October 2010. `https://codebutler.com/2010/10/24/firesheep/` (accessed: 2019-07-14).

[42] Jack Callahan, Don Hopkins, Maek Weiser, and Ben Shneiderman. An empirical comparison of pie vs. linear menus. In *International Conference on Human Factors in Computing Systems*, pages 95–100, 1988.

[43] Pedro Canahuati. Keeping Passwords Secure, March 2019. `https://newsroom.fb.com/news/2019/03/keeping-passwords-secure/` (accessed: 2019-07-11).

[44] Sonia Chiasson and Paul C van Oorschot. Quantifying the security advantage of password expiration policies. *Designs, Codes and Cryptography*, 77(2-3):401–408, 2015.

[45] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. Graphical password authentication using cued click points. In *European Symposium on Research in Computer Security*, pages 359–374, 2007.

[46] Yee-Yin Choong, Mary Theofanos, Karen Renaud, and Suzanne Prior. Case Study – Exploring Children's Password Knowledge and Practices. In *Workshop on Usable Security*, 2019.

[47] Ton J. Cleophas. *Carryover Effects in Clinical Research*, pages 25–36. 1999.

[48] Graham Cluley. Prince William photos accidentally reveal RAF password, 2012. `https://nakedsecurity.sophos.com/2012/11/21/prince-william-photos-password/` (accessed: 2019-07-14).

[49] Jacob Cohen. Statistical power analysis for the behavioral sciences (2nd ed.). *Lawrence Erlbaum Associates*, 1988.

[50] Jose Cordova, Virginia Eaton, Tyler Greer, and Lon Smith. A comparison of CS majors and non-CS majors attitudes and practices regarding password strength. *Journal of Computing Sciences in Colleges*, 33(4):69–75, 2018.

[51] Lynne Coventry. Usable Biometrics. In *Security and Usability*, pages 175–197. O'Reilly Media, 2005.

[52] Sanchari Das, Gianpaolo Russo, Andrew C Dingman, Jayati Dev, Olivia Kenny, and L Jean Camp. A qualitative study on usability and acceptability of Yubico security key. In *Workshop on Socio-Technical Aspects in Security and Trust*, pages 28–39, 2018.

[53] Darren Davis, Fabian Monrose, and Michael K Reiter. On user choice in graphical password schemes. In *USENIX Security Symposium*, 2004.

[54] Antonella De Angeli, Mike Coutts, Lynne Coventry, Graham I Johnson, David Cameron, and Martin H Fischer. VIP: a visual approach to user authentication. In *Working Conference on Advanced Visual Interfaces*, pages 316–323, 2002.

[55] Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies*, 63(1):128–152, 2005.

[56] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Conference on Human Factors in Computing Systems*, pages 987–996, 2012.

[57] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. Now you see me, now you don't - protecting smartphone authentication from shoulder surfers. In *Conference on Human Factors in Computing Systems*, pages 2937–2946, 2014.

[58] Alexander De Luca, Marc Langheinrich, and Heinrich Hussmann. Towards Understanding ATM Security: A Field Study of Real World ATM Use. In *Symposium on Usable Privacy and Security*, 2010.

[59] Alexander De Luca, Emanuel von Zezschwitz, Ngo Dieu Huong Nguyen, Max-Emanuel Maurer, Elisa Rubegni, Marcello Paolo Scipioni, and Marc Langheinrich. Back-of-device authentication on smartphones. In *International Conference on Human Factors in Computing Systems*, pages 2389–2398, 2013.

[60] Rachna Dhamija and Adrian Perrig. Deja vu: A user study using images for authentication. In *USENIX Security Symposium*, pages 45–58, 2000.

[61] Luke Dormehl. Kid unlocks mom's iPhone X with Face ID, 2017. `https://www.cultofmac.com/513995/kid-unlocks-faceid-mom/` (accessed: 2019-09-08).

[62] Paul Dunphy, Andreas P Heiner, and N Asokan. A closer look at recognition-based graphical passwords on mobile devices. In *Symposium on Usable Privacy and Security*, 2010.

[63] D Eastlake, J Schiller, and S Crocker. Request for Comments 4086 - Randomness Requirements for Security. Technical report, 2005.

[64] Claudia Eckert. *IT-Sicherheit.* Oldenbourg Verlag, 2005.

[65] EMC Corporation. RSA SecurID Hardware Tokens, 2015. `https://www.rsa.com/content/dam/en/data-sheet/rsa-securid-hardware-tokens.pdf` (accessed: 2019-01-30).

[66] Mete Eminağaoğlu, Erdem Uçar, and Şaban Eren. The positive outcomes of information security awareness training in companies – A case study. *Information Security Technical Report*, 14(4):223–229, 2009.

[67] ENISA. Information security awareness in financial organisations - Guidelines and case studies. Technical report, 2009.

[68] Entertainment Software Association. Essential Facts About the Computer and Video Game Industry. Technical report, 2017.

[69] Brian W Epps. A Comparison of Cursor Control Devices on a Graphics Editing Task. In *Human Factors and Ergonomics Society Annual Meeting*, pages 442–446, 1987.

[70] Michael Fagan, Yusuf Albayram, Mohammad Maifi Hasan Khan, and Ross Buck. An investigation into users' considerations towards using password managers. *Human-centric Computing and Information Sciences*, 7(1), 2017.

[71] Dinei Florêncio and Cormac Herley. A large-scale study of web password habits. In *International Conference on the World Wide Web*, pages 657–666, 2007.

[72] Dinei Florêncio and Cormac Herley. Where do security policies come from? In *Symposium on Usable Privacy and Security*, 2010.

[73] Dinei Florêncio, Cormac Herley, and Paul C van Oorschot. An Administrator's Guide to Internet Password Research. In *Large Installation System Administration Conference*, pages 35–52, 2014.

[74] Dinei Florêncio, Cormac Herley, and Paul C van Oorschot. Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts. In *USENIX Security Symposium*, pages 575–590, 2014.

[75] Alain Forget, Sonia Chiasson, and Robert Biddle. Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords. In *Conference on Human Factors in Computing Systems*, pages 1107–1110, 2010.

[76] Suzanne Frey. Notifying administrators about unhashed password storage, May 2019. `https://cloud.google.com/blog/products/g-suite/notifying-administrators-about-unhashed-password-storage` (accessed: 2019-07-11).

[77] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. Privacy and Security Threat Models and Mitigation Strategies of Older Adults. In *Symposium on Usable Privacy and Security*, pages 21–40, 2019.

[78] Evan Fuller, Jeffrey M Rabin, and Guershon Harel. Intellectual Need and Problem-Free Activity in the Mathematics Classroom. *International Journal for Studies in Mathematics Education*, 4(1):80–114, 2011.

[79] Kelsey R Fulton, Rebecca Gelles, Alexandra McKay, Yasmin Abdi, Richard Roberts, and Michelle L Mazurek. The Effect of Entertainment Media on Mental Models of Computer Security. In *Symposium on Usable Privacy and Security*, pages 79–95, 2019.

[80] Steven Furnell and Kerry-Lynn Thomson. Recognising and addressing 'security fatigue'. *Computer Fraud & Security*, 2009(11):7–11, 2009.

[81] Xianyi Gao, Yulong Yang, Can Liu, Christos Mitropoulos, Janne Lindqvist, and Antti Oulasvirta. Forgetting of Passwords: Ecological Theory and Data. In *USENIX Security Symposium*, 2018.

[82] Shirley Gaw and Edward W Felten. Password management strategies for online accounts. In *Symposium On Usable Privacy and Security*, pages 44–55, 2006.

[83] Vasileios Gkioulos, Gaute Wangen, Sokratis K Katsikas, George Kavallieratos, and Panayiotis Kotzanikolaou. Security Awareness of the Digital Natives. *Information*, 8(2), 2017.

[84] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. What was that site doing with my Facebook password?: Designing Password-Reuse Notifications. In *ACM Conference on Computer and Communications Security*, pages 1549–1566, 2018.

[85] Paul A Grassi, James L Fenton, Elaine M Newton, Ray A Perlner, Andrew R Regenscheid, William E Burr, Justin P Richter, Naomi B Lefkovitz, Jamie M Danker, Yee-Yin Choong, Kristen K Greene, and Mary F Theofanos. NIST Special Publication 800-63B: Digital Identity Guidelines: Authentication and Lifecycle Management. Technical report, June 2017.

[86] Paul A Grassi, Michael E Garcia, and James L Fenton. Digital Identity Guidelines (Revision 3). Technical report, National Institute of Standards and Technology, 2017.

[87] Josh Grunzweig, Brandon Levene, Kyle Wilhoit, and Pat Litke. SquirtDanger: The Swiss Army Knife Malware from Veteran Malware Author TheBottle, 2018. `https://unit42.paloaltonetworks.com/unit42-squirtdanger-swiss-army-knife-malware-veteran-malware-author-thebottle/` (accessed: 2019-07-28).

[88] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, 30(4):208–220, 2011.

[89] Andreas Gutmann, Karen Renaud, Joseph Maguire, Peter Mayer, Melanie Volkamer, Kanta Matsuura, and Jörn Müller-Quade. ZeTA-Zero-Trust Authentication: Relying on Innate Human Ability, Not Technology. In *European Symposium on Security and Privacy*, pages 357–371, 2016.

[90] Hana Habib, Pardis Emami Naeini, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. User Behaviors and Attitudes Under Password Expiration Policies . In *Symposium on Usable Privacy and Security*, pages 13–30, 2018.

[91] Felix Haeussinger and Johann Kranz. Antecedents of Employees' Information Security Awareness - Review, Synthesis, and Directions for Future Research. In *European Conference on Information Systems*, pages 1–20, 2017.

[92] Ameya Hanamsagar, Simon S Woo, Chris Kanich, and Jelena Mirkovic. Leveraging Semantic Transformation to Investigate Password Habits and Their Causes. In *Conference on Human Factors in Computing Systems*, 2018.

[93] Norman Hänsch and Zinaida Benenson. Specifying IT Security Awareness. *International Workshop on Database and Expert Systems Applications*, pages 326–330, 2014.

[94] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Symposium on Usable Privacy and Security*, pages 213–230, 2014.

[95] Eiji Hayashi and Jason I. Hong. A diary study of password usage in daily life. In *Conference on Human Factors in Computing Systems*, pages 2627–2630, 2011.

[96] Jordan Hayes, Bryan Dosono, and Yang Wang. "They Should Be Convenient and Strong": Password Perceptions and Practices of Visually Impaired Users. In *iConference*, pages 445–451, 2017.

[97] Cormac Herley. So long, and no thanks for the externalities. In *New Security Paradigms Workshop*, pages 133–144, 2009.

[98] Cormac Herley and Paul C van Oorschot. A Research Agenda Acknowledging the Persistence of Passwords. *IEEE Security & Privacy*, 10(1):28–36, 2012.

[99] Robert Buzz Hill. Retina Identification. In *Biometrics*, pages 123–141. Springer, 1999.

[100] Max Hlywa, Robert Biddle, and Andrew S Patrick. Facing the facts about image type in recognition-based graphical passwords. In *Annual Computer Security Applications Conference*, pages 149–158, 2011.

[101] Chiung Ching Ho, C Eswaran, Kok-Why Ng, and June-Yee Leow. An unobtrusive Android person verification using accelerometer based gait. In *International Conference on Advances in Mobile Computing & Multimedia*, pages 271–274, 2012.

[102] Ralph Holz, Thomas Riedmaier, Nils Kammenhuber, and Georg Carle. X.509 Forensics: Detecting and Localising the SSL/TLS Men-in-the-Middle. In *European Symposium on Research in Computer Security*, pages 217–234, 2012.

[103] Shiva Houshmand, Sudhir Aggarwal, and Randy Flood. Next Gen PCFG Password Cracking. *IEEE Transactions on Information Forensics and Security*, 10(8), 2015.

[104] Lin-Shung Huang, Alex Rice, Erling Ellingsen, and Collin Jackson. Analyzing Forged SSL Certificates in the Wild. In *IEEE Symposium on Security and Privacy*, pages 83–97, 2014.

[105] Tom Huddleston. How this scammer used phishing emails to steal over $100 million from Google and Facebook, 2019. `https://www.cnbc.com/2019/03/27/phishing-email-scam-stole-100-million-from-facebook-and-google.html` (accessed: 2019-07-14).

[106] Tehreem Hussain, Kiran Atta, N Z Bawany, and Tehreem Qamar. Passwords and User Behavior. *Journal of Computers*, 13(6):692–704, 2018.

[107] Philip G Inglesant and M Angela Sasse. The true cost of unusable password policies. In *Conference on Human Factors in Computing Systems*, pages 383–392, 2010.

[108] International Organization for Standardization. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts. Technical report.

[109] Iulia Ion, Rob Reeder, and Sunny Consolvo. "...no one can hack my mind": Comparing Expert and Non-Expert Security Practices. In *USENIX Security Symposium*, pages 327–346, 2015.

[110] Blake Ives, Kenneth R Walsh, and Helmut Schneider. The domino effect of password reuse. *Communications of the ACM*, 47(4):75–78, 2004.

[111] Aviv Adam J, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M Smith. Smudge attacks on smartphone touch screens. In *Workshop on Offensive Technologies*, 2010.

[112] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K Reiter, and Aviel D Rubin. The design and analysis of graphical passwords. In *USENIX Security Symposium*, 1999.

[113] Philip Tully John Seymour. Weaponizing data science for social engineering: Automated e2e spear phising on twitter. DEF CON, 2016.

[114] Zach Jorgensen and Ting Yu. On mouse dynamics as a behavioral biometric for authentication. In *Asia Conference on Computer Communications Security*, pages 476–482, 2011.

[115] Miranda Kajtazi and Burcu Bulgurcu. Information Security Policy Compliance: An Empirical Study on Escalation of Commitment. In *Americas Conference on Information Systems*, 2013.

[116] Ehud D Karnin, Jonathan W Greene, and Martin E Hellman. On secret sharing systems. *IEEE Transactions on Information Theory*, 29(1):35–41, 1983.

[117] Michaela Kauer, Sebastian Günther, Daniel Storck, and Melanie Volkamer. A Comparison of American and German Folk Models of Home Computer Security. In *Human Aspects of Information Security, Privacy, and Trust*, pages 100–109. 2013.

[118] Joseph C Kawan. Method and system of contactless interfacing for smart card banking. United States Patent and Trademark Office, 2013.

[119] Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Christian Wiedeman, Lorrie Faith Cranor, and Julio Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *IEEE Symposium on Security and Privacy*, pages 523–537, 2012.

[120] Hamed Ketabdar, Kamer Ali Yuksel, Amirhossein Jahnbekam, Mehran Roshandel, and Daria Skripko. MagiSign: User Identification/Authentication. In *International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2010.

[121] Hassan Khan, Urs Hengartner, and Daniel Vogel. Evaluating Attack and Defense Strategies for Smartphone PIN Shoulder Surfing. In *Conference on Human Factors in Computing Systems*, pages 164–10, 2018.

[122] David Kim, Paul Dunphy, Pam Briggs, Jonathan Hook, John W Nicholson, James Nicholson, and Patrick Olivier. Multi-touch authentication on tabletops. In *Conference on Human Factors in Computing Systems*, pages 1093–1102, 2010.

[123] Sung-Hwan Kim, Jong-Woo Kim, Seon-Yeong Kim, and Hwan-Gue Cho. A new shoulder-surfing resistant password for mobile environments. In *International Conference on Ubiquitous Information Management and Communication*, 2011.

[124] Maria M King. Rebus passwords. In *Annual Computer Security Applications Conference*, pages 239–243, 1991.

[125] Knuddels.de. Vorsichtsmaßnahme Passwortsicherheit, 2018. `https://forum.knuddels.de/ubbthreads.php?ubb=showflat&Number=2916081` (accessed: 2019-25-01).

[126] Vijay Kothari, Ross Koppel, Jim Blythe, and Sean Smith. Password Logbooks and What Their Amazon Reviews Reveal About Their Users' Motivations, Beliefs, and Behaviors. In *European Workshop on Usable Security*, 2017.

[127] Katharina Krombholz, Thomas Hupperich, and Thorsten Holz. Use the Force: Evaluating Force-Sensitive Authentication for Mobile Devices. In *Symposium on Usable Privacy and Security*, pages 207–2019, 2016.

[128] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How Effective is Anti-Phishing Training for Children? In *Symposium on Usable Privacy and Security*, pages 229–239, 2017.

[129] Benedikt Lebek, Jörg Uffen, Markus Neumann, Bernd Hohler, and Michael H Breitner. Information security awareness and behavior: a theory-based literature review. *Management Research Review*, 37(12):1049–1092, 2014.

[130] Canchu Lin and Anand S Kunnathur. Toward Developing a Theory of End User Information Security Competence. In *Americas Conference on Information Systems*, pages 1–10, 2013.

[131] Birgy Lorenz, Kaido Kikkas, and Aare Klooster. "The Four Most-Used Passwords Are Love, Sex, Secret, and God": Password Security and Training in Different User Groups. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 276–283, 2013.

[132] Ijlal Loutfi and Audun Jøsang. Passwords are not always stronger on the other side of the fence. In *Workshop on Usable Security*, 2015.

[133] Karlos Luna. If it is easy to remember, then it is not secure: Metacognitive beliefs affect password selection. *Applied Cognitive Psychology*, 33(5):744–758, 2019.

[134] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse. In *USENIX Security Symposium*, pages 203–220, 2018.

[135] Federico Maggi, Alberto Volpatto, Simone Gasparini, Giacomo Boracchi, and Stefano Zanero. A fast eavesdropping attack against touchscreens. In *International Conference on Information Assurance and Security*, pages 320–325, 2012.

[136] Nathan Malkin, Marian Harbach, Alexander De Luca, and Serge Egelman. THE ANATOMY OF SMARTPHONE UNLOCKING: Why and How Android Users Around the World Lock their Phones. *GetMobile: Mobile Computing and Communications*, 20(3):42–46, 2017.

[137] Pardon Blessings Maoneke, Stephen Flowerday, and Naomi Isabirye. The Influence of Native Language on Password Composition and Security: A Socioculture Theoretical View. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pages 33–46, 2018.

[138] Peter Mayer, Nina Gerber, Benjamin Reinheimer, Philipp Rack, Kristoffer Braun, and Melanie Volkamer. I (don't) see what you typed there! Shoulder-surfing resistant password entry on gamepads. In *Conference on Human Factors in Computing Systems*, 2019.

[139] Peter Mayer, Stephan Neumann, Daniel Storck, and Melanie Volkamer. Supporting Decision Makers in Choosing Suitable Authentication Schemes. In *International Symposium of Human Aspects of Information Security Assurance*, pages 67–77, 2016.

[140] Peter Mayer and Melanie Volkamer. Poster: Secure Storage of Masked Passwords. In *Euro S&P Posters*, 2017.

[141] Peter Mayer, Melanie Volkamer, and Michaela Kauer. Authentication Schemes - Comparison and Effective Password Spaces. In *International Conference on Information System Security*, pages 204–225, 2014.

[142] Tanya McGill and Nik Thompson. Old risks, new challenges: exploring differences in security between home computer and mobile device use. *Behaviour & Information Technology*, 36(11):1111–1124, 2017.

[143] William Melicher, Michelle L Mazurek, Darya Kurilova, Sean M Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Usability and Security of Text Passwords on Mobile Devices. In *Conference on Human Factors in Computing Systems*, pages 527–539, 2016.

[144] Burak Merdenyan and Helen Petrie. Perceptions of the risks of password related activities. In *British Computer Society Human Computer Interaction Conference*, 2017.

[145] Fabian Monrose and Aviel D Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, 16(4):351–359, 2000.

[146] Lewis Morgan. Four embarrassing password leaks on live TV, 2015. `https://www.itgovernance.co.uk/blog/four-embarrassing-password-leaks-on-live-tv` (accessed: 2019-07-14).

[147] Kathleen Moriarty, Burt Kaliski, and Andreas Rusch. PKCS #5: Password-Based Cryptography Specification Version 2.1. Technical report, January 2017.

[148] Edith F Mulhall. Experimental Studies in Recall and Recognition. *The American Journal of Psychology*, 26(2):217–228, 1915.

[149] Hazel Murray and David Malone. Evaluating password advice. In *Irish Signals and Systems Conference*, 2017.

[150] National Cyber Security Centre. Password Guidance: Simplifying Your Approach. Technical report, 2016.

[151] Douglas L Nelson, Valerie S Reed, and Walling John R. Pictorial Superiority Effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2:523–528, 1976.

[152] Stephan Neumann, Benjamin Reinheimer, and Melanie Volkamer. Don't Be Deceived: The Message Might Be Fake. In *International Conference on Trust and Privacy in Digital Business*, pages 199–214, 2017.

[153] James Nicholson, Lynne Coventry, and Pam Briggs. Introducing the Cybersurvival Task: Assessing and Addressing Staff Beliefs about Effective Cyber Protection. In *Symposium on Usable Privacy and Security*, pages 427–441, 2018.

[154] Gilbert Notoatmodjo and Clark Thomborson. Passwords and perceptions. In *Australasian Information Security Conference*, pages 71–78, 2009.

[155] Gizem Öğütçü, Özlem Müge Testik, and Oumout Chouseinoglou. Analysis of personal information security behavior and awareness. *Computers & Security*, 56:83–93, 2016.

[156] Allan Paivio, T B Rogers, and Padric C Smythe. Why are pictures easier to recall than words? *Psychonomic Science*, 11:137–138, 1968.

[157] Fayyaadh Parker, Jacques Ophoff, Jean-Paul Van Belle, and Ross Karia. Security awareness and adoption of security controls by smartphone users. In *eCrime Researchers Summit*, 2015.

[158] Kathryn Parsons, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Computers & Security*, 66:40–51, 2017.

[159] CX Partners. Joypad evolution - Design patterns and innovations in game controllers (visited 30.04.2019), 2010. `https://web.archive.org/web/20101225110143/http://www.cxpartners.co.uk/wp-content/uploads/joypad-evolution-small.pdf` (accessed: 2019-04-30).

[160] PCI Security Standards Council LLC. Payment Card Industry (PCI) Data Security Standard (Version 3.2), April 2016.

[161] Sarah Pearman, Jeremy Thomas, Pardis Emani Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. Let's go in for a closer look: Observing passwords in their natural habitat. In *Conference on Computer and Communications Security*, 2017.

[162] Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568, 2011.

[163] Shari Lawrence Pfleeger, M Angela Sasse, and Adrian Furnham. From Weakest Link to Security Hero: Transforming Staff Security Behavior. *Journal of Homeland Security and Emergency Management*, 11(4):489–510, 2014.

[164] Ugo Piazzalunga, Paolo Salvaneschi, and Paolo Coffetti. The Usability of Security Devices. In *Security and Usability*. O'Reilly Media, 2005.

[165] Josef Pieprzyk, Thomas Hardjono, and Jennifer Seberry. *Fundamentals of Computer Security*. Springer, 2003.

[166] M Kameswara Rao and Sushma Yalamanchili. Novel Shoulder-Surfing Resistant Authentication Schemes using Text-Graphical Passwords. *International Journal of Information Network Security*, 1(3):163–170, 2012.

[167] Real User Corporation. The Science Behind Passfaces. Technical report, 2004.

[168] Karen Renaud. Evaluating Authentication Mechanisms. In *Security and Usability*, pages 103–128. O'Reilly Media, 2005.

[169] Karen Renaud and Joseph Maguire. Armchair authentication. In *British HCI Group Annual Conference on People and Computers*, 2009.

[170] Riva Richmond. The RSA Hack: How They Did It, 2011. `https://bits.blogs.nytimes.com/2011/04/02/the-rsa-hack-how-they-did-it/` (accessed: 2019-07-14).

[171] Caitlin Rinn, Kathryn Summers, Emily Rhodes, Joël Virothaisakun, and Dana Chisnell. Password creation strategies across high- and low-literacy web users. In *ASIST Annual Meeting Information Science with Impact Research in and for the Community*, 2015.

[172] Elena Rocco. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *International Conference on Human Factors in Computing Systems*, pages 496–502, 1998.

[173] Volker Roth, Kai Richter, and Rene Freidinger. A PIN-entry method resilient against shoulder surfing. In *ACM Conference on Computer and Communications Security*, pages 236–245, 2004.

[174] Nader Sohrabi Safa, Mehdi Sookhak, Rossouw von Solms, Steven Furnell, Norjihan Abdul Ghani, and Tutut Herawan. Information security conscious care behaviour formation in organizations. *Computers & Security*, 53:65–78, 2015.

[175] Alireza Sahami Shirazi, Peyman Moghadam, Hamed Ketabdar, and Albrecht Schmidt. Assessing the vulnerability of magnetic gestural authentication to video-based shoulder surfing attacks. In *Conference on Human Factors in Computing Systems*, pages 2045–2048, 2012.

[176] Hirokazu Sasamoto, Nicolas Christin, and Eiji Hayashi. Undercover: authentication usable in front of prying eyes. In *Conference on Human Factors in Computing Systems*, pages 183–192, 2008.

[177] M Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3):122–131, 2001.

[178] M Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. The Great Authentication Fatigue – And How to Overcome It. In *Cross-Cultural Design*, pages 228–239. 2014.

[179] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. Self-Confidence Trumps Knowledge: A Cross-Cultural Study of Security Behavior. In *International Conference on Human Factors in Computing Systems*, pages 2202–2214, 2017.

[180] Florian Schaub, Marcel Walch, Bastian Könings, and Michael Weber. Exploring The Design Space of Graphical Passwords on Smartphones. In *Symposium on Usable Privacy and Security*, 2013.

[181] Bruce Schneier. *Secrets and lies.* John Wiley, 2000.

[182] Bruce Schneier. Sensible Authentication. *Queue*, 1(10):74, 2004.

[183] Bruce Schneier. Write Down Your Password, May 2005. `https://www.schneier.com/blog/archives/2005/06/write_down_your.html` (accessed: 2016-08-09).

[184] Bruce Schneier. "Evil Maid" Attacks on Encrypted Hard Drives, 2009. `https://www.schneier.com/blog/archives/2009/10/evil_maid_attac.html` (accessed: 2019-08-07).

[185] Daniel Schonberg and Darko Kirovski. Iris-based biometric identification. United States Patent and Trademark Office, 2008.

[186] Tobias Seitz and Heinrich Hussmann. PASDJO: Quantifying Password Strength Perceptions with an Online Game. In *Australasian Conference on Computer-Human Interaction*, pages 117–125, 2017.

[187] Adi Shamir. How to share a secret. *Communications of the ACM*, 1979.

[188] Richard Shay, Saranga Komanduri, Adam L Durity, Phillip Seyoung Huh, Michelle L Mazurek, Sean M Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Designing Password Policies for Strength and Usability. *ACM Transactions on Information and System Security*, 18(4):13–34, 2016.

[189] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In *Symposium on Usable Privacy and Security*, pages 88–99, 2007.

[190] Richard M Shiffrin and Walter Schneider. Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending and a General Theory. *Psychological Review*, 84(2):127–190, 1977.

[191] Statista. The Future of Mobile Biometrics (visited 30.04.2019), 2017. `https://www.statista.com/chart/11122/the-future-of-mobile-biometrics/` (accessed: 2019-04-30).

[192] Statista. Share of smartphone shipments with a fingerprint sensor worldwide from 2014 to 2018 (visited 30.04.2019), 2019. `https://www.statista.com/statistics/804269/global-smartphone-fingerprint-sensor-penetration-rate/` (accessed: 2019-04-30).

[193] Elizabeth Stobert. The agony of passwords: can we learn from user coping strategies? In *CHI Conference Extended Abstracts*, pages 975–980, 2014.

[194] Elizabeth Stobert and Robert Biddle. The Password Life Cycle: User Behaviour in Managing Passwords. In *USENIX Security Symposium*, 2014.

[195] Elizabeth Stobert and Robert Biddle. Expert Password Management. In *International Conference on Passwords*, pages 3–20. Springer International Publishing, 2015.

[196] Elizabeth Stobert and Robert Biddle. The Password Life Cycle. *ACM Transactions on Privacy and Security (TOPS)*, 21(3), 2018.

[197] Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer, Peter Mayer, Alexandra Kunz, Philipp Rack, and Daniel Lehmann. Teaching Phishing-Security: Which Way is Best? In *ICT Systems Security and Privacy Protection*, pages 135–149. 2016.

[198] Xiaoyuan Suo, Ying Zhu, and G Scott Owen. Graphical Passwords: A Survey. In *Annual Computer Security Applications Conference*, 2005.

[199] Kishore Swaminathan and Steve Sato. Interaction design for large displays. *interactions*, 4(1):15–24, 1997.

[200] Swype Inc. What is Swype?, 2010. `https://web.archive.org/web/20100715050803/http://swypeinc.com/product.html` (accessed: 2018-04-07).

[201] Peter Szor. *The Art of Computer Virus Research and Defense*. Pearson Education, 2005.

[202] Desney S Tan and Mary Czerwinski. Information voyeurism: social impact of physically large displays on information privacy. In *CHI Conference Extended Abstracts*, pages 748–749, 2003.

[203] M E Thomson and R von Solms. Information security awareness: educating your users effectively. *Information Management & Computer Security*, 6(4):167–173, 1998.

[204] Aggeliki Tsohou, Maria Karyda, and Spyros Kokolakis. Analyzing the role of cognitive and cultural biases in the internalization of information security policies: Recommendations for information security awareness programs. *Computers & Security*, 52:128–141, 2015.

[205] Harshal Tupsamudre, Rahul Wasnik, Shubhankar Biswas, Sankalp Pandit, Sukanya Vaddepalli, Aishwarya Shinde, C J Gokul, Vijayanand Banahatti, and Sachin Lodha. GAP: Game for Improving Awareness About Passwords. In *Joint International Conference on Serious Games*, pages 66–78, 2018.

[206] Barbara Tversky. Encoding Processes in Recognition and Recall. *Cognitive Psychology*, 5(3):275–287, 1973.

[207] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. Design and Evaluation of a Data-Driven Password Meter. In *Conference on Human Factors in Computing Systems*, pages 3775–3786, 2017.

[208] Blase Ur, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Do Users' Perceptions of Password Security Match Reality? In *Conference on Human Factors in Computing Systems*, pages 3748–3760, 2016.

[209] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Symposium on Usable Privacy and Security*, pages 123–140, 2015.

[210] Blase Ur, Sean M Segreti, L Bauer, N Christin, L F Cranor, Saranga Komanduri, Darya Kurilova, Michelle L Mazurek, William Melicher, and Richard Shay. Measuring real-world accuracies and biases in modeling password guessability. In *USENIX Security Symposium*, pages 463–481, 2015.

[211] Valve Corporation. Controller Gaming on PC, September 2018. `https://steamcommunity.com/games/593110/announcements/detail/1712946892833213377` (accessed: 2019-04-30).

[212] Rafael Veras, Julie Thorpe, and Christopher Collins. Visualizing Semantics in Passwords: The Role of Dates. In *International Symposium on Visualization for Cyber Security*, pages 88–95, 2012.

[213] Verizon. 2016 Data Breach Investigations Report. Technical report, 2016.

[214] Verizon. 2017 Data Breach Investigations Report. Technical report, 2017.

[215] Melanie Volkamer, Andreas Gutmann, Karen Renaud, Paul Gerber, and Peter Mayer. Replication Study: A Cross-Country Field Observation Study of Real World PIN Usage at ATMs and in Various Electronic Payment Scenarios. In *Symposium on Usable Privacy and Security*, pages 1–11, 2018.

[216] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. User experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. *Computers & Security*, 71:100–113, 2017.

[217] E von Zezschwitz and A De Luca. Honey, I shrunk the keys: influences of mobile devices on password composition and authentication performance. In *Nordic Conference on Human-Computer Interaction*, pages 461–470, 2014.

[218] Emanuel von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. SwiPIN: Fast and Secure PIN-Entry on Smartphones. In *Conference on Human Factors in Computing Systems*, pages 1403–1406, 2015.

[219] Emanuel von Zezschwitz, Alexander De Luca, Philipp Janssen, and Heinrich Hussmann. Easy to Draw, but Hard to Trace?: On the Observability of Grid-based (Un)lock Patterns. In *Conference on Human Factors in Computing Systems*, pages 2339–2342, 2015.

[220] Chun Wang, Steve T K Jan, Hang Hu, Douglas Bossart, and Gang Wang. The Next Domino to Fall: Empirical Analysis of User Passwords across Online Services. In *Conference on Data and Application Security and Privacy*, pages 196–203, 2018.

[221] Ding Wang, Debiao He, Haibo Cheng, and Ping Wang. fuzzyPSM: A New Password Strength Meter Using Fuzzy Probabilistic Context-Free Grammars. In *APWG Symposium on Electronic Crime Research*, pages 595–606, 2016.

[222] Ding Wang, Ping Wang, Debiao He, and Yuan Tian. Birthday, Name and Bifacial-security: Understanding Passwords of Chinese Web Users. In *USENIX Security Symposium*, pages 1537–1555, 2019.

[223] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. Understanding Password Choices: How Frequently Entered Passwords Are Re-used across Websites. In *Symposium on Usable Privacy and Security*, pages 175–188, 2016.

[224] Daniel Lowe Wheeler. zxcvbn: Low-Budget Password Strength Estimation. In *USENIX Security Symposium*, pages 157–173, 2016.

[225] Zack Whittaker. GitHub says bug exposed some plaintext passwords, May 2018. `https://www.zdnet.com/article/github-says-bug-exposed-account-passwords/` (accessed: 2019-07-11).

[226] Susan Wiedenbeck, Jim Waters, Leonardo Sobrado, and Jean-Camille Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Working Conference on Advanced Visual Interfaces*, pages 177–184, 2006.

[227] Oliver Wiese and Volker Roth. See you next time: A model for modern shoulder surfers. In *International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 453–464, 2016.

[228] Mark Wilson and Joan Hash. NIST Special Publication 800-50: Building an Information Technology Security Awareness and Training Program. Technical Report 800-50, National Institute of Standards and Technology, 2003.

[229] Christian Winkler, Jan Gugenheimer, Alexander De Luca, Gabriel Haas, Philipp Speidel, David Dobbelstein, and Enrico Rukzio. Glass Unlock: Enhancing Security of Smartphone Unlocking through Leveraging a Private Near-eye Display. In *Conference on Human Factors in Computing Systems*, pages 1407–1410, 2015.

[230] Nicholas Wright, Andrew S Patrick, and Robert Biddle. Do You See Your Password? Applying Recognition to Textual Passwords. In *Symposium on Usable Privacy and Security*, pages 1–14, 2012.

[231] Khalil W Yacoub and Anshuman Sinha. Physical access control system with smartcard and methods of operating. United States Patent and Trademark Office, 2011.

[232] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. Password memorability and security: empirical results. *IEEE Security & Privacy*, 2(5):25–31, 2004.

[233] Nur Haryani Zakaria, David Griffiths, Sacha Brostoff, and Jeff Yan. Shoulder surfing defence for recall-based graphical passwords. In *Symposium on Usable Privacy and Security*, 2011.

[234] Yinqian Zhang, Fabian Monrose, and Michael K Reiter. The security of modern password expiration: an algorithmic framework and empirical analysis. In *ACM Conference on Computer and Communications Security*, pages 176–186, 2010.

[235] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. Password advice shouldn't be boring: Visualizing password guessing attacks. In *eCrime Researchers Summit*, 2013.

[236] Leah Zhang-Kennedy, Sonia Chiasson, and Paul van Oorschot. Revisiting Password Rules: Facilitating Human Management of Passwords. In *APWG Symposium on Electronic Crime Research*, 2016.

[237] Huanyu Zhao and Xiaolin Li. S3PAS: A scalable shoulder-surfing resistant textual-graphical password authentication scheme. In *Advanced Information Networking and Applications Workshops*, pages 467–472, 2007.

[238] Yingbo Zhou and Ajay Kumar. Human Identification Using Palm-Vein Images. *IEEE Transactions on Information Forensics and Security*, 6(4):1259–1274, 2011.

[239] Verena Zimmermann and Nina Gerber. The password is dead, long live the password – A laboratory study on user perceptions of authentication schemes. *International Journal of Human-Computer Studies*, 133:26–44, 2020. (pre-print accessible since: 2019-08-23).

[240] Verena Zimmermann and Karen Renaud. Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies*, 2019.