USABLE AUTHENTICATION WITH RECOGNITION-BASED GRAPHICAL PASSWORDS

Martin Mihajlov

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, September 2011

Evaluation Board:

Prof. Vladislav Rajkovič, PhD, Chairman, Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia

Prof. Dejan Dinevski, PhD, Member, Faculty of Education, University of Maribor, Maribor, Slovenia

Prof. Mark Springett PhD, Member, Department of Computing and Multimedia Technology, Middlesex University, London, United Kingdom

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Martin Mihajlov

Usable Authentication with Recognition-based Graphical Passwords

Doctoral Dissertation

Uporabno overjanje na podlagi prepoznavanja grafičnih gesel

Doktorska disertacija

Supervisor: Borka Jerman Blažič, Ph.D.

Co-Supervisor: Tomaž Klobučar, Ph.D.

Ljubljana, Slovenia, September 2011

Index

Abstract ..... **IX**

Povzetek ..... **XI**

Abbreviations ..... **XIII**

1 Introduction **1**
1.1 Hypothesis 2
1.2 Main Contributions 3
1.3 Methodology 3
1.4 Overview of the Dissertation 4
1.5 Related Publications 5

2 Theoretical Background **7**
2.1 Security or Usability? 7
2.2 Usable Security 8
2.2.1 User-centered Design 8
2.2.2 AEGIS 9
2.2.3 Yee's Actor-Ability Model 10
2.2.4 Design Pitfalls 10
2.2.5 Norman's Error Analysis 11
2.2.6 "Safe Staging" & "Metaphor Tailoring" 11
2.3 Usable Security in Authentication 12
2.3.1 Defining Authentication 12
2.4 Graphical Authentication 13
2.4.1 Recall-based Authentication 14
2.4.2 Recognition-based Authentication 16
2.4.3 Hybrid Developments 17
2.5 On the Security of Graphical Authentication 18
2.5.1 Guessing Attacks 19
2.5.2 Capture Attacks 20
2.6 Summary 20

3 Preliminary Analysis of Graphical Authentication Concepts **23**
3.1 Experiment 1: Prototyping Ubiquitous Image Authentication 23
3.1.1 Introduction 23
3.1.2 Mobile Prototyping and Graphical Authentication 25
3.1.3 Research Questions and Hypothesis 26
3.1.4 Participants and Equipment 27

3.1.5 Experiment Design 27
3.1.6 Procedure 27
3.1.7 Results and Discussion 29
3.2 Experiment 2: Image Content 32
3.2.1 Introduction 32
3.2.2 Human Memory 32
3.2.3 Research Questions and Hypothesis 33
3.2.4 Participants and Equipment 34
3.2.5 Experiment Design 34
3.2.6 Procedure 35
3.2.7 Results 36
3.3 Summary 41

4 Designing a Recognition-based Graphical Authentication Mechanism 43
4.1 Defining the ImagePass System 43
4.1.1 Enrollment 44
4.1.2 Authentication 46
4.1.3 Key Replacement 47
4.2 Designing for Security 47
4.2.1 Predictability of ImagePass 47
4.2.2 Graphical Password Space 48
4.2.3 Increasing User Privacy 49
4.2.4 Dealing with Attacks 49
4.2.5 Designing for Usability 51
4.3 Technical Specifications 52
4.3.1 Database Structure 52
4.3.2 Application Layer 53
4.4 Continued System Designs 54
4.4.1 Web Version 54
4.4.2 Mobile Application Version 55
4.5 Summary 56

5 Evaluating Recognition-based Authentication 59
5.1 Experiment 3 – User Evaluation 59
5.1.1 Introduction 59
5.1.1.1 Target Users 59
5.1.1.2 Tasks 60
5.1.1.3 Domains 60
5.1.2 Research Questions and Hypothesis 60
5.1.3 Participants and Equipment 61
5.1.4 Experiment Design 61
5.1.5 Procedure 62
5.1.6 Results and Discussion 63
5.1.6.1 Log Analysis 63
5.1.6.2 Focus Group 65
5.2 Experiment 4: Preliminary Eye Tracking 67
5.2.1 Introduction 67
5.2.1.1 The Eye 67
5.2.1.2 Visual Attention 68

5.2.1.3 Eye Tracking Process and Methods ... 68
5.2.2 Research Questions and Hypothesis ... 69
5.2.3 Participants and Equipment ... 70
5.2.4 Experiment Design ... 70
5.2.5 Procedure ... 71
5.2.6 Results and Discussion ... 71
5.3 Summary ... 75

6 Evaluating Graphical Authentication, Part 2 ... 77
6.1 Experiment 5: Eye Tracking Graphical Authentication ... 77
6.1.1 Introduction ... 77
6.1.1.1 Scanpath Theory ... 77
6.1.2 Research Questions and Hypothesis ... 78
6.1.3 Participants and Equipment ... 79
6.1.4 Experiment Design ... 79
6.1.5 Procedure ... 81
6.1.6 Results and Discussion ... 82
6.2 Experiment 6 – Graphical Password Selection Properties ... 88
6.2.1 Introduction ... 88
6.2.2 Research Questions and Hypothesis ... 89
6.2.3 Participants and Equipment ... 90
6.2.4 Experiment Design ... 90
6.2.5 Procedure ... 91
6.2.6 Results and Discussion ... 92
6.2.6.1 Naming and Categorization ... 92
6.2.6.2 Category Analysis ... 93
6.2.6.3 Color Analysis ... 95
6.2.6.4 Shape Analysis ... 97
6.2.6.5 Password Length & Complexity ... 98
6.2.6.6 Graphical Password Selection Hotspots ... 100
6.3 Summary ... 101

7 Conclusion ... 103
7.1 Images vs. Words ... 104
7.2 Understanding ImagePass ... 105
7.3 Guidelines ... 108
7.3.1 Usability Guidelines ... 108
7.3.2 Security Guidelines ... 108
7.4 Future Work ... 109

8 References ... 111

Index of Figures ... 123

Index of Tables ... 125

Appendix A – Mnemonic Story ... 127

Appendix B – Sample Image Catalog ... 131

Appendix C – Author's Publications.....135

Abstract

The shift towards including human factors as part of system design has a direct impact on the security of the system. The users' misunderstanding of how a secure mechanism works usually results in security failures. People encounter security mechanisms daily, most often required to authenticate themselves using knowledge-based schemes such as passwords, the most common and prevalent type of authentication mechanisms plagued with security and usability problems. As technical solutions have not resolved the usability of passwords many passwords used in practice are either weak and usable or secure and unusable. Hence, in recent years graphical passwords have been proposed as a potential solution due to their improved usability features and the superior human ability to recognize and remember images. This dissertation presents a thorough research on usability and security features of recognition-based graphical passwords and proposes a new approach to authentication suitable for ubiquitous environments.

Povzetek

Pri načrtovanju novih sistemov za dostop do varnih internetnih storitev imajo pomembno vlogo na uporabniško izkušnjo vezani dejavniki, ki vplivajo na stopnjo zagotavljanja varnosti. Nerazumevanje varnostnih mehanizmov s strani uporabnikov pogosto vodi v varnostne incidente. Uporabniki so soočeni z uporabo teh mehanizmov vsak dan, najpogosteje pri dokazovanju lastne identitete s pomočjo gesel. Gesla so najpogostejša metoda overjanja, ki pa je obremenjena s poneverbami in krajo identitete v omrežju. Dosedanje tehnične rešitve niso prispevale k rešitvi problema, zato so danes gesla uporabniško prijazna, vendar ne dovolj varna, ali pa varna, vendar ne uporabniško prijazna. V zadnjih letih so se pojavila grafična gesla, zasnovana na izbiri slik. Grafična gesla so uporabnejša zaradi človekove sposobnosti, da lažje razpozna in si zapomni slike kot črke in številke. Doktorska disertacija vsebuje raziskave, usmerjene k izdelavi grafičnih gesel, ki hkrati zagotavljajo ustrezno stopnjo varnosti v storitvi overjanja in ustrezajo standardom uporabnosti te storitve. Rezultati raziskav so v obliki novih sistemov overjanja, primernih za mobilna in vseprisotna omrežena okolja.

Abbreviations

AOI = Area of Interest
CTA = Concurrent think-aloud
HCI = Human Computer Interaction
LTM = Long-term memory
OTP = One-time password
RTA = Retrospective think-aloud
STM = Short-term memory

1 Introduction

With the increased threats to networked computer systems there is a great need for security service provision. Without a secure authentication system the infrastructure and assets of the system can be easily compromised. To prevent this, security designers create applications with perfect-security which turn out to be rather demanding. The user's needs are not significantly considered, as usability issues have played a very limited role in the development of secure system where security has mostly been treated as a technical issue.

In theory, a secure computer system must employ complex and sophisticated authentication mechanisms to protect the data. As the complexity of the system increases the probability for user errors escalates drastically, which in turn weakens the security of the system (Schultz et. al, 2001). While digital security is hypothetically achievable, in practice, the human factor plays a major role in subverting the best-laid plans of system administrators and security experts. The users are often referred to as the "weak link" in computer security as people generally do not understand risks, and sometimes fail to even understand computers (Schneier, 2000).

The users' naïve approach to security issues is intensified by a mismatch of goals between the user and the system. In authentication the goal of the mechanism is to ensure the identity of the current user and prevent phishing, while the goal of the user is to access the system to carry out a specific task in mind. Usually, increasing the strength of the authentication mechanism also increases the difficulties for the legitimate user to access the desired system. When the system makes authentication burdensome and/or time-consuming, the user will look for a way around it, especially in an uncontrolled environment. In a study by (Friedman et. al, 2001) it was shown that very few users were concerned about security issues with online identities or online interactions. Only technologically-aware users expressed some consideration, suggesting that increased exposure had sensitized them to security risks. Therefore, this raises the importance that security mechanisms should be designed with improved focus towards user needs.

Usable security is concerned with the study of how security information should be handled in the system, both at the user interface and in the back-end process, without discarding consideration for resources and costs (Josang & Patton, 2003). Balancing usability and security to achieve optimal result has been defined by the principle of psychological acceptability (Saltzer & Schroeder, 1975); according to this principle, a security mechanism should not make accessing a resource, or taking some other action, more difficult than it would be if the security mechanism were not present. This means that a security mechanism should add as little as possible to the difficulty of the user's performing some action. Here the perception of "difficult" should account for the abilities, knowledge and mental models of the system users. In essence, for security to be more usable, it has to be less noticeable.

The cognometric approach has been proposed as a viable solution to usable security issues in authentication. Cognometrics is defined as a class of personal authentication techniques based on measuring innate cognitive abilities of the human brain. The core idea behind cognometrics as a usable authentication approach lies in dual-code theory which postulates that language and knowledge of words are represented by functionally

distinct memory systems classified as verbal and nonverbal (Bucci, 1985). Thus, visual authentication relies on the fact that humans have a vast memory for images. This memory, as shown by (Bower et al., 1975), is not affected by the person's cognitive abilities. In addition (Shepard, 1967) has proven that there is a substantial improvement of performance in recall and recognition with pictorial over verbal representations. Hence, cognometric systems have been designed to authenticate the user based on image input with the aid of a mouse, stylus or touch screen.

1.1 Hypothesis

The purpose of this research is to improve and develop authentication mechanisms that objectively satisfy usability guidelines for actual security structures in different environments. By defining user behavior based on previous experiences the dissertation aims to enhance the methods applied in the design of graphical authentication mechanisms.

The main goal of the dissertation is to develop a graphical authentication mechanism that is usable and secure. This goal is achieved by considering both security and usability testing during different phases of development. Specifically, the research presented in this dissertation is focused on recognition-based authentication because of the implications for improved usability and security. By using knowledge accumulated in current graphical authentication research, the proposed system allows for flexibility that is supported by human factors. Since determining how users perceive and understand graphical authentication is a complex problem the development cycle is complemented with several usability experiments as presented in the following chapters. This leads to the main hypothesis of this dissertation which can be stated as:

Recognition-based graphical passwords can be designed to simultaneously support usability and security in authenticating users on the web and in ubiquitous environments.

To verify this statement the following experiments are performed in order to address different aspects of graphical authentication.

Recognition or recall. The first experiment investigates the security and usability strengths and weaknesses of recall-based and recognition-based graphical authentication systems. It helps to verify whether findings in psychological studies regarding cognitive mental processes are applicable in the field of graphical authentication.

Image memorability. The second experiment evaluates the short-term and long-term memorability of images based on content type. This establishes the most suitable image content for graphical authentication.

User performance and behavior. The third experiment is a hybrid study that combines a lab controlled environment for enrollment, and a practical web-based study for successive authentication. The intent of the experiment is to evaluate relevant user behavior over continuous use of the system by considering established usability methods for graphical authentication mechanisms.

Cognitive perception and patterns. This experiment is a two-part eye tracking study. The first part tests the usability of the generic interface components and evaluates the initial user behavior when encountering graphical authentication. The second part tracks subsequent login attempts to analyze the cognitive differences occurring with continuous system use.

Graphical password properties. The password selection properties for a recognition-based password with a single object as image content are evaluated through this

experiment. The users' image preference in password selection is analyzed by determining favorable image content and evaluating memorability performance with differing password lengths.

1.2 Main Contributions

The research presented in this dissertation contributes original ideas and knowledge to the field of usable security. I design and test two novel graphical password schemes along with a novel approach for the implementation of click-based graphical passwords. I also conduct usability and security analysis of both a pre-existing schemes and newly proposed graphical password systems. As part of the work, I examined how design choices affect user behavior, as well as the interaction between usability and security. The original contributions are included bellow:

- A comprehensive review of the field of usable security and the existing evaluation methods (Sections 2.2 & 2.3).
- Detailed analysis and critique of current graphical authentication systems and concepts (Section 2.4).
- Defining a cognitive approach best suited for authentication in emerging mobile environments (Section 3.1, hypotheses 1 & 2).
- Defining the most memorable image content for short-term and long term memorability (Section 3.2, hypotheses 3 & 4).
- Defining a cognitive user mental model for graphical authentication (Section 5.1, hypotheses 5-8, Section 6.1, hypotheses 9, 10).
- Defining graphical password properties based on specific content characteristics and selection grid position (Section 6.2, hypotheses 11-15).
- Developing a usable and secure graphical authentication mechanism suitable for mobile environments (Chapter 4).
- Evaluating proposed concepts in real-life practical scenarios with experiments that assess the sustainability of the proposed system (Chapters 3, 5 & 6).

1.3 Methodology

The research in this dissertation determines a usable and secure approach to graphical authentication suitable for emerging mobile environments. In order to answer the research goals and hypothesis the proposed methodology uses both a theoretical and experimental research approach. The theoretical approach covers a detailed study of properties and features for known graphical authentication mechanisms. It also analyzes the existing methods used to assess usability and security in authentication mechanisms.

The experimental approach utilizes well-established methods and procedures for usability evaluation. Depending on the experiment, the data is collected through different venues. A summary of the methods and data collecting procedures is explained bellow:

- Mobile **paper prototyping** of the system is used to determine the alignment of the proposal to initial user concepts. The paper prototyping is complemented with the **think aloud** protocol to gather user's opinions.
- **Cognitive analysis** and **eye tracking** are used to study the users' mental process. In eye tracking the data is gathered through fixations and gazes in predetermined areas of interests and analyzed with computer software. Cognitive analysis methods are combined with the **retrospective think aloud** method to provide context to relevant data.

- **Log analysis** is used to systematically analyze users' actions through real-life scenarios in different environments. Preprogrammed system logs silently collect data by registering entered information, clicks, page loads and time passed between events. Each log data entry is classified based on the active screen and the clicked page item. For potential data anomalies the log analysis is complemented with informal **post-session interviews** to clarify particular occurrences of specific data anomalies.
- As common approach, **focus groups** are used for post-session analysis. The data is collected in semi-structured sessions with predetermined Likert-scale or directed questions.
- In all experiments, **statistical tools** are used for numerical analysis of the data (ex. descriptive statistics, frequency analysis, t-test, one-way ANOVA, Tukey HSD etc.)

1.4 Overview of the Dissertation

This doctoral dissertation is organized as follows. An overview of the research in usable security and graphical authentication is described as a theoretical background in Chapter 2. This provides relevant knowledge on usability and security issues with this type of mechanisms, concluding with the rationale to develop an appropriate system.

Chapter 3 focuses on the preliminary analysis of graphical authentication concepts. The type of graphical authentication mechanisms best suited for desktop and ubiquitous environments is determined through a paper prototyping experiment. By using a low fidelity prototype test on paper and a limited high fidelity prototype test on a mobile device the generic concepts for recognition-based and recall based graphical authentication mechanisms are evaluated. This is followed by an experiment that determines the most-suitable type of image suitable for recognition-based graphical authentication is presented. Building on existing research, this experiment assesses both short-term and long-term memorability of images based on different image content: abstract, face or object.

Chapter 4 presents the design features of a graphical authentication mechanism called ImagePass, which followed after the first two experiments. Besides being rooted in image-recognition, this system introduces a novel feature based on one-time passwords that increases the security of the system without compromising usability, which makes this system almost immune to guessing attacks.

In Chapter 5 the designed graphical authentication mechanism is subjected to usability and security evaluation through three different experiments. The first experiment is a hybrid study that combined a lab controlled environment for enrollment, and a realistic web-based study for successive authentication. The intent of the experiment was to evaluate relevant user behavior over continuous use of the system. It considered established usability issues for graphical authentication mechanisms that needed to be evaluated. Training and frequent use were explored as group conditions and password creation time, time to login and success rates were measured during data collection. This is followed by a two-part eye tracking experiment. The first part of the experiment tested the usability of the generic interface components and evaluated the initial user behavior when encountering graphical authentication.

Chapter 6 begins with the second part of the eye tracking experiment where subsequent login attempts are observed over different periods of time to analyze the cognitive differences occurring with continuous system use. In the final experiment, the password selection properties for a recognition-based password with single-object as

image content were evaluated. In this experiment the users' image preference in password selection was analyzed by determining favorable image content and evaluating memorability performance with differing password lengths.

Chapter 7 is a discussion on the relevance of the performed studies. It evaluates the results and defines guidelines for the design of usable and secure graphical authentication mechanisms. Furthermore, this section also presents the topics for continuous research such as multiple password interference and system assigned passwords. The main conclusion is that recognition-based graphical authentication mechanisms with single-objects as representative content can complement and sometimes even substitute text-based passwords, especially in keyboard-less systems and environments. This finding has considerable importance for the future study of graphical authentication mechanisms in non-desktop environments.

1.5 Related Publications

The research presented in this dissertation has been peer-reviewed and published in ranked academic conferences and impact factor scientific journals.

The list of published applications is as follows:

- Mihajlov M., Jerman-Blazic B., 2011. Memorability, Performance and Perception in Graphical Authentication. Interacting with Computers, Elsevier. (2010 impact factor: 1.192). http://dx.doi.org/10.1016/j.intcom.2011.09.001
- Mihajlov M., Jerman-Blazic B., Ilievski M., 2011 (in press). Recognition-based Graphical Authentication with Single-Object Images. In Proceedings of the 4th International Conference on Developments in eSystems Engineering.
- Mihajlov M., Jerman-Blazic B., Ilievski M., 2011 (in press). ImagePass – Designing Graphical Authentication for Security", In Proceedings of the 7th International Conference on Next Generation Web Services Practices. (ERA Rank: C).
- Mihajlov M., Jerman-Blazic B., Josimovski S., 2011. A Conceptual Framework for Evaluating Usable Security in Authentication Mechanisms – Usability Perspectives. In Proceedings of the 5th International Conference on Network and System Security, pp. 332-336 (ERA Rank: B).
- Mihajlov M., Jerman-Blazic B., Josimovski S., 2011. Quantifying Usability and Security in Authentication. In Proceedings of the 35th Annual IEEE Computer Software and Applications Conference, pp. 626-629 (ERA Rank: B). http://dx.doi.org/10.1109/COMPSAC.2011.87
- Mihajlov M., Jerman-Blazic B., Saikayasit R., 2010. ImagePass - Developing A Graphical Authentication Mechanism Based on Usable Security. Human Factors in Information Security Inagural Conference Poster Session.

A full list of author's publications can be found in Appendix C.

2 Theoretical Background

A fundamental principle of designing security mechanisms is that complex systems are difficult to manage. They become harder to configure, maintain and implement, with an increased probability for errors which in turn can weaken the security of the system. In essence, the complexity designed to add security has the opposite effect as it significantly decreases the usability of the system.

Maximizing both usability and security of a system is a venerable predicament. According to the principle of psychological acceptability (Saltzer & Schroeder, 1975) a security mechanism should not make accessing a resource, or taking some other action, more difficult than it would be if the security mechanism were not present. This means that a security mechanism should be least intrusive adding as little difficulty as possible to the user performing an action. However, due to the diversity of users and their differing conceptions of "added difficulty" applying this principle requires taking into account the abilities, knowledge, and mental models of all the people who will use the system. In reality, the developers often design the mechanism to meet their own expectations and models of the system which are invariably different from the expectations and models of the system's users. Consequently, security mechanisms are cumbersome and less effective than their intended purpose.

This chapter focuses on the area of usable security and graphical authentication. The first section observes the fundamentals of usable security and outlines how human factors can be applied to increase usability. It presents the need to look beyond the user interface of security tools where most of the current research and development effort is focused. The second section focuses on usable security research in the specific sub-field of authentication mechanisms and describes proposed evaluation methods. The final section presents a tighter focus on graphical authentication by exploring evaluations of proposed systems.

2.1 Security or Usability?

Security research over the past twenty years has tried to increase system security by identifying common flaws and errors, and then proposing solutions that could be used in a variety of different circumstances. But although this approach has been applied to common security problems such as buffer overflows (Cowan et al., 1998) and the transmission of passwords by clear-text protocols (Ylonen, 1996), it has seldom been applied to techniques for promoting secure Human Computer Interaction (HCI). In-depth examination of computer security literature reveals that usability issues have been largely ignored by the security community until the beginning of this century. Conversely, an examination of usability literature shows that usable security solutions were also not actively researched in the same period. It seems that researchers were busy exploring questions in both fields that didn't require knowledge and methods from the other discipline.

Cryptography is one of the elements responsible for the absence of focus to the issue of usability in security. Researchers' attention was on developing cryptographic techniques for protecting information in computer systems. Cryptography is an

intellectual and mathematically interesting problem which excludes real-world scenarios that involve users (Koops, 1999). Another factor is the industry's emphasis on bug fixing rather than secure design in software development. The standard approach for running secure systems is to make sure that virus definition files are up-to-date and that all of an operating system's latest patches and bug-fixes are downloaded and installed on a regular basis. However, this short-term approach to long-term problems fails to address the underlying causes (SANS, 2004). The final factor that contributes to avoiding usability research in security is the research field itself. HCI research requires the experimenter to experiment on human beings through user studies which are usually not needed for success in computer science (Smetters & Grinter, 2001).

From the opposite direction, there are several factors that contribute to ignoring issues in computer security by usability researchers. Primarily, HCI is a field that emerged in the 1980s and 1990s from practitioners of social and behavioral computing. In the beginning researchers were busy exploring more fundamental usability questions, hence the early work on usability simply ignored security issues, even when security was part of the overall problem (Nielsen, 1993). In addition, security research was not a priority in the pre-networked world as most computer systems were protected through physical security. Another factor is the dissonant claim that most usable security problems can be viewed as conventional usability problems that can be using conventional usability approaches (Sasse, 2003).

2.2 Usable Security

The previous section showed that usability and security are viewed as competing goals. On one hand, security experts have the tendency to reject concepts for improving usability as assistance to users might also function as assistance to a potential attacker. This leads to the refutation of more usable mechanisms because of the potential for introducing additional system vulnerabilities. Nevertheless, a usable mechanism should not be dismissed on these grounds as such a dismissal ignores the importance of human factors and economic realities, and misses the key goal of security "building systems that are actually secure, as opposed to theoretically secure" (Tognazzini, 1995).

This section reviews some of the techniques for aligning usability and security.

2.2.1 User-centered Design

User-centered design of security mechanisms is more than just user interface design because interaction with security policies and mechanisms is not limited to the point of contact. Designers can make the simplest security mechanism, but the users can still fail to use it correctly by attempting to "cut corners". In order to make an effort towards security, the users must believe that their assets are under threat, and that the security mechanism will provide effective protection against that threat (Karat, 1989). The process of building a usable secure system involves assessing security needs, elaborating use policies and designing countermeasures. In designing secure systems it must be clear that security is not the primary goal or task for users, but instead it has a supportive role in protecting users' assets. The effective performance of tasks is a key principle for designing successful systems. The performed tasks can be divided into two types: production tasks, required to achieve a goal and supporting tasks, those that enable production tasks to be carried out, but are not essential to achieving the goal. Production tasks are the reason why a system exists, and if they can't be completed effectively, the system becomes irrelevant. Therefore, production tasks are always put first and security is viewed as a supporting task. When security mechanisms are chosen without considering

productivity, individual users are left to choose between complying with security regulations and performing their job. When users understand and accept the need for security their natural inclination to circumvent security measures can be avoided.

2.2.2 AEGIS

Current security mechanisms often make unreasonable demands on all users whose work is exposed to increasing complexity in order to keep systems secure (Zurko & Simon, 1997). System administrators have to secure systems at all possible levels, developers have to understand the complexity of security implications for the systems they develop and end-users compound these problems by demanding working functionality as early as possible. To address this issue, the AEGIS integrated development method for secure systems has been proposed (Flechais et al., 2003). Appropriate and Effective Guidance for Information Security (AEGIS) is a socio-technical software engineering methodology for creating secure systems based on asset modeling, security requirements identification, risk analysis, and context of use. The purpose of this methodology is to provide system developers with simple and intuitive tools for producing a secure system that considers end-user needs (Figure 2.1).

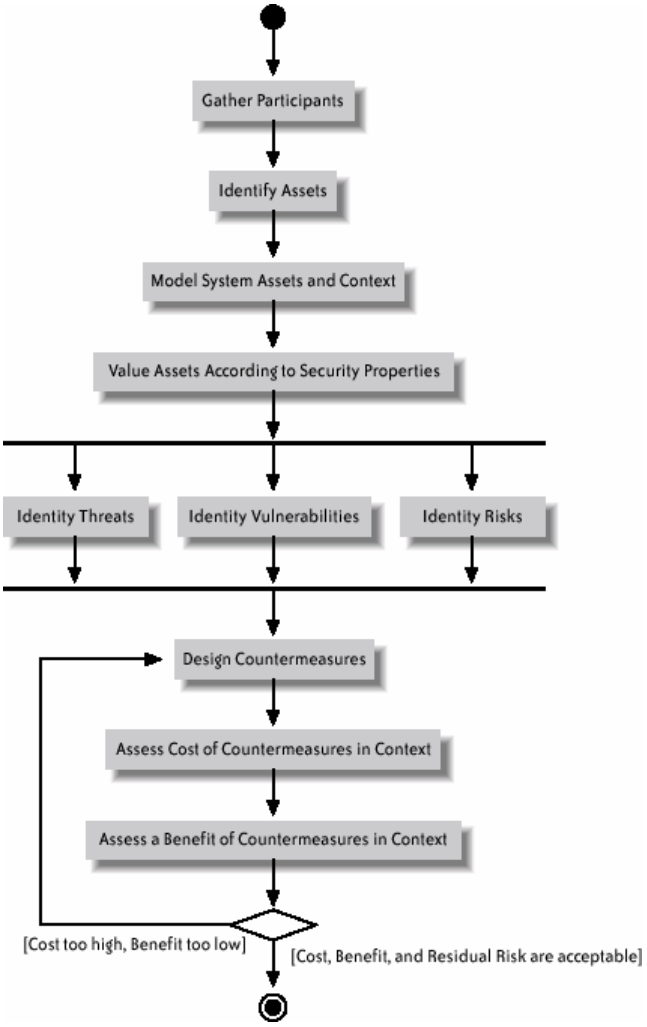


Figure 2.1: *AEGIS activity diagram*

The core AEGIS process consists of the following steps:

- Gather design process participants.

- Identify the system's assets.
- Model assets in the context of system operation.
- Identify security requirements for the assets.
- Conduct a risk analysis to identify vulnerabilities, threats, and risks.
- Design the security of the system.

This process is based on security requirements identified by the participant and risk analysis which highlights areas where the system is unacceptably vulnerable. Countermeasures are proposed and evaluated against their cost and their effectiveness, while contextual information is used to assess the cost of the countermeasure to the system as a whole. Identifying the benefit of the countermeasure depends on an assessment of the effectiveness of that measure at preventing, detecting, or reacting to risks. Based on a better understanding of the impact of the countermeasure, it can then either be accepted or rejected as a part of the architecture.

The AEGIS process provides increased security awareness for all participants. This allows them to identify a number of security problems and issues, and to provide a wealth of information about the needs of all user profiles. In addition, the security features of the system become more accessible which is makes them valuable in combating security apathy and the lack of security motivation. By providing a simple model through which the security properties of the system can be discussed the communication during the security design is improved which consequently improves the security decision-making.

2.2.3 Yee's Actor-Ability Model

According to Yee secure usability is a system property, observing that correct use of software is just as important as the correctness of the software itself. The issue is not the mere abilities of a program or a process, but how those abilities compare with the expectations of the user. This insight is the basis of Yee's Actor-Ability Model (Yee, 2003), which he uses to describe the apparent conflict between the way that users expect their computers to operate and the ways that they can actually operate.

The Actor-Ability Model is based on the capabilities available to the discrete actors' resident on the user's computer. The computer's primary actor, A0, is the computer's user. But all computers have other actors—programs which are capable of their own actions noted in this model as actors A1 . . . An. There is also a set of perceived abilities that the user believes each actor can perform, which he calls Pi. The range of actions available to the actor doesn't necessarily match the range of actions that the user believes the actor can perform. The range of actions that the actor can actually perform is defined as Ri. There is also a no-surprise condition that is true when the user is more powerful than expected and the other actors on the system are less powerful than believed.

Using this model as a basis, 10 principles for secure interaction design are formed. These principles are divided into Fundamental Principles, Actor-Ability State, and Input and Output principles. Each principle can be augmented with a test that can be used to determine if the principle is realized in a piece of secure, usable software. These principles are not the result of a systematic investigation, but are instead based on discussion with "security experts about their experiences designing software that had to be both usable and secure.

2.2.4 Design Pitfalls

According to Lederer, there are "pitfalls" in the design of applications that are designed to

protect privacy (Lederer et al., 2004). These pitfalls were defined while working on a program which controls the presentation of personal information in a ubiquitous computing environment. The identified "pitfalls" include:

Obscuring potential information flow. Systems that maintain and attempt to protect personal information perform poorly when explaining that the potential for information flow exists. Making the scope of a system's privacy implications clear would help users understand its capabilities and limits which in turn would provide grounds for comprehending the actual flow of information through the system.

Obscuring actual information flow. Systems do not clearly identify the direction and participants when information is transferred. For example, web browsers do not tell users about the existence of cookies and web bugs, let alone report when these devices are used to report personal information from the user back to the primary or a third-party web site.

Emphasizing configuration over action. Many users are unable to clearly articulate their privacy needs in advance, and even if users could predict their future privacy preferences, users are then forced to specify those preferences in detail using some kind of rule-based logic that is far removed from the day-to-day task of using a computer and then being frustrated by a privacy (or security) setting.

Lacking coarse-grained control. Users frequently want a simple, obvious control that they can use to "make it safe" or "make it private"—even if pressing this button results in making their computer generally unusable.

Inhibiting established practice. Systems frequently inhibit techniques for providing security and privacy developed by society and individuals. Examples of systems that preserve such nuance are instant messing systems that allow a user to avoid responding to an invitation to chat without having to explain why.

2.2.5 Norman's Error Analysis

Norman has observed that many users new to a computer system will make the same common errors (Norman, 1983). As errors are usually results of design flaws, they are classified as being either mistakes or slips. A mistake occurs when a user's intended action was the error. A slip occurs when the user's intention was correct but an error was made in the execution. Because mistakes are frequently the result of poor training, Norman's further analysis concentrates on slips. Slips are classified into three categories, each of which is subdivided into further subcategories. Many slips with computers arise from either the existence of modes or the inability of people to correct their errors. One of Norman's most important observations is that most errors can be overcome through the use of memory aids. When the error is an inappropriate action being performed or an inappropriate action being activated, memory aids in the form of on-screen notices, status indicators, or pop-up warnings can overcome these errors by activating the correct response.

2.2.6 "Safe Staging" & "Metaphor Tailoring"

"Safe Staging" and "Metaphor Tailoring" were specialized user interface design techniques proposed by Whitten and Tygar as a reaction to the user interface of PGP 5.0. Safe staging is a way to structure a user interface so that users may safely postpone learning how to use a particular security technology until they decide they are ready to do so (Whitten & Tygar, 2003). Safe Staging is implemented as a series of help screens that appear when the user attempts initiates a security feature. Metaphor tailoring uses conceptual model specifications that have been augmented with security risk information to create visual representations of security mechanisms and data that incorporate as many

desirable visual cues as possible.

2.3 Usable Security in Authentication

Although general usability principles have been defined there has been very little research on usability and security, especially with user authentication mechanisms. Most of the time the secure system requirements are in conflict with the appropriate usability measures, since there is no set of recognized usability principles and standards for user authentication methods and protocols. As recommended, improving usability of authentication security should not focus on educating the users about security management. Instead, there should be zero user interaction, meaning that the security mechanism should be unobtrusive, requiring no input/feedback from the user. Usable security is essentially concerned with the study of how security information should be handled in the system, both at the user interface and in the back-end process (Jøsang & Pattton, 2001), without discarding consideration for resources and costs.

Users are the "weakest link" in the security chain (Whitten & Tygar, 1999). When they fail to comply with the behavior required by a secure system, the security mechanism does not work as originally intended. The reasons for this misconduct are that users are either unable to or don't want to behave as required. Authentication mechanisms are a good example of the users' inability to comply with standard password policies. While remembering a single, frequently used password is a perfectly manageable task, most people have many different passwords to deal with. There are multiple and frequently changed passwords at work or passwords and personal identification numbers outside work, with a varied frequency of use. The limitations of human memory prevent the user from coping with high memory demands, and as a result users behave in discordance with the defined security policies (Sasse et al., 2001). Users write passwords down, stick notes with passwords onto their screens, or maintain a list of current passwords on the nearest whiteboard, share passwords with other users, choose repetitive or simple passwords etc.

2.3.1 Defining Authentication

One of the biggest challenges in digital society is the provision of secure and usable access to electronic systems. Authentication is defined as the confirmation of a claimed entity at the entry of a system in which a set of services is available to users (Ka-Ping, 2004). Before a reliable security system gives a legitimate user access to a secured system it must determine the identity of the user, confirm the user's authenticity and then authorize the user for access, as presented in Figure 2.2. Conversely, the access control of the secure system demands three distinct elements: identification, authentication and authorization.

- Identification requires the user to establish his identity by means of a token or an identification string (e.g. username).
- Authentication consists of making an accurate identity determination through a provided authentication key (e.g. password).
- Authorization determines the level of access a particular authenticated user should be granted to the secured resources controlled by the system.

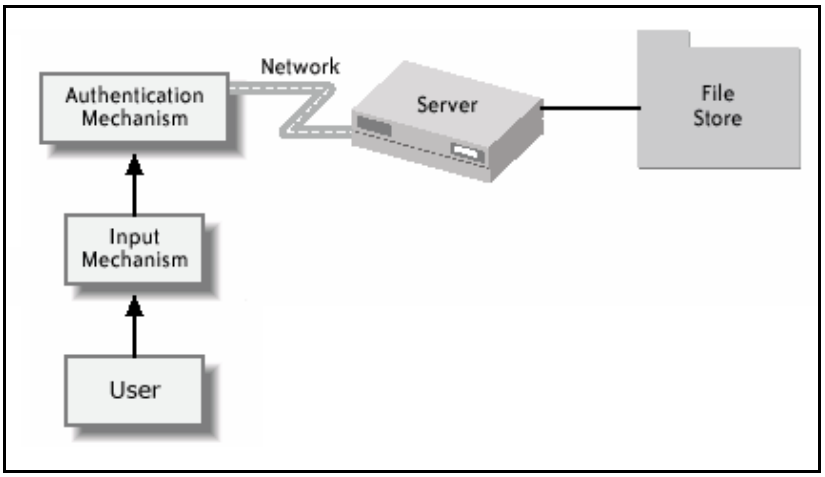


Figure 2.2: *Authentication process*

Once initiated, authentication systems have two modes of operation: enrollment and verification. The enrollment process consists of acquiring the authentication key and storing it as a template, while in the verification process, the submitted authentication key is compared against the stored template.

User access control can be established in different ways depending on authentication factors. Authentication factors determine the basis for the user's authentication. An authentication factor can be something the users knows (password), something the user is (biometrics) and something the user has (token). Authentication factors determine the strength of the authentication process and the difficulty of compromising the security of the system. In addition, the chances of an erroneous positive authentication can be reduced by requiring that users perform multiple authentications with different authentication methods in order to gain access. Multifactor authentication is deployed to prevent security breaches when the authentication key has been compromised. The most common strategy is to use two factors for authentication. This combines two system approaches to authentication, allowing the strength of each system to counterpart for the other's weakness. The familiarity with multifactor authentication comes from ATM use where the card is the storage token (what the user holds) and the PIN is the password (what the user knows). The token and associated password can be loaned or stolen, however, simply knowing the password without the token, or vice versa, is not sufficient for the attacker to successfully authenticate to the system. Both authentication keys must be used together in order to access the desired system which improves security, however the user is still inconvenienced as two things need to be remembered instead of only one. Other multifactor authentication systems combine two or more biometric authentication mechanisms (Adams & Weiner, 2002), combining biometrics with tokens (Jin et al., 2004; Ho & Armington, 2003) or biometrics with passwords (Monrose & Rubin, 2000). The usability implications differ with each combination of authentication mechanisms, requiring separate usability studies to determine the effectiveness of use.

2.4 Graphical Authentication

A graphical password is an authentication technique that has it roots in cognometrics where the innate cognitive abilities of the human brain are measured. Since the end of the last century, several graphical authentication schemes have been proposed with the objective of improving both usability and security of access-based systems. The main idea behind graphical passwords is to utilize the vast human memory for visual information and relating it to authentication by generating passwords based on images or sketches. It

has been proposed that this approach reduces memory burden and facilitates the selection and use of more secure and less predictable passwords (Suo et al., 2003).

The human brain has a superior memory for recognizing and recalling visual information as opposed to verbal or textual information. The suggested explanation for this is dual-coding theory (Bucci, 1985) which had evolved from earlier psychology concepts and studies (Shepard, 1967; Paivio et al., 1968; Madigan, 1983). This theory suggests that verbal memory and visual memory are represented by functionally distinct memory systems. Verbal items, such as text, are represented symbolically with associated meaning, while visual items, such as images are represented perceptually deriving the meaning from direct observation. The apparent additional effort necessary to render verbal memory makes this a more difficult cognitive task.

Generally, there are two popular approaches in graphical authentication based on the memory task necessary to memorize and reproduce the password: recall and recognition. In recall items are directly retrieved from the users' memory with or without any cues. In recognition the user is provided with the item and has to decide whether this matches what has been memorized. The differences between recognition and recall memory have been explained as two unique processes (Anderson & Bower, 1972), where recognition is perceived as an easier cognitive task than recall (Tulving & Watkins, 1973).

2.4.1 Recall-based Authentication

Recall based authentication mechanisms can be subdivided into cued-recall and uncued recall. Cognitively, uncued recall is a more demanding memory task than cued recall, as the absence of memory prompts or cues increases the necessary user effort to retrieve the password. Cues can improve the usability of graphical authentication as items in human memory may be available but not accessible for retrieval while previously inaccessible information in a pure recall situation can be retrieved with the aid of a retrieval cue (Tulving & Pearlstone, 1966).

Cued recall-based authentication schemes include selecting different locations on a still image as the authentication key, whereas the selected sequence constitutes the password (Blonder, 1995). The user utilizes a mouse or a stylus to click on predetermined tap locations defined with invisible boundaries (e.g. face) in a particular order (e.g. left eye, mouth nose) in order to authenticate to the system. The intent of this location specific feature is to reduce the cognitive effort required from the user. Ideally, the cue is beneficial only to genuine users.

The PassPoints system (Wiedenbeck et al., 2005b) has taken this idea in a different direction by eliminating predefined content boundaries and allowing for the use of arbitrary images. In PassPoints, the user can click anywhere on the image, with a tolerance square defined around each click-point to avoid the expectations of replicating pixel-perfect selections. These types of authentication schemes have security and usability weaknesses which make graphical passwords predictable. The users tendency to select similar locations on images allow for the forming of image hotspots (Dirik et al., 2007; Thorpe & Oorschott, 2007). Furthermore, in (Chiasson et al., 2009) the authors have shown that click sequences follow predictable geometry patterns.

Variants to PassPoints have been suggested to help create more secure passwords and improve the usability aspects of the concept. A shoulder-surfing resistant modification to PassPoints has been proposed by (Suo et al., 2005), where image sections are blurred and the user has to decide whether the focused area falls within the password range. In Cued Click-Points (CCP) (Chiasson et al., 2007) users select a click-point on 5 different images presented sequentially. Analysis of user choice in CCP passwords revealed that users had a tendency to select hotspot click-points. To help create more secure passwords, in

Persuasive Cued Click Points (Chiasson et al., 2008) the system assists users during password selection by highlighting a random part of the image where the click has to occur. PassPoints and all its variants use a grid-based discretization algorithm to determine whether the click-points fall within a tolerance range.

The most popular uncued recall approach to graphical authentication is Draw-A-Secret (DAS), drawing a free form on a 2D reference grid in a single or multiple strokes where the grid crossings represent the password (Jermyn et al., 1999). During authentication, if the drawing of the image touches the same grid crossings in the correct sequence the user is granted access to the system. In a 5x5 grid, the full password space of DAS is larger than that of the full text password space. In a study of DAS graphical dictionaries it was shown that the memorability of DAS passwords is greater when user-selected (Thorpe & Oorschot, 2004a). However, users have a tendency to draw symmetrical forms in the center of the grid with few strokes, which essentially decreases the potential password space. This is further supported by a finding in a follow-up study that showed stroke count as having a larger impact on DAS password space than stroke length (Thorpe & Oorschot, 2004b).

Emerging alternatives to DAS have improved on some usability issues. BDAS (Dunpy & Yan, 2007), adds background images to the grid in order to encourage the users to create more complex passwords and decrease the symmetries in the password form. YAGP, Yet Another Graphical Password, (Gao et. al, 2008) implements a grid with a higher detail level by using distance string matching to allow the acceptance of approximately correct drawings. Passdoodle (Varenhorst, 2004), uses a more complex matching process and discards the grid completely. PassShapes (Weiss & De Luca, 2008) translates forms into alphanumeric characters based on stroke direction which releases the drawing of the correct password from its location on the grid. In the Pass-Go scheme (Tao & Adams, 2008) users draw their passwords using grid intersection points, thus eliminating the impacts of small trace variations. Suggestions for further improvement have suggested the use of additional characteristics such as pen color, drawing speeds, pen pressure.

There are only a few recall-based authentication schemes for ubiquitous devices. GrIDSure (GrIDSure Corporation, 2010), is a commercial PDA-tested product where the user memorizes a pattern of digits in a 5x5 grid. The digit position in the grid is shuffled at every login hence each successful authentication attempt creates a different shape. Pattern-lock (Tafasa, 2010) is used for unlocking a Blackberry screen and a similar system has been adapted to Google Android cell phones. The user touch-draws the password on a 3x3 grid, however once a grid intersection is passed it cannot be revisited again. Generally, in touch-screen devices the attacker can discern the users' password through the finger smudges left on the phones' surface.

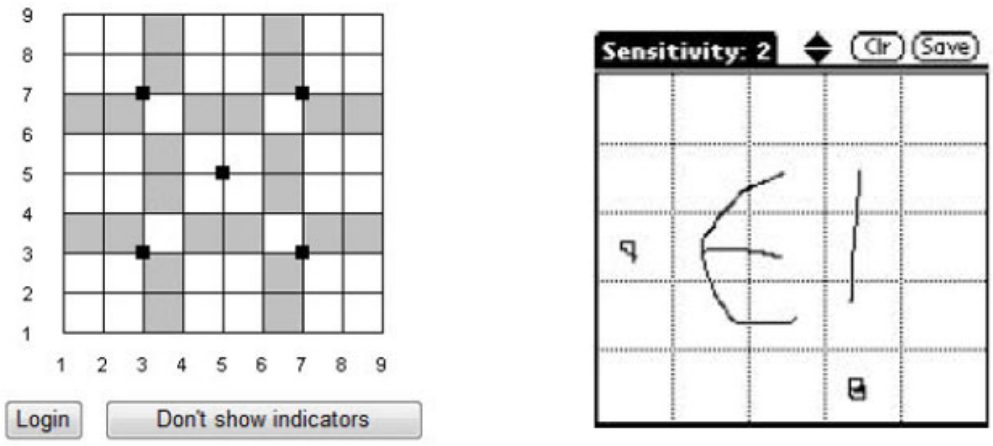


Figure 2.3: *Recall-based authentication*

2.4.2 Recognition-based Authentication

In recognition-based authentication, the graphical password is entered by the user recognizing a sequence of images from a given set of real and decoy items. The mechanisms discussed in the current research literature have varying differences in authentication approach and system interface. For the purpose of this dissertation we observe the proposed mechanisms by the type of content the image contains.

Passfaces is a commercial graphical password scheme trademarked by Passfaces Corporation which uses photographic facial images as part of the authentication key (Real User Corporation, 2008). In order to authenticate to the system the user has to recognize preselected faces through four successive image grids that contain decoys. A usability study of Passfaces (Davis et al., 2004) has demonstrated that face recognition graphical passwords have impressive success rates, while exhibiting gender, race and culture biases that create patterns in user-selected passwords. This bias greatly diminishes the password space which becomes easily exhausted, and exposes the system to an automated brute-force attack. In addition, according to (Brostoff & Sasse, 2000) the memorability rates of recognition-based passwords although higher than text-based passwords drop significantly when users use different graphical passwords to authenticate to different systems. Findings in the same study imply that users will not accept the increased execution times for frequently used passwords. Furthermore, the use of multiple Passfaces graphical passwords to authenticate to different systems causes an interference effect, which in turn initiates a dramatic decline in memorability and performance (Everitt et al., 2009).

In Story (Davis et. al, 2004) users select a sequence of photographic images for their portfolio and have to identify those images among decoys by constructing a mnemonic story. Although graphical password selections were more varied, the authors show the existence of exploitable patterns, differences between male and female choices, greater difficulties in remembering passwords and frequent ordering errors.

Photographic images are also used in the Photographic Authentication system (Pering et al., 2003), where users have to provide their own set of photos during enrollment and then identify those photos among decoys during authentication. The decoy images originate from randomly selected password images belonging to other users. Use Your Illusion (Hayashi et al., 2008) has the same general approach, but it additionally distorts the selected images after selection. The purpose of the distortion is the prevention of illegitimate access, as distorted images would be difficult to recognize by an unknown

party.

In (De Angeli et al., 2005) the authors introduce VIP, a graphical authentication prototype with clipart images representing simple, familiar and concrete everyday objects on a white background, implying that "concrete, nameable, and distinctive color images are easier to remember". This prototype is suggested as a replacement for PIN based authentication in ATM machines and compares memorability and efficiency of images against numbers.

Clipart objects as image content are also used in several graphical authentication schemes proposed by (Sobrado & Birget, 2002). In one scheme the user has to recognize preselected object images in a large grid and click inside the convex hull formed by those objects. In the second scheme the user moves a frame around until the password images align. The main drawbacks of these schemes are the rather long authentication times as well as the necessity for displaying a large amount of images on screen in order to gain a sufficiently large password space. To avoid the vulnerability to shoulder surfing (Man et al., 2003) proposed a scheme where each object image has a unique code and the user types in the string of codes for the recognized images. This password scheme is resistant to shoulder surfing as much as traditional text-based passwords, significantly increases the cognitive load on the user as the requirement is to remember both text and images.

Déjà Vu uses abstract images generated by a hash visualization algorithm for authentication (Dhamija & Perrig, 2000). The user selects a number of random computer-generated images and later identifies those images to gain access to the system. The authors claim that the system offers better security due to the inability to accurately write down the authentication key or share it with other users, as abstract images cannot be easily described with words. Nevertheless, the system requires longer enrollment and authentication times, while a security disadvantage is the storing of portfolio images on the server as plain text.

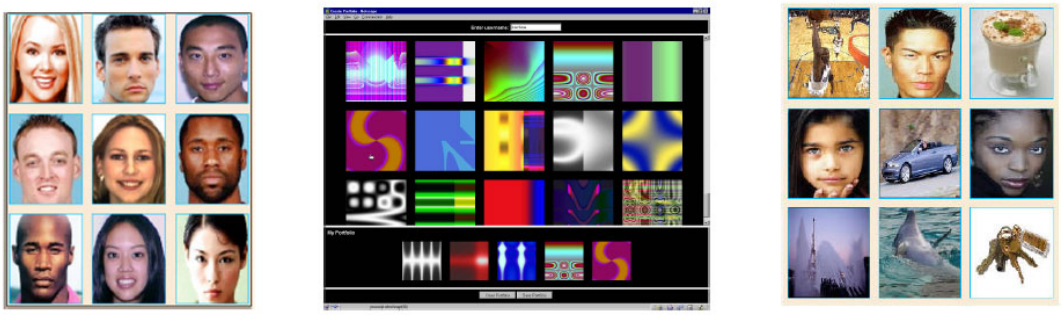


Figure 2.4: *Recognition-based authentication examples*

Regarding ubiquity, two recognition authentication schemes applicable in a mobile environment have been proposed. In (Jansen et al., 2003), a thematic image is divided into thumbnail segments and the user then selects a sequence of these segments as a graphical password. A numerical value is assigned to each segment, thus the selected sequence generates a numerical password. In a variation to this approach, the user can select their favorite image for authentication (Takada & Koike, 2003); however this increases the predictability of the password (Davis et al., 2004).

2.4.3 Hybrid Developments

Ongoing authentication projects include graphical password schemes based on image interpretation, where the password is derived from word associations based on a series of

inkblot images, like in psychological tests. In (Stubblefield & Simon, 2004) the authors believe that the entropy of this scheme is sufficiently high for practical use; nevertheless this has yet to be subjected to a detailed study. In practice, as shown in Figure 2.3., this study has been used by Microsoft to deploy an online application that uses a series of Rorschach-style inkblots to help users generate secure passwords. The user is required to go through a series of inkblot images, associate each image with a particular word and then use certain letters from the word to generate the password (McDougall, 2008). The word associations users come up with are being collected and stored for further research purposes.

Figure 2.5: *Rorschach-style inkblot*

Other developments include: graphical password techniques based on repeating a sequence of actions (Boroditsky, 2009) and a map authentication system that is based on navigating through a virtual world. There is also a proposal for a 3D authentication scheme where user actions in a virtual world, such as clicking areas, drawing, and interacting with the environment, are interpreted as their password (Alsulaiman & El Saddik, 2007).

The most current and thorough research overview of graphical authentication can be found in (Biddle et al., 2011). The authors describe the innovative features of selected schemes and discuss methodological issues in empirical evaluation. They also define the usability and security perspectives that need to be included in presentations of knowledge-based authentication schemes as they apply to graphical passwords. These perspectives include considerations for target users, domains, applications, password space, user password choice, password interference, vulnerability to attacks and user evaluation.

2.5 On the Security of Graphical Authentication

An authentication system must provide adequate security for its intended environment in order to meet its primary goal. The system should be evaluated against common attacks to determine if it satisfies security requirements. Attacks which exploit software vulnerabilities to bypass the authentication system entirely are not considered at this point as they are software-specific and depend on timely administration of the computer system. The types of attacks that can be undertaken for cracking authentication mechanisms can be categorized into two general categories:

- **Guessing attacks,** where the attackers can either make an exhaustive search through the entire password space, or create password dictionaries in order to get within a manageable range of guesses. Small password spaces and password selection patterns are identifiable vulnerabilities of authentication systems to these types of attacks.
- **Capture attacks,** where the password is obtained directly by either capturing login credentials during authentication or by getting the user to reveal the password. Shoulder-surfing, phishing and malware are common methods for capture attacks.

Standard attacks on text passwords can be easily converted to attacks on graphical password schemes. Existing graphical password systems usually have a strong performance against one type of attack, but remain vulnerable to the other types thus providing inadequate security in everyday environments.

2.5.1 Guessing Attacks

Guessing attacks can generally be subdivided into online guessing attacks and offline guessing attacks. Online guessing attacks require direct interaction with the system. Attackers enter password guesses sequentially in turn until the login is successful. Usual protection measures against such attacks are either CAPTCHAs or limiting the number of allowed incorrect login attempts before disabling the user account. However, limiting the number of incorrect attempts is a fallible solution as it decreases both the usability of the system, (the legitimate user who forgets their password is permanently locked out), and the security of the system (enables denial-of-service attacks which intentionally provide incorrect passwords to disable accounts). Other defenses consider user choice issues and include password rules or policies such as: disallowing weak passwords at creation, encouraging stronger password choices (Bicakci et al., 2009), and using both reactive and proactive password checkers (Bergadano et al., 1998). Offline guessing attacks are possible if the attackers gains access to verifiable text (Gong et al., 1993). Attackers don't need to continuously interact with the live system, trial guesses are processed faster, and pre-computed data structures or special hardware may be used to aid the cracking procedure. Defensive techniques against offline guessing attacks include processing password with a one-way hash function or using iterated hashing (Morris & Thompson, 1979) before storing. This increases the time required to process password candidates or use pre-computed dictionaries.

From another perspective, based on the guessing method, guessing attacks can be divided into: brute-force attacks and dictionary attacks.

Brute-force attacks try all elements in a search space and are usually used in offline password cracking. However, in practice a full search of a large password space is limited by the available processing power. As user-selected passwords are not equally probable along the complete password space, brute-force attacks are aided by exhaustive-search optimizations such as: coarse sequencing, guessing shorter passwords first, fine sequencing, ordering passwords in decreasing expected probability, Oechslin's rainbow tables (Oechslin, 2003), trading pre-computation time for storage, and favoring subsets (Oorschot & Thorpe, 2008) preprocessing higher probability passwords. To minimize the threat of brute-force attacks it is necessary to have a large theoretical password space or to use complementary mechanisms.

Dictionary attacks involve guessing passwords from a relatively short pre-compiled list (dictionary) of high-probability candidate password, based on empirical data or assumptions about previous user behavior. Smart dictionary attacks use algorithms to

combine time-memory trade-offs with higher success probabilities of prioritized dictionaries (Narayanan & Shmatikov, 2005). Dictionary attacks exploit skewed password distributions resulting from password subsets that are more attractive to users. Successful attacks on user-selected passwords are from predictable, relatively small subsets of the theoretical password space which can be enumerated, are small enough to search, and contain a significant fraction of passwords chosen in practice. A theoretical space too large to be exhaustively attacked does not guarantee security, if the effective password space is small. This is where many graphical password proposals fall to dictionary attacks due to predictable selection patterns in user choice.

2.5.2 Capture Attacks

Capture attacks intercept passwords by one of the following methods:

Shoulder-surfing refers to the gaining knowledge about users' credentials by either direct observation or by using recording devices (Roth et al., 2004). Shoulder-surfing is a real concern if the attackers targets a specific user and has access to the users' location. Most graphical passwords can be recovered from observing one successful login. With some graphical passwords it is necessary to observe multiple successful logins in order to deduce the full password as usually only a subset of user portfolio images is displayed during each login. Graphical schemes believed to be resistant or immune to shoulder-surfing have significant usability drawbacks (Komanduri & Hutchings, 2008). The reported drawbacks are concerned with the increased time and effort required to authenticate, which diminishes the usability of the proposed systems and makes them less suitable for everyday authentication.

Malware or malicious software includes any unauthorized software installed or downloaded without a user's informed consent. Malware code is silently installed on the users' computer and can capture different aspects of computer use. Keystroke-loggers record keyboard input, mouse-loggers capture mouse actions, and screen scrapers record screen activity. All of the captured data is usually sent remotely to a predefined virtual destination or made available for retrieval. Many graphical password schemes require a mouse-logger and/or a screen scraper in order to capture the authentication key, and a keystroke-logger to collect the username. If graphical passwords gain popularity, more specific malware will likely follow.

Phishing is a way of attempting to acquire sensitive information by masquerading as a trustworthy entity. In phishing attacks the user is tricked into entering their credentials at a fraudulent website, for example, by having the user follow a link in an email. Phishing attacks on graphical passwords usually require for the specific images to be presented to the user. Consequently, a graphical password phishing site needs to conduct earlier server probes to either collect the images or retrieve and relay information from the legitimate site through a man-in-the-middle (MITM) attack.

Social engineering refers to the direct manipulation of people into performing actions that divulge confidential information. Social engineering methods may require targeted background work which is often easier than otherwise breaking into a system (Mitnick & Simon, 2002). While text passwords are relatively easy to share, graphical passwords require a frame of reference in order to convey sufficient detail about the password.

2.6 Summary

There are many specific strategies and methods for attacking graphical authentication systems in existence. Since no system offers perfect security, authentication graphical authentication schemes must be designed and evaluated according to their prospective

vulnerabilities. Although for a particular attack strategy, it is possible to compare the susceptibility of different schemes, in practice, the likelihood of such attacks cannot be accurately predicted.

The following chapter presents the initial research required for the development of a graphical authentication mechanism that satisfies both usability and security metrics.

3 Preliminary Analysis of Graphical Authentication Concepts

The studies presented in this section are focused on conceptualizing and developing graphical authentication mechanism appropriate for ubiquitous environments. As users can only make informed choices when the proposals being discussed are meaningful to them, enabling users to envision and make sense of those proposals is an essential element of all approaches to system design.

In the first section the paper prototyping method is used to effectively simulate graphical authentication and distinguish the user preference between two system approaches in ubiquitous environments. Complementary, both a high fidelity paper and a mobile prototype is used in order to evaluate popular graphical authentication concepts which are based on separate cognitive functions: recall and recognition. This experiment uses a between group design where each participant evaluates a prototype for both authentication concepts in the same medium. The results of the study demonstrate that recognition-based graphical authentication mechanisms are more suitable for ubiquitous environments than recall-based systems. Hence, the purpose of the second experiment is to determine the most suitable image content that can be used in a recognition-based graphical authentication mechanism. To assess this claim the short-term and long-term memorability of images are analyzed based on three different types of content: abstract, face, and single-item. The findings show that single-item images are easier to remember and retain in memory that other content types.

3.1 Experiment 1: Prototyping Ubiquitous Image Authentication

3.1.1 Introduction

Paper prototyping is a widely used and validated technique for exploring, communicating, and evaluating early interface designs (Snyder, 2003). It is an established approach in designing interactive systems for the development of basic software versions that help users and designers to understand the possible alternatives (Ehn & Kyng, 1991; Preece et al., 2002). Conversely, prototypes are used to examine the aesthetics, content and interaction techniques from the perspective of the user.

By gathering data on user mistakes and comments, designers and usability professionals can trace usability problems at an early stage. In (Nielsen, 2003) the author states that paper prototyping allows for testing early design ideas at an extremely low cost, which allows for fixing usability problems before implementing something that doesn't work. This implies the many benefits of paper prototyping during the design process, such as (Rudd et al., 1996):

- externalizing design ideas with low investment,
- generating and testing numerous alternatives early in the design cycle,
- iterating a design many times prior to committing to an implementation, and
- focusing evaluation on macro-level issues such as major interface screens and overall interaction.

Alternatively, paper prototyping has some inherent limitations, such as the lack of complete realism in the interaction or understanding and revising dynamic behaviors (O'Neill et al., 1999). However, it is generally considered that these limitations are a worthwhile trade-off for the ability to explore numerous alternatives early in the design cycle.

The variance in fidelity between a prototype and the final product can be along several dimensions. They include:

- breadth of features,
- degree of functionality,
- similarity of interaction, and
- aesthetic refinement.

In (Virzi et al., 1996) the authors note that a prototype that compromises on one or more of these four dimensions in a way that is obvious to the user is a low-fidelity prototype. Hence, as described in (Sefelin et al., 2003), a low fidelity prototype is simple and built with the intent of visualizing design ideas at very early stages of the design process. Comparatively, high-fidelity prototypes are representations composed of predefined components and scripted interactions. Thus, without compromising the noted dimensional aspects and the increased investment needed for development, the users can directly interact with the prototype. Another implication in prototyping is the choice between paper and computer as the prototype medium, which has implications in the realism of the representation, the type of available usability testing methods and the ability of users to participate in the design process (Walker et al., 2002). When compared to a digital medium, paper obviously doesn't respond to either mouse or keyboard input, thus the differing mediums support different aspects of the interaction process.

In order to demonstrate the interaction process in low-fidelity prototyping tests, the examiner needs to create a pretend-environment that simulates a fully functional system. With a paper medium this includes manipulating sheets of paper in response to the user's behavior. Furthermore, paper prototypes allow for quick changes while exploring interactions by sacrificing some realism. On the other hand, computer prototyping requires a better defined interaction flow before user testing as it usually allows for pre-programmed responses to user behavior and remote recording of user interactions. However, high-fidelity computer prototypes may reduce design effectiveness as development tools limit the creative flow and slows down the prototyping process (Vaidyanathan et al., 1999). Low-fidelity computer prototyping require less skill and allow for quick changes by sacrificing the range of interaction techniques available in the prototype.

The need and effectiveness of low and high fidelity prototypes have been analyzed in several studies. (Virzi et al., 1996) compared usability problems discovered by low and high fidelity prototypes, and conclude that substantially the same sets of usability problems were found under both conditions. (Sefelin et al., 2003) also reached the same conclusion with the added knowledge that subjects prefer computer over paper prototypes. When considering ubiquitous environments, in (Liu & Khooshabeh, 2003) the authors determine that low fidelity paper prototyping is insufficient for supporting requirements, such as scalability, but a prototype with higher fidelity and automation levels can enhance the quality of interaction data available for evaluation.

Snyder (Snyder, 2003) describes how to prepare, conduct and interpret the results of paper prototyping. A common technique for evaluating paper prototypes is user evaluation (Rettig, 1994), where a user informally works through several controlled tasks

and the design team identifies where the prototype met or didn't meet user expectations. This can be complemented with other evaluation techniques such as cognitive walkthroughs (Polson et al., 1992) and heuristic evaluations (Nielsen & Molich, 1990). Most evaluation techniques utilize the Wizard of Oz approach (Dahlback et al., 1993), where when a user interacts with the interface, a facilitator changes the representation in the background to simulate an interaction with a fully functional application.

3.1.2 Mobile Prototyping and Graphical Authentication

Although paper prototyping is a reputable method for evaluating usability in early design iterations, the definite principles of paper prototyping for evaluating applications deployed in ubiquitous environments have not yet been established. Ubiquitous environments are dynamic and typically complex with an "on-the-go" context of use, hence usability problems are best discovered in situations representative of the real world (Kjeldskov & Stage, 2004).

Mobile prototyping studies are rarely performed due to the challenges inherent to mobile use (Bertini et al., 2006). In an on-the-street study (Hendry et al., 2005) the authors used a table PC-sized cardboard box to simulate screens of a mobile user interface. They reported a great difficulty in using this prototype in the field as prototype components were difficult to manage in such an environment. (Sa & Carrico, 2006) discovered that the fragility of just using cards and paper misled participants about the form factor of the final product and used card-holding wooden frames with dimensions approximating the target device. Extending these studies, a comparison of traditional paper prototype to pseudo-paper prototype used in a mobile contextually relevant lab-based protocol showed that the later allows participants to identify more unique usability problems (Lumsden & MacLean, 2008).

Reflecting on authentication mechanisms, there is no research of either paper or mobile prototyping during the development process of such mechanisms. Considering the specificity of ubiquitous environments and the most current research overview of usability and security in graphical authentication (Biddle et al., 2011), it is necessary to test graphical authentication concepts using both a high fidelity paper prototype and a high fidelity mobile prototype in order to encapsulate different aspects of the interaction process during evaluation.

The prototyped graphical authentication concepts evaluated in this experiment differ on the mental task necessary to memorize and reproduce the graphical password: recall and recognition. The recall-based graphical authentication concept, codenamed ImagO, has its roots in Blonder-style passwords (Blonder, 1995). This approach is common for many of the existing recall-based graphical authentication mechanisms such as Passpoints, CCP, PCCP, and even GrIDSure. In ImagO the user is presented with a large photographic image that contains a sizeable amount of distinct elements. The elements in question represent either artificial objects or living beings, each defined as a separate clickable region. To authenticate, the user clicks on the distinct elements in the image, thus entering the password (Figure 3.1.).

Figure 3.1: *ImagO concept example*

The recognition-based graphical authentication concept, codenamed ImagePass, is essentially based on VIP (De Angeli et al., 2005), but follows the general concepts of its sibling graphical authentication schemes such as Passfaces, DejaVu, Story, etc.. In ImagePass the user is asked to enter the graphical password by selecting a series of images representing objects from a selection grid which contains both password and decoy images (Figure 3.2.). If both the sequence and the selected images are correct the user is granted permission to access the system. The improved memorability and the suitability of photographic images for authentication have been demonstrated in (Mihajlov & Jerman-Blazic, 2011).

Figure 3.2: *ImagePass concept example*

3.1.3 Research Questions and Hypothesis

With the current advent of ubiquitous technologies and devices graphical passwords can be considered as a complementary replacement for traditional passwords in specific situations. The purpose of this experiment is to determine whether one graphical authentication concept is more suitable for ubiquitous environments than the other. As both usability and security need to always be considered when evaluating authentication mechanisms the following hypothesis have been formulated:

H1: There is a difference in users' perception of usability between recognition-based and recall-based graphical authentication mechanisms in ubiquitous environments.

H2: There is a difference in users' perception of security between recognition-based and recall-based graphical authentication mechanisms in ubiquitous environments.

As ubiquitous environments can differ significantly based on device and context of use, there are two reasons for using both a paper prototype and an interactive prototype in the evaluation: to gather more issues and to simulate different screen sizes and situation by using a larger paper prototype and a small interactive mobile prototype.

3.1.4 Participants and Equipment

This study investigates user behavior in ubiquitous environments where the selected participants belong in the generation X and generation Y cohorts. Hence, thirty-eight participants (20 male, 18 female, age range 17-36, average age 27.3) were recruited for this study. All of the participants had normal vision, varied touch-screen gadget experience, and used mobile Internet for professional or leisure purposes on a frequent basis.

The paper prototype versions of the graphical authentication mechanism were created with A4 paper. The interactive prototype versions utilized a Samsung Galaxy S touch screen mobile phone with Android 2.2 OS. All experiment sessions were conducted at the Laboratory for Open Systems and Networks at the Jozef Stefan Institute and the E-business Laboratory at the Ss. Cyril and Methodius University.

3.1.5 Experiment Design

For the purpose of this experiment, six large images with a substantial number of recognizable distinct elements were prepared for the ImagO prototypes and fifty small single-object images were prepared for the ImagePass prototype. For the paper prototype the images were printed in color on A4 sized paper. Single page color printouts were used for each image for prototyping the ImagO concept while the ImagePass concept used multiple color cut-outs that were managed by the examiner during the session as necessary. The paper prototype was laid out on a table with the examiner performing the role of the computer by moving parts of the interface in accordance with users' actions. In the mobile prototype a mock-up app was used which showed images with undefined click areas and a pre-programmed tap response regardless of where on the screen the participant decided to press.

In order to shift the participants focus to authentication as a secondary task and also to give context to the authentication process, in the initial version of the experiment three different authentication scenarios were developed: email login, forum login and system/phone unlock. However, this approach was discarded as the nature of the scenarios introduced a strong bias in the results. For example, during the email scenario the participants had a strong dislike for the authentication mechanism, as they immediately felt that the authentication mechanism was inappropriate for mail use. This approach could be feasible in future studies when the appropriate environment is determined for specific systems.

During the session the participants were asked to think aloud while the examiner occasionally focused the direction to their security concerns. All of the participants' comments were recorded with a digital voice recorder. The content of the collected comments was analyzed to specifically define the variables to be evaluated.

3.1.6 Procedure

The participants had to use a novel authentication mechanism in order to login to a system in a mobile environment. For each experiment session the participant was randomly assigned either to the paper or to the mobile prototype evaluation. They then proceeded to

evaluate both ImagO and ImagePass graphical authentication concepts. The order in which the concepts were presented and evaluated was alternated for every subsequent participant. All of the participants were required to work through two tasks, enrolment and authentication, with both graphical authentication concepts.

While prototyping ImagO, on both mediums, the participant was presented with three different images containing multiple elements and was asked to select the preferred image for authentication. If the participant felt a strong dislike for all the images the other three available images were presented as an alternative. Once an image was chosen the participant was then asked to select his authentication key by pointing or tapping on different image sections. There was no limit imposed on the authentication key length. In order to determine the expected area tolerance during password selection and the nature of the clicked areas, after successful enrolment the participant was asked a few follow-up questions to elaborate the selection. *Eg. If a person/animal was clicked the participant was asked whether it was expected that tapping different body segments should count towards the same selection or treated differently.* In ImagO the authentication task was essentially the same as enrolment as the user had to repeat the authentication key, by interacting with the same image.

In ImagePass different interface components were prepared as paper cut-outs for the paper prototype while interface screenshot images were used for the mobile prototype. During enrolment the participant was presented with a selection grid containing 30 images and was asked to choose a password by pointing/tapping a sequence of the available images. If the images were disliked, the participant could click the New Images button which in the paper medium prompted the examiner to change the selection grid image, while in the mobile medium a new screenshot image was loaded onto the screen. After the graphical password selection, the participant was asked for the reasons behind selecting particular images, while the second examiner prepared the authentication grid screenshot image containing selected and decoy images. The grid image was then printed and loaded into the mobile device before proceeding with the authentication task. During authentication the participant was asked to enter the previously chosen password by pointing/tapping the correct images in the authentication grid (Figure 3.3)

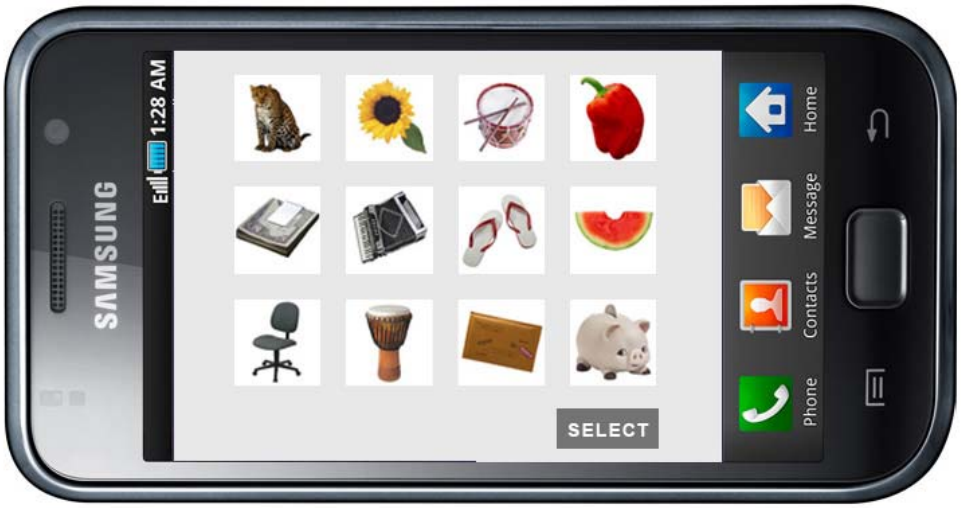


Figure 3.3: *Example for the ImagePass mobile prototype.*

During the think-aloud protocol, the participants didn't generally confer information

about security issues unless they were specifically lead to that direction of thought by the examiner. To initiate the process the examiner would ask a question such as: “Do you feel that the graphical password you have selected is secure?”, “Do you think that someone could compromise your graphical password?” or something similar. In addition, at the end of each prototyping session the examiner asked questions to clarify some observations made during the enrolment or authentication task.

3.1.7 Results and Discussion

Based on the audio recordings each session was transcribed and a content log of participants’ commentaries was created. The comments were aggregated and analyzed for each graphical authentication concept using the prototype medium as an independent variable. There was no significant difference for the number of comments between mobile and paper prototype either for the ImagO or for the ImagePass concept, thus a medium independent data analysis was conducted.

Qualitative analysis methods such as defect categories (Lindgaard, 1994) or mobile heuristics (Bertini et al., 2006) have not been specifically defined for authentication mechanisms. Hence, based on content analysis it was decided that comments will be coded in the following four categories:

- **usability issues,** comments that focus on usability deficiencies of the concepts;
- **security issues,** comments that focus on security deficiencies of the concepts;
- **positive remarks,** any affirmative expression for the concepts;
- **improvement recommendations,** comments that express constructive suggestions about possible system improvements.

Observing the positive remarks through usability and security perspectives was considered, however many of the comments failed to be classified in any of the categories. A summary of the data is presented in Table 3.1 and Figure 3.4.

Table 3.1: *Number of comments for both prototype concepts across different categories.*

	Usability			**Security**			**Positive remarks**			**Suggestions**			**Total**		
	N	**M**	**SD**	**N**	**M**	**SD**	**N**	**M**	**SD**	**N**	**M**	**SD**	**N**	**M**	**SD**
ImagO	188	4.95	2.31	217	5.71	2.89	61	1.61	0.91	93	2.45	1.29	543	14.29	4.37
ImagePass	128	3.37	1.74	144	3.79	2.59	45	1.18	0.63	56	1.47	0.89	389	10.24	3.26
Total	316	8.32	3.52	361	9.50	5.12	106	2.79	1.38	149	3.92	2.32	932	24.53	7.10

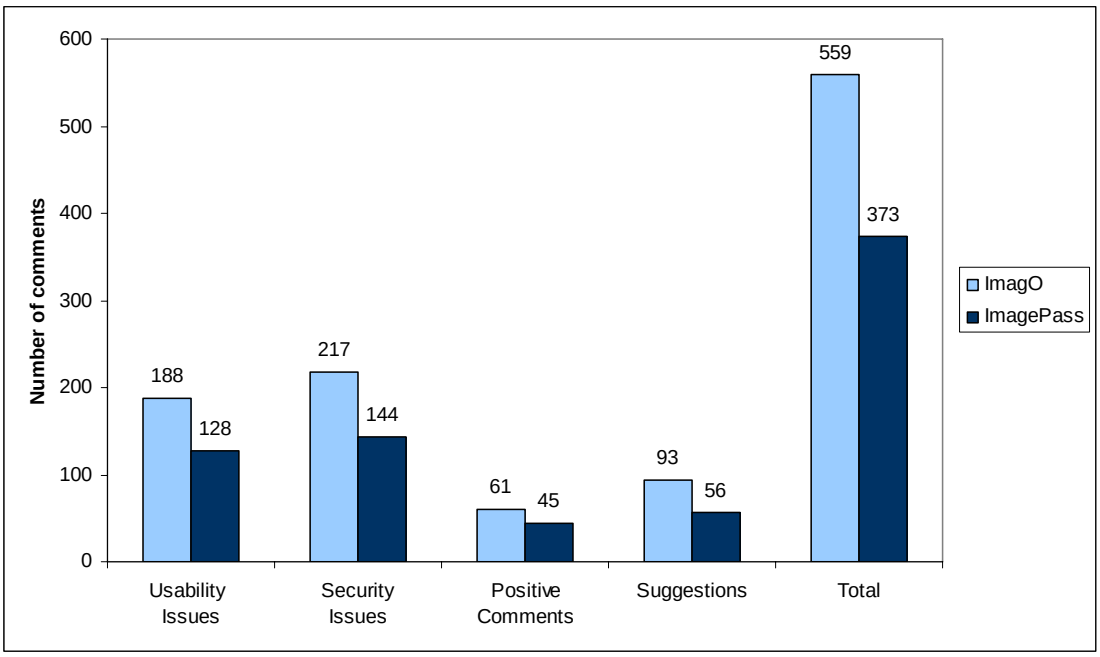


Figure 3.4: *Total number of comments grouped according to category and prototype.*

In general, the 38 participants made 932 comments (M=24.53, SD=7.1). On average the participants made more comments about the ImagO prototype, whereas a one-way ANOVA revealed a significant difference for the number of comments between graphical authentication concepts $F(1, 38) = 13.184$; $p=0.00 < 0.05$. These results partially support both hypotheses, showing that ImagO was generally commented more than ImagePass.

A total of 316 comments (M=8.32, SD=3.52) regarding usability were made by all the participants. A one-way ANOVA found a significant difference for the number of usability comments between graphical authentication concepts $F(1, 38) = 14.350$; $p=0.00 < 0.05$. The participants discussed usability more during the ImagO prototype (M=4.95, SD=2.31), than the ImagePass prototype (M=3.37, SD=1.74). Clustering the comments identified 24 unique usability issues, 14 issues for ImagO (M=0.37, SD=0.67) and 10 issues for ImagePass (M=0.26, SD=0.54). This supports the first hypothesis by showing a difference in the perception of usability in favor of ImagePass.

When discussing security, a total of 361 comments (M=9.50, SD=5.12) were made by all the participants. A one-way ANOVA using the number of security comments as the dependent variable found a significant difference between graphical authentication concepts $F(1, 38) = 11.297$; $p=0.00 < 0.05$. Security issues were discussed more during the ImagO prototype (M=5.71, SD=2.89), than the ImagePass prototype (M=3.79, SD=2.59). Clustering the comments revealed 17 unique security issues, 11 issues for ImagO (M=0.29, SD=0.52) and 6 issues for ImagePass (M=0.16, SD=0.77). This supports the second hypothesis by showing a difference in the perception of security in favor of ImagePass.

The participants expressed 106 positive comments about the concepts. A one-way ANOVA using the number of positive comments as the dependent variable discovered a significant difference between graphical authentication concepts $F(1, 38) = 10.791$; $p=0.01 < 0.05$. More positive comments were mentioned during the ImagePass prototype (M=1.61, SD=0.91), than the ImagO prototype (M=1.18, SD=0.63). As positive comments could not be strictly categorized into usability or security categories these results partially support both hypotheses by showing a positive prevalence for the ImagePass concept.

While working on the prototypes the participants made 149 suggestive comments (M=3.92, SD=2.32) on what they considered to be system improvements. A one-way ANOVA using the number of suggestive comments as the dependent variable discovered no significant difference between graphical authentication concepts, although more suggestions were made for the ImagO prototype (M=2.45, SD=1.29), than the ImagePass prototype (M=1.47, SD=0.89).

Generally, the participants were amused by both concepts for graphical authentication. Regarding medium specific unique usability issues, during the interactive prototype session while entering the graphical password the participants expected some sort of a system response to each selection for both concepts. Participants suggested either screen vibrations or briefly increasing the lighting of the selected area. The lighting proposition was prevalent in the ImagO concept suggested as adding an increased awareness to the user of the clickable area tolerance. This follows to the second medium specific unique usability issue, which was the difference in responses between the paper and the mobile prototype for the tolerance of the selection area. In ImagO, mobile prototype participants expected a larger tolerance and fewer regions when compared to participants working with the paper prototype. This is understandable as essentially the paper medium tentatively presented the user with a larger "screen size". In addition, during ImagePass enrolment most participants either thought that "the number of available images is too large for the available screen space," or "the size of the available images is too small", which in essence refers to the same issue of cluttered screen content. In both prototype concepts during graphical password entry the participants expected additional information as to what they have entered before submitting the password. On the subject of security there were no unique medium-specific issues. The main security concern of participants was shoulder-surfing, which was expressed strongly on almost every occasion during the paper prototyping sessions.

By prototyping both conceptual approaches to authentication it can be safely concluded that recognition-based systems have lower usability and security problems as well as a higher user preference. In addition, the results of this study passively support previous findings (Rudd et al., 1996; Liu & Khooshabeh, 2003; Lumsden & MacLean, 2008), confirming that although a higher fidelity prototyping medium does potentially discover more issues, it doesn't significantly improve the outcome of the process. Nevertheless, it could be suggested that for ubiquitous environments when only one type of prototype can be developed due to project constraints, an interactive prototype would maximize the comprehension of the developed system. As prototypes were compared for only two graphical authentication concepts it could be beneficial to repeat the evaluation for additional concepts in order to further generalize these findings.

Summarizing the results it is evident that there is support for both tested hypothesis. There is a difference in users' perception of both usability and security between recognition-based and recall-based graphical authentication mechanisms in ubiquitous environments. ImagePass, the recognition-based concept, prevails with fewer usability and security issues and was selected as the basis for the further development of a usable and secure graphical authentication mechanism. The next experiment presented in this chapter focuses on the memorability of images in order to determine the most memorable image content for recognition-based systems.

3.2 Experiment 2: Image Content

3.2.1 Introduction

A wide array of human memory properties can function as potential approaches to cognitive authentication. The human mind has a considerable capacity for imprinting complex experiences which can be recognized without effort. Such phenomena can be applied to secure authentication protocols providing a novel approach to usable security. In (Weinshall, 2004) the authors present a study of exploiting natural characteristics of the human memory and integrate those characteristics into cryptographic protocols. Furthermore, the authors define three cognitive phenomena useful for authentication: image recognition, referring to remembering photographic and non-photographic images (Cave, 1997), pseudo-word recognition, referring to remembering words and their variants (Tulving et al., 1982), and language recognition, referring to distinguishing grammatical from ungrammatical statements (Perruchet & Pacteau, 1990; Reber, 1967).

The human ability to recognize previous experiences is effortless. A person learns and has learned a great number of things and needs to reveal only a few in order to be uniquely identified. Unlike recognition, unaided recall such as providing a text-based password requires additional conscious effort and functions at a lower capacity. The advantage to using images over text is that they require little conscious awareness to be remembered while simultaneously are usually complex to describe.

3.2.2 Human Memory

Humans receive information from their senses which in the mind is interpreted in terms of previous experience. After interpretation the information passes from the sensory short-term storage to short-term memory (STM). If the information is exposed to further processing it will be encoded within the long-term memory (LTM) (Norman, 1969). This additional processing involves organizing the new information in relation to previously encoded information and categorizing the new material. Conclusively, the information is stored in LTM only if it has been understood and interpreted (Figure 3.5.).

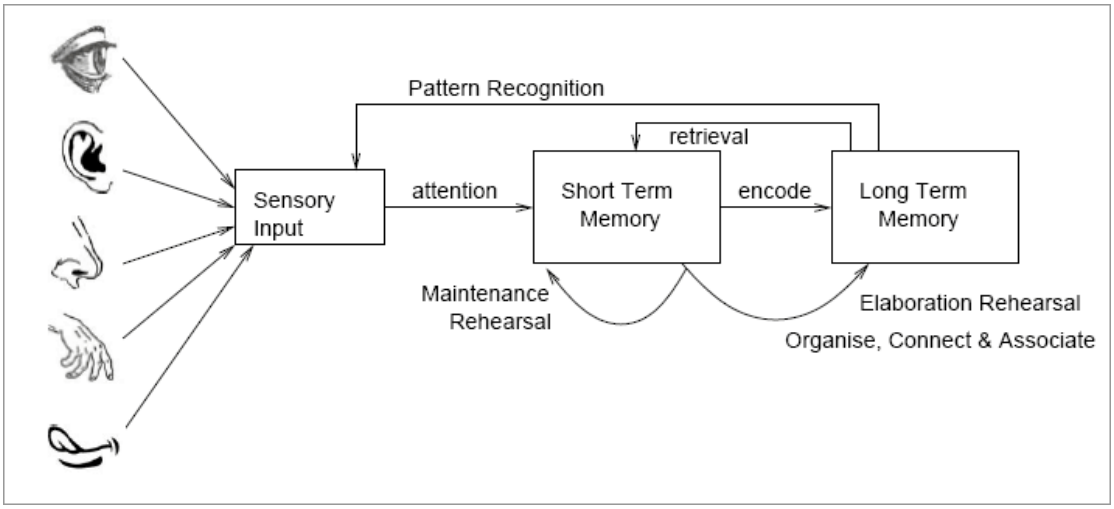


Figure 3.5: *Human information processing*

The encoding process is relevant for authentication as it is necessary to understand the circumstances which will support long-term recording and retrieval of authentication

keys. After a certain period of time, an authentication key can be retrieved only if it is stored in LTM, as information stored in STM dissipates when new information is presented to the senses. Information can be maintained in STM for longer only if the user engages in a specific activity to retain it (e.g. verbal repetition of a particular number), but it will not be sent to LTM store unless some meaning is attached to the information. Verbal repetition is referred to as maintenance rehearsal, and is essentially a cursory and superficial processing of the information. In order for the authentication key to be retrieved at a later time, it is necessary to invest some specific encoding effort, also known as elaboration rehearsal.

The cognitive activity involved in the authentication encoding process can be measured by the depth of processing necessary to store the information in LTM. The depth of processing at encoding time is determined by the amount of attention paid to the activity (Gregg, 1986). Therefore, superficial processing of a knowledge-based item leads to memory decay within minutes or hours. When the item is not meaningful and there is invested effort to learn the item, it will be encoded in memory and will be lost in a matter of days or months. Meaningful items are encoded in relation to previously-learnt or known items. The finest encoding results regarding memory retention occur when the user is able to come up with some scheme, such as mnemonics to encode an item that is not meaningful (Craik & Tulvig, 1975).

After encoding and memorizing the item the person has to retrieve the remembered item from memory at a later time. Retrieval is facilitated by means of establishing meaningful connections with previous structures in LTM. Meaningful retrieval can be measured by the mental effort the user has to apply to retrieve and deduce the authentication key. This is affected by whether the key is self-assigned or system-generated. Since users find it easier to recognize than to recall (Gutmann & Grigg, 2005), deducible recognition of the key would require less effort than cued recall, with the system optioning reminders, or uncued recall, where significant effort is spent. Uncued recall is the weakest type of retrieval mechanism, and becomes more difficult as people age (Parkin, 1993).

Retrieval is obviously dependent on the encoding process. When a new item can be encoded in LTM by linking it to a previously learnt items, the meaningfulness of the new item increases which in turn makes it easier to commit to memory. Deducible items are even easier to retrieve, since the user can reconstruct the activities at encoding time, and retrieve the item that way (Eysneck & Eysneck, 1979). A non-meaningful item can be learnt effectively if the person puts some effort into learning it, but it will probably decay within a month (Ebbinghaus, 1964). Essentially, once an item has been remembered it can be forgotten due to one of the following reasons:

- **Decay.** This happens when the item was not encoded specifically enough and the person cannot retrieve it.
- **Interference.** This happens when an item in memory interferes with another item. When an older item interferes with a newly-learnt item, it is called proactive interference, and when a new item interferes with a previously-learnt item, it is called retroactive interference (Erricsson & Kintsch, 1995). Similar items are most likely to interfere with one another.
- **Memory failure.** This happens when the person does not succeed in reproducing the item from memory.

3.2.3 Research Questions and Hypothesis

As different image types may affect the people's ability to remember and recognize graphical passwords, the main research question in this experiment was: "How does

image content affect memorability?" Most of the research on image memorability has been focused on comparing success rates of remembering images against words or numbers. Some studies have investigated the memorability of images for cued recall (Wiedenbeck et al., 2005a), but image content studies that would have implications for recognition based systems are scarce. In one of the few available study (Weinshall & Kirkpatrick, 2004) the authors evaluate the memorability of photographic images and discover that images with a clear central subject or action have greater memorability rates than other images. Additionally, they evaluated the memorability of images with single drawn objects, however their results were less successful. In Isola et al. (2011), the authors measure the subjective properties of photographs in order to extract memorability properties of images. They discover that the semantics of object content are important in determining whether an image will be remembered.

To determine the image type best suited for recognition-based graphical authentication, this experiment focused on a cognitive analysis of how users remember different images in relation to their content. For this purpose three categories of images were defined as follows:

- **abstract**, images with a non-specific shapes and patterns,
- **single-object**, photographic images representing a single object on a clear background, and
- **face**, photographic images of a single human face.

The purpose of the experiment was to determine whether particular images are more suitable for graphical authentication. As arbitrary content has less meaning for a user than semantic content (Norman, 1988) there is an expected decreased performance when remembering abstract images. The difference between single-object and face images was uncertain; hence the hypotheses were formulated as follows:

H3: There is a difference in short term memory image recognition success between abstract, single-object and face images.

H4: There is a difference in long term memory image recognition success between abstract, single-object and face images.

3.2.4 Participants and Equipment

Twenty-four Caucasian participants (11 male, 13 female, age range 24-38, average age 29.6) were recruited based on poster advertisements and direct contact. All of the participants had normal vision. As the experiment had a cognitive nature, the level of computer and Internet experience was not measured for any of the participants.

A 15.6" Dell Inspiron 6400 laptop with a screen resolution of 1600x900 was used to run the experiment in various casual environment settings.

3.2.5 Experiment Design

For the purpose of the experiment, a total of 305 royalty-free images: 97 abstract, 101 single-object and 107 face, with a resolution of 150x150 pixels were prepared and loaded into a database. The abstract images were aesthetically interesting, visually distinct with different color combinations and a varying level of complexity. The face images were a head shot of people distinguishable by race, gender, age, facial expression and hair style, while single-object images could be discerned by color and purpose.

An application that successively displayed a series of ten random images from a

specific category was developed with the Adobe Flash CS5 software. The application displayed an image in the center of a screen with a white background for 0.5 seconds. The screen remained blank for one second between image displays. After the ten-image series presentation was completed and following a ten second pause, a 6x5 grid containing a random display of the 10 previously shown images and an additional 20 decoy images from the same image category appeared on the screen. A sample grid is presented in Figure 3.6.

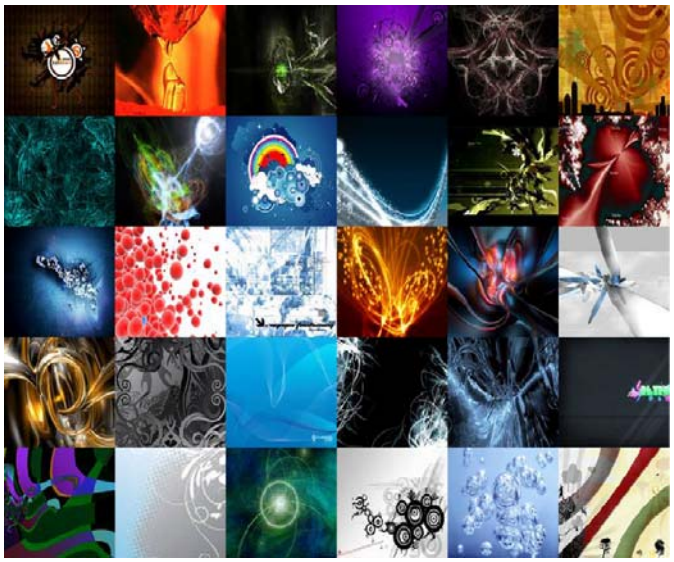


Figure 3.6: *Sample grid presenting abstract images from the experiment*

During the grid display, the participants had to physically point and vocally identify the recognized image; an approach derived from a trial run. When using a mouse to point and/or click to complete the recognition process, the time required to complete the process was slightly longer and participants mostly kept to themselves. With the physical interaction participants responded quicker and were more vocal in expressing comments during the session.

To capture the recognition process and the personal comments every session was recorded with a camera focused on the screen, while the examiner simultaneously kept notes for the whole process. The variables that were collected and evaluated for each image category were: number of correctly recognized images, number of incorrectly recognized images, time to complete the recognition process, personal comments and a 1-to-3 ranking scale based on the users' success perception for each session. At a later date, under the same setting the participants were presented with grids containing images from their respective experiment session. They were asked to recognize the images from the first session, with same data collected as in the first session.

3.2.6 Procedure

The experiment was divided in two phases. The first phase was performed in different casual environment settings over the course of two days. At each instance of the experiment, during the introduction, the participants were told that there would be three brief sessions of testing. It was explained that in every session they were going to be shown a series of images on a computer screen after which they had to recognize the shown images from a new larger set. The number of the images in the series, the existence of different image categories and the display dynamics of the application was

not disclosed.

Each session started with a blank white screen and the image display sequence was initiated when the participants expressed their preparedness to start the experiment. After the sequence, during the 10-second pause, the participants were informed that a large grid containing more images was going to appear on the screen and they have to point out the images that were previously seen. This procedure was subsequently repeated two more times for the other two image categories. After the image recognition was completed for all categories, the participants were asked to rank the sessions based on their perception of how well they performed.

To avoid biased results from the learning effect the experiment used a within-group design, dividing the participants into three 6-people groups. Each group was shown the series of images in a different order: group 1 (abstract, object, face), group 2 (face, abstract, object) and group 3 (object, face, abstract). Every session lasted between two and three minutes, thus precluding the possibility for the results to be offset by fatigue. All of the participants were asked to return on a later date to complete the experiment.

The second phase of the experiment took place approximately ten days after the initial session. The participants were told that they are going to be presented with an image grid containing images from their previous session and that they are going to have to identify the images they recognize. Every participant was presented with a grid containing the same images as the images in their respective session displayed in a different order. This process was then repeated two more times for the other image categories. To avoid biased results regarding category content, the image grids were displayed in a random order.

3.2.7 Results

The results in this section are presented separately for the short-term memory (STM) test and the long-term memory (LTM) test. The first results presented are from the analysis of the number of correctly and incorrectly recognized images, which is then followed by the analysis of completion time, rank and comments.

In the STM test, the recognition rate for object images was the highest at 93.3%, followed by abstract images, 72.9% and face images, 66.7%. Regarding error rate, object images had the lowest score at 0%, followed by abstract images, 5% and face images, 11.3%. A summary of the results is presented in Table 3.2.

Table 3.2: *Descriptive statistics for recognized images in the STM test*

Recognized	Type	M	SD	Min	Max
Correctly	Abstract	7.29	1.546	3	9
	Face	6.67	1.736	5	10
	Object	9.33	0.868	7	10
Incorrectly	Abstract	0.5	0.59	0	2
	Face	1.13	1.076	0	4
	Object	0	0	0	0

As the sample size is not large the Kolmogorov-Smirnov test (K-S test) was used to compare the scores in the sample to a normally distributed set of scores. The K-S test reveals that the dataset deviates significantly from a comparable normal distribution the number of correctly recognized images ($D(24) = 0.177$, $p<0.05$ for abstract, $D(24) = 0.248$, $p<0.05$ for face, $D(24) = 0.279$, $p<0.05$ for object). The K-S test also show a significant deviation for the number of incorrectly recognized images, $D(24) = 0.343$,

$p<0.05$ for abstract, $D(24) = 0.338$, $p<0.05$ for face, $D(24) = 0$, $p<0.05$ for object. A summary of the results for the K-S test is presented in Table 3.3.

Table 3.3: *Kolmogorov-Smirnov for normal distribution of recognized images in the STM test*

Recognized	Type	Statistic	df	Sig.
Correctly	Abstract	0.177	24	0.048
	Face	0.248	24	0.001
	Object	0.279	24	0.000
Incorrectly	Abstract	0.343	24	0.000
	Face	0.338	24	0.000
	Object	0	24	0.000

As the data is non-normally distributed, to test for difference between three related groups the Friedman's ANOVA test for significance was used. A summary of the results is presented in Table 3.4.

Table 3.4: *Results of Friedman's ANOVA for the STM test*

Recognized	Type	Rank	χ^2	Df	Sig.
Correctly	Abstract	1.71	31.69	2	0
	Face	1.46			
	Object	2.83			
Incorrectly	Abstract	2.04	28.1	2	0
	Face	2.56			
	Object	1.4			

The results of Friedman's ANOVA in the short-term memory test show that the number of correctly recognized images is significantly different between image types ($\chi^2(2) = 31.692$, $p=0.00<0.05$. It also reveals that the number of incorrectly recognized images is significantly different between image types (χ^2 $(2) = 28.107$, $p=0.00<0.05$. As there is a significant difference for both correctly and incorrectly recognized images it is necessary to do a non-parametric post-hoc analysis with a Wilcoxon signed-ranks test to correctly determine the ranking. There are three groups, so to cover all possibilities three comparisons are made: abstract-face, abstract-object and face-object. To accept something as significant the significance threshold is corrected by the number of comparisons, $p=0.5/n=0.05/3=0.0167$. A summary of the results is presented in Table 3.5 and Table 3.6. After the Wilcoxon-signed rank test the effect size r was calculated as $z/\sqrt{}$number of observations. As two conditions with 24 observations each are compared there are a total of 48 observations.

Table 3.5: *Wilcoxon signed-ranks test for all pair variables in the STM test*

		Face – Abstract			Object – Abstract			Object – Face		
Recognized		**N**	**Mean Rank**	**∑ of Ranks**	**N**	**Mean Rank**	**∑ of Ranks**	**N**	**Mean Rank**	**∑ of Ranks**
Correctly	Negative Ranks	11	9.09	100	0	0	0	0	0	0
	Positive Ranks	5	7.2	36	20	10.5	210	20	10.5	210
	Ties	8			4			4		
	Total	24			24			2		
Incorrectly	Negative Ranks	0	0.00	0.00	11	6.00	66.00	18	9.50	171.00
	Positive Ranks	9	5.00	45.00	0	0.00	0.00	0	0.00	0.00
	Ties	15			13			6		
	Total	24			24			24		

Table 3.6: *Significance of Wilcoxon signed-rank for the STM test*

Recognized		**Face – Abstract**	**Object - Abstract**	**Object - Face**
Correctly	Z	-1,695[a]	-3.958[b]	-3.948[b]
	Sig.	0.09	0	0
Incorrectly	Z	-2.810[b]	-3.207[a]	-3.898[a]
	Sig.	0.005	0.001	0

[a] based on positive ranks
[b] based on negative ranks

In the STM test there was no significant difference between the number of correctly recognized abstract images (M=7.29, SD=1.546) and the number of correctly recognized face images (M=6.67, SD=1.736), T=5, p= 0.09>0.0167, r=-0.24. However, the number of correctly recognized object images (M=9.33, SD=0.868) was significantly higher than the number of correctly recognized abstract images (M=7.29, SD=1.546), T=0, p= 0.000<0.0167, r=-0.57. Also, the number of correctly recognized object images (M=9.33, SD=0.868) was significantly higher than the number of correctly recognized face images (M=6.67, SD=1.736), T=0, p= 0.000<0.0167, r=-0.57.

For incorrectly recognized images the Wilcoxon signed-rank test revealed a significant difference in all comparisons. The difference was significantly lower for abstract than face images, T=0, p= 0.005>0.0167, r=-0.41, significantly lower for object than abstract images, T=0, p= 0.001<0.0167, r=-0.46 and significantly lower for object than face images, T=0, p= 0.000<0.0167, r=-0.56.

The LTM test showed results similar to the STM test. The recognition rate for object images was the highest at 72.9.3%, followed by abstract images, 60.4% and face images, 51.1%. Regarding error rate, object images had the lowest score at 0.8%, followed by abstract images, 5.8% and face images, 7.1%. A summary of the results is presented in Table 3.7.

Table 3.7: *Descriptive statistics for recognized images in the LTM test*

Recognized	Type	M	SD	Min	Max
Correctly	Abstract	6.04	1.367	2	8
	Face	5.21	1.285	4	8
	Object	7.29	0.69	6	8
Incorrectly	Abstract	0.58	0.584	0	2
	Face	0.71	1.122	0	4
	Object	0.08	0.282	0	1

The K-S test showed a significant deviation from a comparable normal distribution for the number of correctly recognized images ($D(24) = 0.238$, $p<0.05$ for abstract, $D(24) = 0.231$, $p<0.05$ for face, $D(24) = 0.264$, $p<0.05$ for object), and a significant deviation from a comparable normal distribution for the number of incorrectly recognized images ($D(24) = 0.304$, $p<0.05$ for abstract, $D(24) = 0.314$, $p<0.05$ for face, $D(24) = 0.533$, $p<0.05$ for object). As the data was again non-normally distributed the Friedman's ANOVA test was used to examine the significance. A summary of the results from both tests is presented in Table 3.8 and Table 3.9.

Table 3.8: *Kolmogorov-Smirnov for normal distribution of recognized images in the LTM test*

Recognized	Type	Statistic	Sig.
Correctly	Abstract	0.238	0.001
	Face	0.231	0.002
	Object	0.264	0.000
Incorrectly	Abstract	0.304	0.000
	Face	0.314	0.000
	Object	0.533	0.000

Table 3.9: *Results of Friedman's ANOVA for the LTM test*

Recognized	Type	Rank	χ^2	df	Sig.
Correctly	Abstract	1.9	21.42	2	0
	Face	1.44			
	Object	2.67			
Incorrectly	Abstract	2.23	12.04	2	0
	Face	2.19			
	Object	1.58			

The results of Friedman's ANOVA in the long-term memory test show that the number of correctly recognized images is significantly different between image types ($\chi^2(2) = 21.522$, $p=0.00<0.05$) and that the number of incorrectly recognized images is also significantly different between image types (χ^2 $(2) = 12.040$, $p=0.02<0.05$). As with the STM test, this was followed by a non-parametric Wilcoxon signed-group test for post-hoc analysis correcting the significance threshold by the number of comparisons ($p=0.0167$). The results are presented in Table 3.10 and Table 3.11 respectively.

Table 3.10: *Wilcoxon signed-ranks test for all pair variables in LTM test*

		Face - Abstract			Object – Abstract			Object - Face		
		N	**Mean Rank**	**∑ of Ranks**	**N**	**Mean Rank**	**∑ of Ranks**	**N**	**Mean Rank**	**∑ of Ranks**
Correctly recognized	Negative Ranks	12	8.67	104	4	6.5	26	2	3.5	7
	Positive Ranks	3	5.33	16	18	12.61	227	20	12.3	246
	Ties	9			2			2		
	Total	24			24			24		
Incorrectly recognized	Negative Ranks	7	6.86	48	11	6	66	9	5	45
	Positive Ranks	7	8.14	57	0	0	0	0	0	0
	Ties	10			13			15		
	Total	24			24			24		

Table 3.11: *Significance of Wilcoxon signed-rank test for LTM test*

Recognized		**Face - Abstract**	**Object - Abstract**	**Object – Face**
Correctly	Z	-2,532[a]	-3.333[b]	-3.965[b]
	Sig.	0.011	0.001	0
Incorrectly	Z	-0.299[b]	-3.207[a]	-2.810[a]
	Sig.	0.765	0.001	0.005

[a] based on positive ranks
[b] based on negative ranks

The Wilcoxon signed-rank test for correctly recognized images revealed a significant difference in all comparisons. The difference was significantly higher for abstract (M=6.04, SD=1.367) than face images (M=5.21, SD=1.285), T=3, p= 0.011<0.0167, r=-0.37, significantly higher for object (M=7.29, SD=0.690) than abstract images (M=6.04, SD=1.367), T=4, p= 0.001<0.0167, r=-0.48 and significantly higher for object (M=7.29, SD=0.690) than face images (M=5.21, SD=1.285), T=2, p= 0.000<0.0167, r=-0.57. For incorrectly recognized images there was no significant difference between abstract and face images, T=7, p= 0.765>0.0167, r=-0.04. However, the difference was significantly lower for object than abstract images, T=0, p= 0.001<0.0167, r=-0.46 and significantly lower for object than face images, T=0, p= 0.005<0.0167, r=-0.41.

Completion time and rank have a normal distribution. A one-way ANOVA reveals no significant difference for the time to complete the recognition process between image categories.

Regarding the users' perception on how well they performed the test, a one-way ANOVA shows a significant difference for rank between images. A post-hoc analysis Tukey's HSD test revealed that there is a significant difference in how users perceive the images between single object and abstract images (p<0.01), as well as between single-object and face images (p<0.01). There was no significant difference between abstract and face images (p>0.05).

As for the personal comments during the image recognition session, it was noticeable that most participants were reserved about their ability to remember faces. However it is

worth to note that according to (Baron, 1981; Going & Read, 1974) facial recognition is an innate human ability despite the declared perceptions of subjects. Some example comments included: “Remembering faces is not my strong point.”, “This session is hard. I don’t remember faces well.”, etc. The participants expressed apathy and disinterest toward the abstract images, for example, “Dull pictures.”, “They all look the same.”, “These images are so confusing.”. For single-object images the participants had a mildly positive attitude with several comments focused on the image content like “Simple, but different.”, “Hmm… food.”

The presented results partially confirm Hypothesis 3 as there is a difference in image recognition success between single-object and abstract images and single-object and face images, but there is no difference between face and abstract images. Regarding cognitive effort, the results show that single-object images are easier to recognize. In addition, unlike face and abstract images, single object-images are difficult to confuse. There were no incorrect guesses for this category and on average participants couldn’t remember only 1.125 of the 10 single-object images. The participants’ ranking of the categories and their personal comments also support the primacy of the single-object image category over face and abstract images. Abstract and face images had similar recognition rates, however, surprisingly, abstract images had a lower recognition error rate than face images which would imply they are easier to remember.

Hypothesis 4 is confirmed more clearly as there is a difference in image recognition success between all image types. The results show that single-object images are retained in long-term memory better than abstract images which in turn are retained better than face images. Hence, it could be concluded that images with content represented by a single-object have a significantly higher memorability. Therefore, this type of content will be used further in the development of the proposed graphical authentication mechanism.

3.3 Summary

As the initial intent was to determine the graphical authentication concept best suited for ubiquitous environment, the first study presented in this chapter evaluated two such concepts through a high fidelity prototype test on paper and a limited high-fidelity test on a mobile device. The evaluation used a between group design, where each participant evaluated the both concepts with a random approach to the medium of the prototype. The findings of the experiment illustrate that recognition-based graphical authentication has a greater potential for ubiquity than recall-based systems. The second study focused on distinguishing the memorability of images based on image content. Both a short-term memory and a long-term memory test were performed on three different types of image content: abstract, face and single object. The results of the study show that single-object image content is more memorable than abstract of face image content.

The following chapter of the dissertation presents the design and development of the ImagePass graphical authentication mechanism. In further chapters, this mechanism will be refined through usability and security evaluations of different aspects of the system. In addition, these studies will reveal indications for the potential usefulness of ubiquitous graphical authentication.

4 Designing a Recognition-based Graphical Authentication Mechanism

Current studies show that while dominant on the usability side, because of their intuitiveness and potential password space, graphical authentication mechanisms are not necessarily more secure. Also, the necessary training and user support for this approach are rather costly. Furthermore, it is difficult to get users to abandon the traditional text-password approach for something unfamiliar, therefore limiting graphical passwords to isolated environments rather than mass market solutions.

Considering the conducted literature on authentication mechanisms and the specificity of the World Wide Web, two graphical authentication schemes seemed feasible for authentication mechanisms that would also include a web environment. Therefore, in the previous chapter two concepts codenamed ImagePass and ImagO were developed for further consideration. In the ImagO concept the user was presented with a large image which contained a sizeable amount of distinct elements. The elements represented either artificial objects or living beings, each defined as a separate clickable region. To authenticate, the user had to click on the distinct elements in the image, to enter the graphical password. In ImagePass the user is asked to enter the graphical password by clicking on a series of recognized images representing objects. The designs for both concepts were developed and evaluated through a paper prototyping experiment. As the ImagePass concept received a considerably more favorable response, it was further defined and developed in more detail.

This chapter focuses on designing a prototype for a graphical authentication mechanism based on image recognition that can be deployed in a web environment and is supported by ubiquitous devices. It presents the conceptualization of the ImagePass system that follows usable security guidelines in the development process.

4.1 Defining the ImagePass System

As mentioned previously, ImagePass uses the concept of graphical password as an authentication key. The user enters the graphical password by clicking on a series of images representing objects. The object images are presented in a 4x3 grid which contains both images from the graphical password and decoy images (Figure 4.1.). If both the sequence and the clicked images are correct the user is granted permission to access the system.

Figure 4.1: *ImagePass grid*

As it will be displayed in the remainder of this chapter, ImagePass presents a novel approach in image recognition designed by employing usable security principles in the system development process. At first, the phases of the authentication process will be explained in the next section.

4.1.1 Enrollment

The enrollment to the ImagePass system is a simple and straightforward procedure consisting of three phases: username choice, graphical password selection and graphical password confirmation. In username selection, the user enters a preferred username in the Username text-field. The system then checks the availability of the desired choice. If the requested username is available, the user is taken to the graphical password selection screen. If the requested username is not available, the user is prompted to make a different selection (Figure 4.2.).

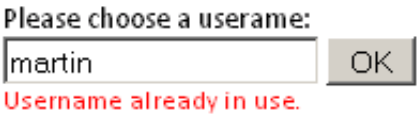


Figure 4.2: *Username selection*

In the graphical password selection screen, the user selects the authentication key from a given set of images. This screen contains an information section that briefly explains the graphical password selection process, a graphical password-selection grid, a selected graphical password panel and function buttons (Figure 4.3).

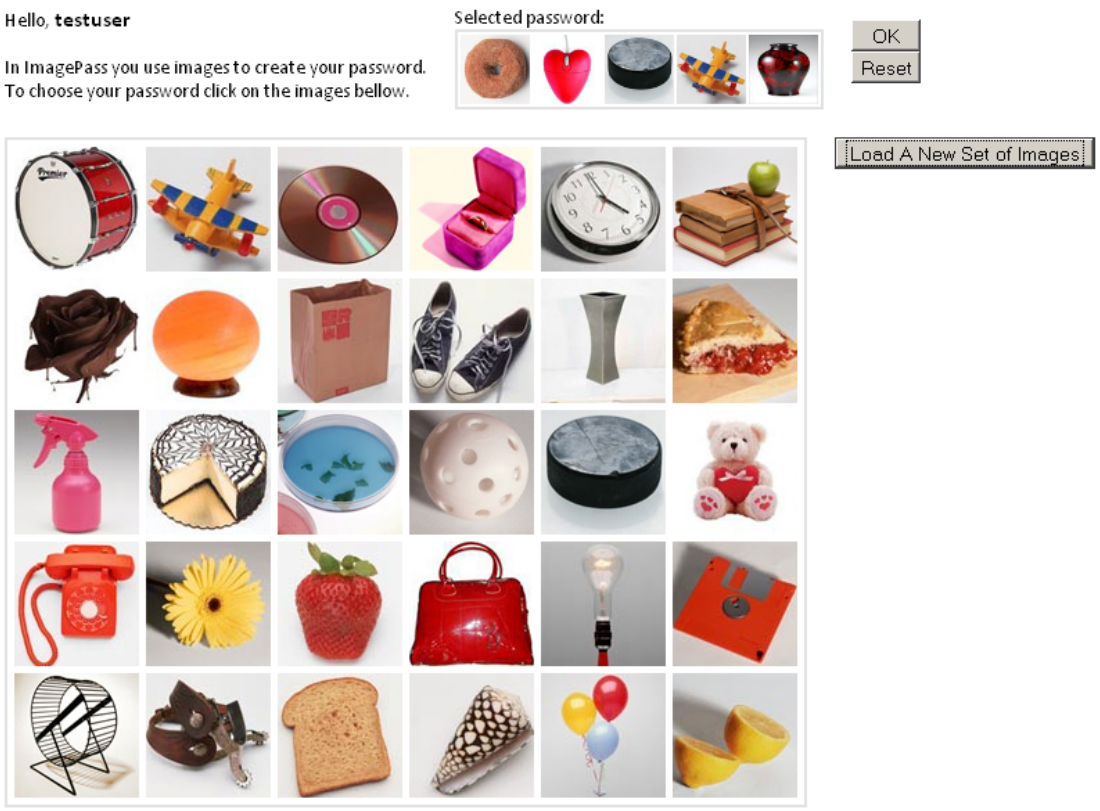


Figure 4.3: *Graphical password selection*

The images in ImagePass are color images, sized at 90x90 pixels, with the content always a single object on a light background. There is a large updatable image database that supports the graphical password selection process. However, for convenience, during graphical password selection the user is presented with a selection grid containing a random subset of 30 images. Showing all the available images would be futile as it is highly unlikely that anyone would consider all the possibilities. In the case where the user is not satisfied with the available image choice, clicking the Load a new set of images button will load a different random subset of images.

To select a graphical password, the user clicks on x number of images in a particular order. As the images are clicked, the Selected Password panel, placed on the top-right of the screen, displays the sequence of the selected images. If not satisfied with the selection the Selected Password panel allows the user to discard a particular image by moving over the image with the mouse cursor and then clicking on the Remove Image icon that appears in the top-right corner of the image (Figure 4.4.). The user can also Reset the selection and start the graphical password selection process from the beginning. Once the graphical password has been selected, the user has to go through the graphical password confirmation phase before the enrollment process can be completed.

Figure 4.4: *Current Selection panel*

Due to the specificity of the system there is a limitation imposed on the maximum number of different images that the graphical password can contain. The limit is set at twelve, as this is the maximum number of images that can appear in the authentication grid. As confirmed by the experiments presented in the following chapter, the expected user behavior is the selection of fewer than twelve images for a graphical password. Regardless of the selected graphical password length, a fixed set of twelve specific images containing both user-selected real images and system-selected decoy images are permanently attached to the username. For example, if in the enrollment process the user selects 5 different images for the graphical password sequence, the system will randomly assign 7 more images which will become a permanent part of the users' authentication imageset. If the decoy images were to change at every authentication session it would be very easy to learn the graphical password content just by eliminating the non-repeating images through several attempts.

The graphical password confirmation screen is a simulation of actual authentication. The screen is similarly divided as the Select Graphical Password screen with an information section that briefly explains the confirmation process, a graphical password selection grid, a selected password panel and function buttons. The selection grid contains 12 real and decoy images out of which the user has to select the correct graphical sequence. Initially, the Selected Password panel is inactive, but the user can switch it on by clicking on it. In case the user cannot remember the graphical password there is a Show Password button that will briefly display the graphical password in the same panel.

The enrollment process is completed after only one successful repetition as this is the common practice for text passwords. Requiring the user to reenter the graphical password several times in order to better memorize the sequence and/or to better learn how to use the system in essence decreases the practicality of the system and limits the environments into which it can be deployed.

4.1.2 Authentication

For the authentication process to begin, the user is first taken to the login screen (Figure 4.5.). The login screen contains a username text-field where the user enters a valid username and an OK button that confirms the username entry. If an invalid username is entered, the information "User does not exist" appears in red bellow the textbox. When a valid username is entered, clicking the OK button will take the user to the authentication screen.

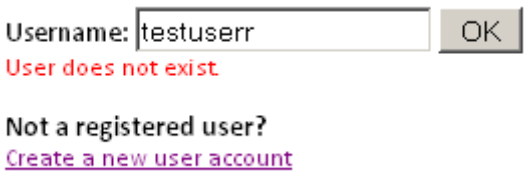


Figure 4.5: *Login screen*

The authentication screen is almost identical to the confirmation screen in the enrollment process, without the Show Password button. The information section instructs the user how to select the graphical password from the 4x3 selection grid with the Selected Password panel inactive unless clicked. The selection grid contains the same real and decoy images that were assigned to the username during the enrollment process which are preloaded before the page content appears. As explained previously, the images in the selection grid are unique for every username and are system-determined during

graphical password selection. During authentication, the user is always shown the same set of images, but the image positions in the display grid are randomly permuted in order to slightly diminish the possibilities for positional shoulder surfing (Figure 4.6.).

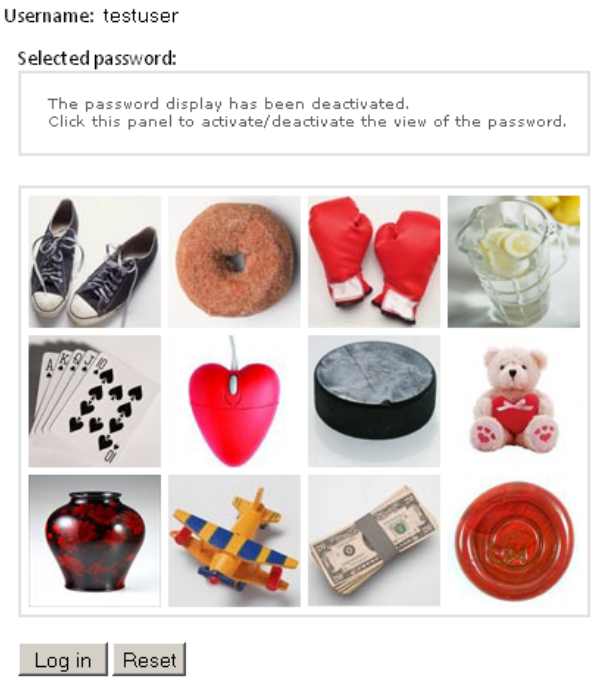


Figure 4.6: *Authentication screen*

Clicking the correct sequence of images successfully enters the graphical password and the user is granted access to the system. If a mistake is made, or the graphical password is invalid, the user is allowed to repeat the authentication procedure. There is no limit on the number of allowed unsuccessful logins; however, there is a system change in the authentication process taking place after every 5 unsuccessful attempts; this is explained in more detail later in this chapter in the section describing security.

4.1.3 Key Replacement

For small scale Intranet deployments ImagePass uses physical key replacement. This means that if a user forgets the graphical password currently there is a need to personally contact an administrator in order to change the authentication key. For a wider deployment the key replacement must be based on email identity confirmation as it would be the only cost-effective solution.

4.2 Designing for Security

During the conceptualization and design of ImagePass, many of the security issues relating to usable security qualities that were discussed in Chapter 2 were specifically addressed.

4.2.1 Predictability of ImagePass

The predictability of the authentication key revolves around one basic principle: creating a key that is unguessable within a reasonable amount of time. It is a large issue as confirmed by the plethora of password choice recommendations in professional and

popular literature (Burnett, 2005). The difficulty of intentional unpredictability arises with the poor concept of randomness that human beings posses. Randomness in itself is a strange concept that is difficult to define and is most closely associated with the absence of order. Nevertheless, to determine if a sequence is random it is necessary to observe several properties. Even distribution, distinctness and uniqueness are parameters that in essence determine the predictability of the authentication key (Renaud, 2004).

- **Even distribution.** The equal probability of distribution over the entire set of data. Most passwords are clustered together in similar groups.
- **Distinctness.** The nonexistence of relationship with previous or subsequent data. Using similar passwords with different systems decreases distinctness.
- **Uniqueness.** The inability to randomly produce the same sequence of data. The uniqueness of the key increases with the key length.

In adherence to these factors the image selection concepts and processes in ImagePass fulfill the following claims regarding the randomness and predictability of the authentication key:

1. When more ImagePass systems are used in different environments to prevent authentication key relationship between systems, it is sufficient to utilize different images on different systems. This supports the distinctness of the graphical password.
2. No two users will select and share the same graphical password, while some might just click the same image multiple times when selecting their authentication key. This supports the uniqueness of the graphical password.
3. When changing the graphical password, no user could use the exact same password as different sets of images will be offered during graphical password selection. The effort invested into choosing the same graphical password is greater than choosing a new graphical password, as the user will have to repeatedly load new sets of images until he discovers the same key units. This supports the uniqueness of the graphical password.

While the first claim is rather obvious, the second and third claims are tested in the usability study presented in the following chapter. The issue regarding even distribution is also analyzed in a separate experiment as evident from data analysis of specific system logs and presented in the next chapter.

4.2.2 Graphical Password Space

ImagePass has a varied authentication key space. During enrollment the user is presented with 30 images, a selection which can be expanded up to the number of images in the database. However it can be expected that most users will make their graphical password selection from the given subset. During authentication the display grid is smaller, with only 12 images, thus decreasing the available choices. As the system allows repetitive use of the same image and it requires the correct sequence of images for authentication, the graphical password space can be calculated with a permutations formula. The resulting data is presented in Table 4.1. The graphical password space at enrollment has been calculated by considering only the initial selection grid.

Table 4.1: *Graphical password space comparison*

Password length	Minimal password space at enrollment	Password space at authentication
3	27000	1728
4	810000	20736
5	24300000	248832
6	729000000	2985984
7	21870000000	35831808

On a first glance, the minor authentication grid potentially increases the susceptibility of the system to a cracking attack. This has been considered with a design countermeasure presented later in this section.

4.2.3 Increasing User Privacy

In limited Intranet environment the only personal data required by ImagePass is the full name of the user as it is used during key replacement. In a web environment, ImagePass will require the users' email address in order to support the key replacement process.

In all authentication mechanisms the identity of the user is defined by the username. Since most systems display the username actively and publicly, the identity of the user is partially compromised even before the authentication process begins. ImagePass introduces an option which makes a distinction between actual and displayed username. Actual username is the username that the specific user selects during enrollment. This username is not disclosed to third parties and is used only during authentication. Displayed username is the username displayed as a representation of the authenticated entity in all online communications. It is the username that becomes knowledgeable to other users of the system (e.g. forums, social networks) or users that have been contacted (e.g. email). Besides increasing the privacy of the user, this feature drastically improves the security of the system, as the potential attacker becomes unaware of the actual username that can be used to initiate the identity comparison.

4.2.4 Dealing with Attacks

The largest concern for graphical passwords is guessing attacks, as usable graphical passwords usually have a lower password space than traditional text passwords. Enhancing this password space by increasing the required password length is not advised as it affects the memorability of the authentication key and consequently the usability of the system (Stobert et al., 2010).

One of the proposed solutions for dealing with guessing attacks is the deployment of multifactor authentication. The most common strategy is to use two factors for the authentication, which combines two system approaches where the strength of each system counterbalances for the other's weakness. The main security feature in ImagePass that prevents most guessing attacks is adding another security layer that is based on one-time passwords. One-time passwords (OTPs) are used to create a one-time authentication code during system registration (Piazzalunga, 2007). In OTP systems the user has to own a OTP token, such as a smart card, a USB token or a mobile phone, which is used to connect to a service provider that will display the authentication data on the device’s display (Wu et al., 2003; Shelfer & Procaccino, 2002). The user needs to enter the given input manually, before the token expires, to gain access to a particular system. The authentication data changes for each new, user-authentication session. With the additional

authentication factor in ImagePass the graphical password can never be revealed at the system level. This greatly increases the immunity of the authentication mechanism to guessing attacks, dictionary brute-force attacks, both automatic and physical and transmission sniffers that collect text-based passwords. The strategic approach to implementing this feature is as follows.

Before the images from the user-specific imageset are sent to the browser, the system assigns a random number to each image. The random number and the corresponding image are stored as a temporary record in the database, which is valid only for the duration of the authentication session. In addition, as the images are displayed via HTML code, their filenames are hashed with a random key. For example, let the users' graphical password consist of images with the following sequential content: apple, orange, cat, house and cake. As each image is saved in the database under a unique numerical identifier, let us assume that the images from the graphical password are stored as follows: apple (ID=1), orange (ID=2), cat (ID=3), house (ID=4) and cake (ID=5). Clicking the correct image sequence essentially sends the numbers 1, 2, 3, 4 and 5 to the system to authenticate the user. If this is left as such, it would make the system extremely vulnerable to brute-force attacks, with everyday computers being able to crack numbered keys in minutes. With the preventive OTP measure, before sending the images to the browser, the system assigns random numbers to the images as temporary records for the duration of the authentication session, for example, apple (AuthID=356), orange (AuthID=34), cat (AuthID=12735), house (AuthID=956) and cake (AuthID=11). Consequently, clicking the right sequence would send 356, 34, 12735, 956 and 11 to the system which would then be subsequently interpreted as 1, 2, 3, 4, and 5, when compared to the temporary records. The example is presented in Table 4.2.

Table 4.2: *Sample table with random ID's*

ID	Image	AuthenticationID
1	Apple	356
2	Orange	34
3	Cat	12735
4	House	956
5	Cake	11

If the graphical password is entered incorrectly 5 times, the system randomly changes the assigned numbers to the images. When the graphical password is entered correctly, the temporary records are erased. This effectively prevents an automated brute-force attack. The programmed cracking application cannot discover the authentication key if the numbers keep changing and it is being reset after every five attempts. Even traditional transmission sniffers are redundant as they would record a sequence that is not valid for any subsequent attempts. Only a custom transmission sniffer which would successfully parse all the incoming images can be successful for infiltrating the system.

Since the images are displayed in XHTML code, besides the ID number that is sent back as a part of a graphical password sequence, there is also input on the displayed image filename. The normal source code for the image should look something like this:

<img sac="images\apple.jpg" id=1">

Nevertheless, to prevent matching of image filenames with generated ID, the filenames are hashed with a random key, hence the resulting XHTML code would be similar to:

<img sac="ejxhy.jpg" id="356">

To prevent a physical brute-force attack, where the attacker tries to physically click

every possible combination of images, the image positions in the grid are randomly switched after 5 failed attempts. The attacker could not make a structured attack unless he physically records all image clicks in all attempts. This significantly increases the cognitive effort of the attacker as it becomes necessary to keep track of mixing image sequences.

Regarding shoulder-surfing risks, ImagePass shares the same advantages and disadvantages as other recognition-based systems, higher vulnerability when using a mouse, and lower vulnerability when using a keypad (Tari et al., 2006). Like other graphical authentication mechanisms, the system has a high immunity to malware, unless the spying software captures the screen and mouse clicks simultaneously. Regarding social engineering, ImagePass graphical passwords are more vulnerable; due to their simplistic nature they are easy to describe and share.

4.2.5 Designing for Usability

During the conceptualization and design of ImagePass most of the usability issues relating to usable security qualities were considered and addressed. The basic idea behind using images rather than text as a foundation for authentication is the lower cognitive processing requirement and higher meaningfulness of images over text (Shepard, 1967). In addition, users find it easier to recognize than to recall (Gutmann & Grigg, 2005), therefore the ImagePass prototype has been conceptualized with recognition-based precepts. From the experiment presented in the previous chapter it was determined that users find it easier to memorize content that is familiar, simple and straightforward in order to minimize the cognitive effort required for recognition, therefore single-object images are used as the authentication key.

If graphical authentication mechanisms are intended to eventually replace text-based passwords the usability advantages and performance levels have to be conserved where possible. During enrolment, text-based passwords require only one password confirmation; consequently, in ImagePass the password is considered as being confirmed if it is successfully re-entered only once. By evaluating the graphical authentication mechanism under the same conditions as text-based passwords it is possible to get comparable usability results.

All graphical authentication mechanisms are generally highly inaccessible for visually impaired users. In ImagePass, the constant repositioning of the images in the graphical password-selection grid further affects the accessibility of the system. Nevertheless, one of the advantages of using simple object images for authentication is that they can be described in simple words. Implementation in further system iteration such as, simple image tagging with familiar words that quickly describe each image (for example, trumpet, red apple, autumn leaf) could partially alleviate this accessibility issue.

The described ImagePass prototype is intended for web use; however, the increased internet access in ubiquitous environments implies that design considerations have to be made for mobile devices. Hence, there is an imposed limit of 12 images for the authentication grid, which could be easily adapted to correspond to a mobile keypad and deploy the system in different environments. For touch screen devices this limit can be increased to 16 images. As it is a new mechanism its suitability would initially be only for low-risk domains. Usability issues such as authentication time, password resets and memorability defects are studied in the experiment presented in the next chapter. Larger usable security topics, subject to further research, are referred to in the discussion chapter.

4.3 Technical Specifications

Based on the concepts presented in the previous section, a prototype for ImagePass was jointly developed with a hired programmer in order to test several usability and security aspects of the given propositions. To minimize development costs, the prototype excluded some security-specific functionalities that were not relevant for the study.

Since the intended use for studying the ImagePass prototype is the World Wide Web, the application was developed using the .NET 3 Framework in conjunction with AJAX supported by a Microsoft SQL Server 2005 database. Besides the previously specified functionalities, the application also logged all of the user activities at all phases as a support to the study. The recording logs included all clicks and time intervals for each user who initiated a browsing session with ImagePass. Using graphical password is obviously more bandwidth demanding than plain text. Therefore, all images were highly optimized for web use. The average image size in ImagePass is 14.32KB, which averages the weight of the authentication grid at 171.84KB. This translates to a 1 second wait on a 1Mbps line, implying that the system will have a slight decrease in usability only at non-broadband Internet speeds.

4.3.1 Database Structure

To support all the features of the system and the subsequent testing and evaluation, several database tables were created that store user and log information (Figure 4.7). In addition, stored procedures for data manipulation were also created to speed up system performance and response times between the database and the application server.

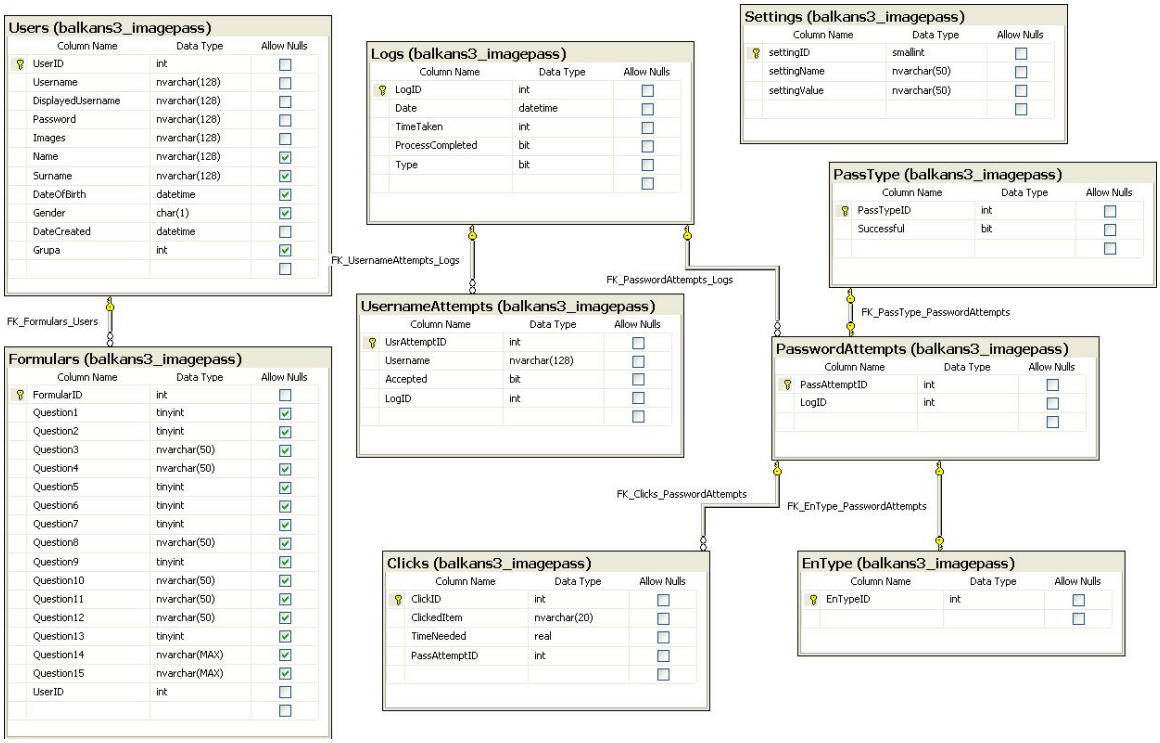


Figure 4.7: *Database structure*

All the user information is recorded in the Users table. It contains typical User fields such as first/last name, username, and date of creation. The graphical password images are recorded in the Password field, while Images contains all the images from the authentication selection grid. DateOfBirth and Gender were added for the purpose of

collecting demographic data during the usability study and are not meant to be a part of the system. From the stored procedures it is important to mention CheckUsername, which checks whether an entered username exists in the database and VerifyUser which checks whether the username and graphical password match during authentication.

ImagePass images are physically stored on the disk in a protected folder which has permissions to be accessed only through the database. Only references to the original image files are stored in the Images table. Before any images are loaded into the graphical password selection grid temporary filenames are generated to display the images. The img tags are created with the faux name attributes. Once the graphical password is selected, the user-specific authentication grid is stored in the Users_Images table. The password is stored as a hashed string containing ID's from the selected images.

The Logs table records either the enrollment or authentication process, determined by the Type field. It stores data on when the process was started, how long the process lasted and whether the process was successfully completed or aborted at a particular stage. Additional tables are related to the Logs table in order to record the details of each process. All the username entries: valid usernames, existing usernames during an enrollment session and nonexistent usernames during an authentication session, are recorded in UsernameAttempts. The graphical password entries are more complex as it is necessary to record the clicked images, clicked options (e.g. Reset, Load a new image set), time between clicks, entry phase (graphical password selection, graphical password confirmation or authentication) and successful/unsuccessful attempts. All this information is stored in the PasswordAttempts, Clicks, PassType and EnType tables respectively. All of the stored procedures for these tables are used for manipulation of log data such as CreateLog, UpdateTimeTaken, InsertUsernameAttempt, InsertPasswordAttempt, RegisterClick, etc.

The purpose of the Formulars table is to support an evaluation questionnaire needed for the usability study. In addition, the Settings table deals with some administrative privileges such as minimal password length and from which folder to load the images used in this system.

4.3.2 Application Layer

The communication between the database and the web interface was established through a .NET application. This application is essentially defined through three classes: MyMembership, Logger and ImagePass. As this is not a dissertation about programming structures, the application code and algorithms have been excluded and only a brief explanation is given for each class in the following paragraphs.

The MyMembership and Logger classes are used for communication with the database. MyMembership manages the users of the system and contains functions that either read or write data from/to the database. Both read and write functions follow a similar execution process, where a static variable describes the connection string to the database while both functions accept arguments compatible with the database nomenclatures. The Logger class follows the same concepts of the MyMembership class with the functions limited only to data entry.

The third class, ImagePass, contains the functions and procedures to generate and operate ImagePass authentication and enrollment selection grids. The behavior of this class is primarily defined through two functions: GenerateImagePassGrid, that generates the XTHML code representing the authentication or enrollment graphical password selection grids, and GenerateImagePassForUser which is used to complete the 12 item graphical password imageset with decoy images. The GenerateImagePassGrid function contains a Randomize sub-function that deals with randomization of the authentication

image position before display.

4.4 Continued System Designs

Over the three years of analysis and experiments the ImagePass system went through several iterations, each an improvement based on experiment results. A brief overview of the changes between the ImagePass version presented in this chapter and the final versions are presented next.

4.4.1 Web Version

The final ImagePass prototype has the following visual and system changes (Figure 4.8):

- The prototype has a redefined visual interface environment that addresses identified interface issues.
- The display of visual content is adapted to the device used for access through separate CSS files.
- The user authentication grid was increased from 12 to 16 images. This increases the password space for the authentication key and also increases the compatibility of the system with mobile devices as they regularly display sixteen objects per screen..
- During authentication entering the username and graphical password takes place on the same screen. The username dependent image loading process is supported by a separate AJAX component.
- The minimum graphical password length is set at 4, as that is the optimal "comfortable" length as shown in the last experiment presented in this dissertation.
- Images can be tagged by the administrator through several content categories such as color, shape and purpose.

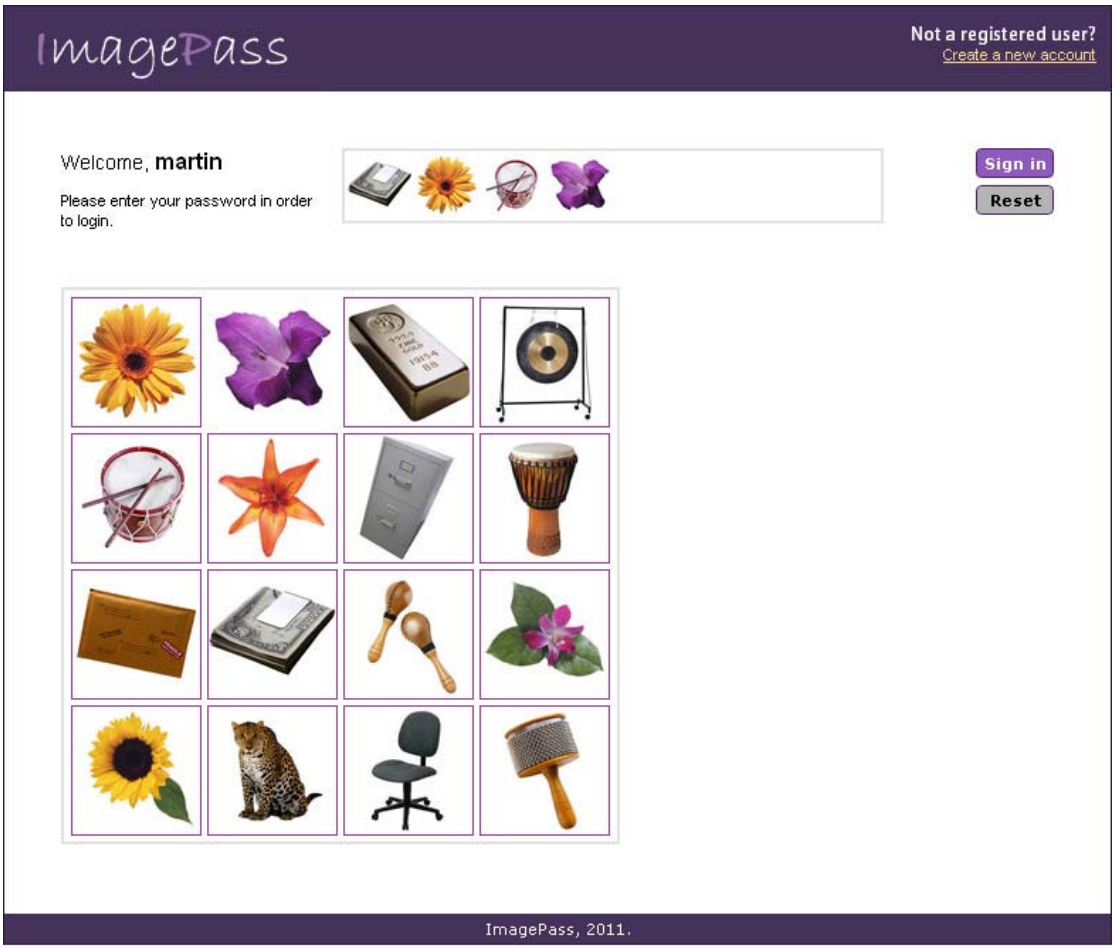


Figure 4.8: *Final web version*

The most up-to-date web version is available at http://www.martin.com.mk/imagepass

4.4.2 Mobile Application Version

A special mobile application was developed for ubiquitous devices that use the Android operating system. This application was developed by using the Android SDK which is publically available on the android web site. The application is used to protect the device from unauthorized access. Currently, the application requires an Internet connection in order to load the images for enrollment, however once the enrollment process is complete the imageset is stored and hashed locally on the device.

The design of the mobile ImagePass application is modeled on the classic Android 2.0+ application screen, displays 16 items in a grid. During graphical password selection the user can swipe through the available images on 5 screens. The maximum password length is set at 5 images.

Figure 4.9: *ImagePass mobile application on a Samsung Galaxy S*

4.5 Summary

This chapter presented the development process of a graphical authentication mechanism, codenamed ImagePass, by following usable security quality criteria. Conceptually, in ImagePass a user uses several images, each containing a single object, from an image database as their authentication key. During enrollment, the user chooses a username and is then presented with individual images, each containing a single object, to choose the graphical password sequence. During authentication, the user enters the graphical password by clicking on the previously selected sequence of images placed among decoy substitutes.

As discussed in Chapter 2, researchers have developed various graphical authentication mechanisms, with varied results in different environments. In retrospect, the most successful and widely researched recognition-based graphical authentication mechanism is Passfaces, a graphical-password system which employs facial photographs for authentication. In Passfaces, the user creates an authentication key by selecting five faces, one each from five successive grid displays of nine faces. By following usable security guidelines, ImagePass improves upon the conceptually similar applications like Passfaces in several areas:

Single screen authentication. Unlike Passfaces, ImagePass does not guide the user through several panel screens in order to enter the graphical password, instead, all the images are presented in a single grid. On the one hand, using multiple panels guides the graphical password sequence, but on the other it decreases the usability of the system as the user has to wait for the panels to switch.

Increased authentication key space. Besides having a lower authentication key space, in Passfaces the potential attacker is informed about the graphical password length by the number of panel switches between image selections.

Equitable images. Another drawback of the Passfaces system is the limitation of image content to human faces, which has shown potential for racial and gender bias.

The following chapters present a thorough usability analysis on many different aspects of the system. They contain four experiments that: a detailed user study, a two part eye tracking study and a graphical password analysis based on grid position and image content.

5 Evaluating Recognition-based Authentication

As users can only make informed choices when the proposals being discussed are meaningful to them, enabling users to envision and make sense of those proposals is an essential element of all approaches to system design. Therefore, the design and development of the first functional ImagePass prototype was followed by a thorough investigation of the usability of the system. The evaluation was divided into two phases intermitted with a system redesign based on the discovered issues.

The research presented in this chapter focuses on analyzing different aspects of the users' perception and behavior when interacting with a recognition-based system. In the first section, ImagePass is deployed in a live environment and subjected to a large six-month long user study. Several control groups are observed under different conditions in order to determine the effects of training and frequency of use. The experiment also addresses memorability dependence on password length, as well as potential cognitive differences between male and female users. This was followed by an exploratory eye tracking study which evaluates cognitive aspects of user when interacting with the system.

5.1 Experiment 3 – User Evaluation

5.1.1 Introduction

Essentially, there has been no coordinated work towards an accepted standard for evaluating the usability of graphical password schemes. Every evaluated system presents tests and results using different criteria. Even when apparently similar measures are reported, they are often calculated with different methods. Context and specific recommendations have been suggested only recently (Biddle et al., 2011). These recommendations focus on three topics: target users, tasks and domains.

5.1.1.1 Target Users

Expectantly, the characteristics of the target users for the graphical authentication mechanism must be considered when designing a graphical password scheme. Two important characteristics are:

- **Expertise level**, which will define the acceptable complexity of interaction with the system, and the training required in order to use the system.
- **Cognitive effort level** as defined by the frequency of use. Frequently used graphical authentication mechanisms may rely more heavily on users' memory, whereas infrequently accessed systems must be especially memorable since memory decays over time.

Issues of accessibility may also arise since different user populations have different requirements. Graphical authentication mechanisms implicitly require users with good vision, potentially including good color vision for recognizing cues, and good motor skills for entering sketches or accurate clicks on an image. Hence, the design of graphical

password systems needs to either address these issues or provide suitable alternatives.

5.1.1.2 Tasks

Although ease of login is the most frequently examined task when evaluating authentication mechanisms, the usability of the system should be explored along several dimensions. Essential elements to measure and report include: time to create a password, time to login; success and error rates over an extended period, and multiple password interference. Typically the usability evaluation focuses on all segments of authentication: enrollment, login and password recovery.

During **enrollment** the graphical password can be either system-assigned or user-selected. User-selected passwords have improved usability as a password with a personal meaning might be easier to remember. On the other hand, this can lead to password predictability and password reuse across systems. Systems which assign randomly selected graphical passwords prevent attacks which exploit predictability, and eliminate cross-account password reuse. However, such systems require time-consuming training to help users remember their passwords and even then passwords may remain more difficult to remember (Weinshall, 2004). Password confirmation is a common occurrence in enrollment procedures which ensures that users do not make trivial entry errors, and can accurately remember and enter their graphical password after a short period of time.

The **login** procedure should be quick and simple as it is the most common task completed by the users of the authentication system. The time to enter a graphical password, error and success rates on login are common usability measures most often reported in user studies of graphical passwords, nevertheless, they are often calculated in different ways and measured at different times. Memorability issues are important when discussing login performance, as memorability is a main factor determining login success. Measures of memorability address whether passwords can be remembered over short- and long-term and with varying login frequencies.

Password reset is not typically examined during usability testing of a new graphical password scheme. The process may involve the user interacting only with the system, or may require contact with help desk personnel. Both would require the user to confirm his/her identity through some secondary means, and issuing a new password which can be easily communicated by phone or through email. Graphical passwords cannot be communicated as easily, which poses a unique usability challenge.

5.1.1.3 Domains

The performance constraints and goals for a graphical authentication mechanism will differ depending on the intended environment of use. In a new scheme, the target environment should be clearly declared, to allow comparison of systems intended for similar conditions, and to avoid deploying systems in inappropriate domains.

In high-risk domains it may be acceptable to have a system that is slightly more difficult to use in order to achieve the desired level of security. Conversely, in low-risk domains, it may be acceptable to have very usable, but lower security schemes. It is unlikely that a single scheme will be suitable for all domains, thus, specifying the target environments and applications for newly proposed schemes is important.

5.1.2 Research Questions and Hypothesis

The first evaluation experiment was conceived as a usability study of the users' interaction with the ImagePass graphical authentication mechanism in a web environment. The performed evaluation aimed to uncover the potential difficulties as the

users went through realistic usage scenario tasks. As user reported views do not reflect actual performance (Etezadi-Amolia & Farhoomand, 1996), the primary source of information for the experiment were system logs, with information gathered from focus groups and post-session interviews used as complementary data. Due to the cognitive nature of this authentication approach, the differences between male and female users were also considered. Conversely, user performance was evaluated over continuous use of the system, and the effect of training was assessed.

The formal statements of the hypothesis tested in this experiment are as follows:

H5: There is a difference in performance for graphical authentication between male and female users.

H6: There is a difference in performance for graphical authentication between users based on frequency of use.

H7: System performance is not affected by password length when the password is selected by the user.

H8: Graphical password memorability is affected by mnemonic training.

5.1.3 Participants and Equipment

Initially, 211 participants, 82 male and 129 female, were recruited to participate in the study, most of them being undergraduate students from the Faculty of Economics or the Faculty for Information Sciences at Ss. Cyril and Methodius University in Skopje. Their age varied between 21 and 24, with a mean age of 23.1. The participants were either casual or experienced computer users who use the Internet for study and leisure purposes on a daily basis.

The ImagePass application was deployed in a network environment on a Dell Poweredge 2950 Server with a Microsoft Windows Server 2008 SP2 operating system, SQL Server 2008 database engine and IIS 7 web server running .NET Framework 3.5 SP1. The supervised phase of the experiment was completed in two computer labs at the Faculty of Economics in Skopje. A total of 88 PC's running Windows XP Professional on Pentium IV processors with 15 LCD monitors set on a 1024x768 resolution were used. Each PC had both Internet Explorer 7 and Mozilla Firefox 3 web browsers installed, with the web browser choice left to the discretion of the participant. The unsupervised phase of the experiment was performed from faculty computer labs, home computers, Internet cafes and mobile devices.

5.1.4 Experiment Design

The ImagePass authentication mechanism was deployed online to a public IP address that was accessible through a domain name. The experiment was a hybrid study that combined a supervised lab-controlled environment for evaluating ImagePass enrollment, and an unsupervised web-based study to evaluate the continuous use of the system. To test Hypothesis 6 and study the effects of how repeated use of graphical authentication affects user performance, the participants were observed under two different group conditions. The conditions were assigned before the beginning of the experiment and they were based on how frequently the participants were going to use the system. The participants in the first control group would access the system approximately once per week (Group 1), while participants from the second control group were intended to access the system approximately once a month (Group 2). To test Hypothesis 8 and explore whether any

sort of training could influence the users' perception and performance with ImagePass, participants from both frequency groups, Group 1 and Group 2, were subdivided into two more groups. The first subgroup, Group A, received no additional input that could be considered as training, while the second subgroup, Group B, received mnemonic instructions on remembering a series of items (Buzan, 2000, Appendix A). Hence, Group 1A would access the system once per week and receive no mnemonic instructions, Group 1B would access the system once per week and receive mnemonic instructions. Conversely, Group 2A would access the system once per month and receive no mnemonic instructions, while Group 2B would access the system once per month and receive mnemonic instructions.

The main data was collected through system logs and focus groups, while secondary data was collected through brief post-session interviews. System logs were programmed as a part of the application and silently collected data in the background by registering data entered, left-clicks, page loads and time passed between events. Each log data was classified based on the active screen and the clicked page item. As it was expected that the log analysis would yield some anomalies or peculiar behavior, informal post-session interviews took place within a 10-day time span after the experiment. When necessary a participant was contacted by phone to clarify the particular occurrences of specific data.

The system log data was compiled and processed to define several variables necessary for the analysis of task performance measured by time and errors. In the context of authentication, failed login attempts are user costs and should be minimized where possible. If two people have the same number of failed logins, but different numbers of successful logins, then counting the absolute number of failed login attempts would be misleading. Therefore, a measure for login failure has been defined by Brostoff & Sasse (2001) as $L_{fr}=L_f/(L_f + L_s)$, where L_{fr} is the login failure rate, L_f is the number of failed logins and L_s is the number of successful logins. Other variables that were compiled from the system logs and analyzed were: phase-completion time (the time the participant spent to complete a phase of the experiment), task-completion time (the time the participant spent to complete a specific task) and authentication-key length.

Two focus groups interviews took place after the Enrollment phase of each session with each group containing 10 participants. The focus groups followed a semi-structured process based on pre-determined questions that lasted between 20 and 30 minutes. Audio recordings and examiner notes were used to collect the participants' data which was later compiled and analyzed as presented in the next section.

5.1.5 Procedure

The supervised, enrollment phase of the experiment was completed in four sessions over two days, with each session lasting around 30 minutes, on average. Each session took place in a computer-lab environment and was supervised by a researcher. Every participant worked individually on a personal computer and had to complete four tasks: Create a new username (T1.1), Select a graphical password (T1.2), Confirm graphical password (T.1.3) and Enter personal data (T1.4). Before each session the participants were told that they were going to work with an authentication mechanism based on images. Prior to two of the sessions the participants were asked to read a two-page text with mnemonic instructions on how to remember things efficiently before proceeding with the tasks. As the ImagePass application was publicly available online, to prevent random artifacts the IP addresses of the participants were tracked and only data generated in the supervised environment was considered for further analysis.

The second, unsupervised phase of the experiment, continuous web-use of the system, was completed at different time intervals over a period spanning five months. To motivate

participants to continuously use the system and to avoid evaluating authentication as a primary task, study materials for two courses were published online and were available to the participants once they successfully authenticated through ImagePass. To control the frequency of use, materials for one of the courses were published weekly and materials for the other course were published monthly. The availability of new materials was announced to all participants via email. To complete a session in this phase of the experiment, the participants had to perform two tasks: login using ImagePass (T2.1) and download study materials (T2.2).

5.1.6 Results and Discussion

As users in this study can be clustered into different groups based on frequency of access and received instructions, descriptive statistics and an ANOVA analysis explored the user behavior based on the clustering parameter. Data anomalies were clarified in the informal post-session interview and the conclusions are presented along with the data, where appropriate.

From the 211 recruited participants, a total of 151 showed up for the Enrollment phase of the experiment in a supervised environment. From the remaining participants, 52 completed the tasks from a different environment, while 8 withdrew from participation. The demographics of the participants, per group, are presented in Table 5.1. Except for group 2A, there is a small prevalence of female users, with a 60/40 ratio to male users.

Table 5.1: *Gender distribution for participants per clustering group*

Group	Male	M % of Group	M % of Total	Female	F % of Group	F % of Total	Total
1A	18	40.91%	11.92%	26	59.09%	17.22%	44
1B	16	32.00%	10.60%	34	68.00%	22.52%	50
2A	14	51.85%	9.27%	13	48.15%	8.61%	27
2B	12	40.00%	7.95%	18	60.00%	11.92%	30
Total	60		39.74%	91		60.26%	151

A total of 114 users continued using the system after the activities in the controlled enrollment, with 11 requesting graphical password resets. Based on frequency of use, 62% of the participants were members of group 1, with weekly access to the system, while 38% of the participants were members of group 2, with monthly access to the system. Regarding the subdivision based on received instructions the distribution is 53% to 47%.

5.1.6.1 Log Analysis

During the Enrollment phase, 248 usernames were entered in the username field, with 214 username entries accepted as valid and 34 entries rejected as either duplicate or blank attempts. As clarified in the post-session interviews, most of the duplicates were generated by two participants. The first circumvented the lab environment by downloading and installing a beta version of the Google Chrome browser, which was incompatible with the deployed authentication mechanism. This was a browser the participant was most comfortable with and he had a habit of always installing it on machines if such an action was allowed. The other participant tried to continuously enter her name as a username, regardless of the “Username already in use” message. As this was a new system she was a little confused by the message and was hoping that the username rejection was due to a system error.

Out of the 214 accepted username entries, 203 continued with the enrollment procedure, as 9 entries were from participants playing-around with the system after they had completed the experimental tasks. The accepted usernames initiated session logs that registered 499 enrollment passwords: 203 selected passwords, 89 reset passwords (17 unique), 203 confirmed passwords, 4 mismatched passwords (1 unique), and 2065 image clicks. The reset-password anomaly was caused by a single participant who compulsively clicked the Reset button during graphical password selection as he found the task uninteresting. Most participants using the Reset function reset the password once. All the mismatched password logs were caused by a single participant. Out of 203 completed enrollments, 26 participants used only one image repeatedly as a password sequence, 17 participants had a repetition of an image in their password sequence and 160 had an authentication key with completely different images.

During the Continuous use phase, 837 usernames were entered in the username field for user identification, with 695 usernames accepted as valid and 142 rejected as nonexistent. The accepted usernames initiated 586 session logs. Most of the rejections were generated by a few participants accessing the system via unsupported browsers. The session logs registered 1061 graphical password attempts with 5090 image clicks and 101 graphical password resets. The discussed data is presented more concisely in Table 5.2.

Table 5.2: *System log summary*

	Entered usernames	**Accepted usernames**	**Session logs**	**Password attempts**	**Image clicks**
Enrollment	248	214	203	499	2065
Identification / Authentication	837	695	586	1061	5090
Total	1085	909	789	1560	7155

Regarding the authentication key, the minimum length was 4 images, the minimal allowed, while the maximum was 10 (M=4.34, SD=0.799), with most participants using a 4–5-image-long graphical password. A one-way ANOVA analysis showed that there was no significant difference for the length of the chosen authentication key between groups A and B, $F=2.067$, $p=0.153>0.05$, Hence, the mnemonic instructions had no influence on graphical password selection, which is partially not supportive of Hypothesis 8. Even though users could theoretically better memorize a series of items, they still selected an authentication key with the same length as other users. A one-way ANOVA analysis showed that there was no significant difference in the length of the chosen authentication key between genders as well, $F=0.47$, $p=0.828>0.05$.

The minimum time to complete the first three tasks of the enrollment session was 25 seconds, and the maximum was 327 seconds (M=130.83, SD=60.7). A one-way ANOVA analysis suggested that there was no significant difference for the completion time between groups A and B, $F=2.856$, $p=0.153>0.05$. This implies that mnemonic instructions had no influence on performance time, which further supports the alternate hypothesis to Hypothesis 8. On the other hand, a one-way ANOVA analysis using completion time as the dependent variable and gender as the independent variable suggests a significant difference between genders, $F=4.84$, $p=0.029<0.05$, which is in support of Hypothesis 5.

To check whether a relationship exists between password length and enrollment completion time, a Pearson's correlation test was performed. The results showed no significance for all the participants ($r=0.083$, $p=0.301$). An additional Pearson's correlation test examined the relationship between password length and login failure rate,

showing no significance for all the participants (r=0.062, 0=8.421). These results sustain Hypothesis 7, as password length does not influence the users' performance.

From the continuous-use sessions the following variables were compiled from the system logs and were subjected to further analysis: password attempt time, and login failure rate. The average time to enter the password for all participants is 10.48s (SD = 27.957). When password attempts are divided into successful password attempts and unsuccessful password attempts the average times are 9.53 (SD=22.865) and 12.75 (SD=37.350) respectively. The descriptive statistics for the clustering groups are presented in Table 5.3.

Table 5.3: *Descriptive statistics for login analysis based on clustering groups*

Login		**Group 1**	**Group 2**	**Group A**	**Group B**	**Male**	**Female**
Unsuccessful	Mean	9.101	18.326	11.933	14.163	15.978	10.321
	SD	9.0088	18.7658	14.1528	14.8425	16.9817	11.6144
Successful	Mean	9.6	13.2225	10.3725	10.8	9.45	11.235
	SD	11.358	25.004	17.7155	14.0195	12.9613	17.8523

A one-way ANOVA analysis revealed the following results for successful and unsuccessful logins. There was a significant difference for the completion time in successful logins between the participants from group 1 and the participants from group 2, $F=4.598$, $p=0.033<0.05$. There was no significant difference for the completion time in successful logins between the participants from group A and the participants from group B, $F=0.79$, $p=0.778>0.05$. There was no significant difference for the completion time in successful logins between the male and female participants, $F=1.338$, $p=0.248>0.05$. There was a significant difference for the completion time in unsuccessful logins between the participants from group 1 and the participants from group 2, $F=12.556$, $p=0.001<0.05$. There was a significant difference for the completion time in unsuccessful logins between the male and female participants, $F=4.532$, $p=0.035<0.05$. There was no significant difference for the completion time in unsuccessful logins between the participants from group A and the participants from group B, $F=0.35$, $p=0.723>0.05$. Regarding the login failure rate, a one-way ANOVA revealed a significant difference between groups 1 and 2, $F=11.826$, $p=0.03<0.05$, and no significant difference between groups A and B or the genders.

An interpretation of these results leads us in the following direction. Hypothesis 5 is only partially supported as the differences in gender-based performance are only significant during unsuccessful logins. Hence, when failure is introduced there might be a gender-specific cognitive behavior that influences the process of authentication. However, before any claims can be specified and elaborated this occurrence needs to be studied further. Hypothesis 6 is supported as there is a difference in performance between the users for all the tested variables. Participants who use the system more frequently have more successful logins, make fewer authentication mistakes and complete the authentication procedure more efficiently. Hypothesis 7 is not supported as there was no relationship between password length and enrollment completion time and between password length and login failure rate. Hypothesis 8 is not supported under any circumstances, as no improvement was measured for participants that received mnemonic instructions for any variable.

5.1.6.2 Focus Group

The semi-structured focus-group sessions were concentrated around the participant's

perception of the graphical authentication process and the user interface. The question structure of the sessions is presented in Table 5.4.

Table 5.4: *Focus group session structure*

Perception of graphical authentication	**User interface**
- Were you aware that all the images showed a picture of a single object?	- Did you understand the interface?
- How memorable was the choice of images available for the password selection?	- How obvious was it that you had to click the Create Account link to enroll to the system?
- Why did you choose the images you chose for your password?	- How acceptable was the separation of username and password in different screens?
- Would you consider using a graphical password system? Please explain why.	- How obvious was the graphical selection process?
	- How useful did you find the display showing your password selection?
	- How obvious was the reclicking of the password in order to successfully complete the enrollment process?

After settling down, the first question of each session was intended to explore the cognitive awareness of the participant about the type of images. Surprisingly, only around 40% of the participants were cognitively aware that images represented single objects. There were no general complaints about the images themselves, with most of the images remembered as impressionable and a few remembered as indistinct. The participants were asked to rate their perception of password memorability on a 5-point Likert scale and the data was analyzed using one-way ANOVA. The analysis showed a significant difference, $F(1, 78) = 5.079$, $q=0.026<0.05$, between the participants from group A and group B. This implies support for Hypothesis 8, as the participants who had some knowledge of how to better remember items found the system more memorable.

When asked to explain their image choice, the initial answer of most participants was that they chose items they can relate to and/or have some personal meaning. Interestingly, after a brief discussion a significant number of participants found themselves related to food items, remembering objects like cakes, porridges, fruit, etc. When asked whether they would consider using ImagePass or a similar graphical authentication mechanism in the future, there were divided opinions, with around half of the participants expressing some kind of an affirmative answer. The participants that expressed a negative opinion listed security concerns and memory difficulties as the reasons for not feeling comfortable with the system.

As for the participants' understanding of the interface, they were asked to rate their impression of the system on a 5-point Likert scale. A one-way ANOVA analysis showed a significant difference, $F(1, 78)=3.389$, $q=0.048<0.05$, between the participants from group A and group B. This result is in support of Hypothesis 8, as participants who had instructions on how to better remember items found the system more usable. Regarding specific functionalities, participants found no difficulties with starting the enrollment process. They found the separation of username and password untraditional, as their

habits and expectations were not to change screens in the enrollment process during this phase. The graphical selection process was somewhat intuitive with the textual information on the screen described as helpful. The re-entering of the graphical password was acceptable, with a few participants suggesting the disabling of the entered password display field during confirmation.

5.2 Experiment 4: Preliminary Eye Tracking

5.2.1 Introduction

The screen search behavior of users is governed by expectations about what is being looked for and where it might be found (McCarthy et al., 2003). Evidence from search studies indicate that unless the sought object has a unique feature such as color, contrast or motion, the search proceeds by selecting elements one by one (Treisman & Gelade, 1980).

Eye tracking research in the field of graphical authentication is rather scarce. Some eye tracking studies have been performed on mechanisms such as Passpoints (Le Blanck et al, 2008). The authors investigated whether eye fixations can predict the location points for graphical passwords, whereas they concluded that eye gaze is not a good predictor of passwords. In a follow-up study with a modified experiment design (LeBlanck et al, 2010), the predicting success was slightly improved, but it still remained rather limited.

In order to understand the reasoning behind studying eye movements, it is first necessary to introduce some basic facts about the human vision and eye tracking methods.

5.2.1.1 The Eye

Light reflected from an object or a scene travels into our eyes through a lens which concentrates and projects the light on to a light sensitive surface located on the back of a closed chamber. This surface, also known as the **retina**, is not equally sensitive everywhere. We can only see clear details in a limited part of our visual field, the **foveal area**, while a larger part of our visual field is dedicated to blurry and less colorful images in the **peripheral area**. The region of transition between these areas is called the **parfoveal area** where the image gradually changes from blurry to focused.

The main reason for the differences in our visual field is the two different kinds of light receptor cells available in the eye: the **rods** (94%) and the **cone cells** (6%) (De Valois & De Valois, 1988). Rods require little light in order to work, but they only provide a blurred and colorless image. For detailed clear vision, our eyes are equipped with cone cells. Cone cells are most often available in three different varieties based on the color they can register: red, green and blue. The peripheral area is mostly covered by rods, and cones are mostly found within the fovea.

The human visual field spans around 220 degrees. The visual data is primarily registered through the foveal area, constituting 8% of the visual field, but 50% of what is sent to the brain via the optic nerve. The peripheral area is only good for registering movements and contrasts. By letting the foveal area register the image, the brain gets the highest resolution for the area of interest. Besides having a very limited sharp field of vision, our eyes are fairly slow at registering changes in images. The retina needs approximately 80 ms to register a new image in normal light conditions (50-60ms for perceiving a word, 150ms to interpret a picture).

Eye movements have three main functions which are deemed as relevant when processing visual information. To focus on the area of interest the eye uses fixations,

(pauses on a specific area), and saccades (rapid movements between pauses). Tracking a moving focused area in the retina is controlled via smooth pursuit movements, with the eyes' speed adjusting to the target speed. Perceptual fading of stationary objects is compensated by microsaccades, tremors and drift.

5.2.1.2 Visual Attention

The perceptual selection process of consciously or unconsciously focusing on a fraction of the total information we could potentially process is called **attention** (Heijden, 1992). Moving our eyes from one point in the visual field to another is referred to as **overt attention**. Moving the mind's attention to peripheral areas of the visual field without eye movements is called **covert attention**. The most frequent use of these two mechanisms is simultaneous.

5.2.1.3 Eye Tracking Process and Methods

The eye tracking process and methods are best described in Tobii (2009). Eye tracking studies identify and analyze patterns of visual attention of individuals performing specific tasks such as reading, searching, scanning an image, driving, etc. The eye movements are typically analyzed in terms of fixations and saccades. As we perceive the world visually only through fixations, during each saccade visual acuity is suppressed and, as a result, we are unable to see at all, the brain integrates the visual images that we acquire through successive fixations into a visual scene or object. We are only able to combine features into an accurate perception when we fixate and focus our attention on them. The more complicated, confusing or interesting those features are the longer we need to process them and, consequently, more time is spent fixating on them. In most cases we can only perceive and interpret something clearly when we fixate on an object or are very close to it. This eye–mind relationship is what makes it possible to use eye movement measurements to tell something about human behavior.

The basic idea behind eye tracking is that our eye movements can be used to make inferences about our cognitive processes (Peyrichoux & Robillard-Bastien, 2006). An eye tracker follows the user's eye movements by reflecting infrared light onto the eye and then, using a geometrical model, determines the exact gaze point of the user.

Eye tracking is a promising method for detecting usability problems in websites (Cowen et. al, 2002; Ehmke & Wilson, 2007). Nevertheless, when using eye tracking in usability studies it is important to select the most suitable methodology in order to extract relevant and useful data from the participants. Eye tracking data should be combined with additional qualitative data because eye movements cannot be clearly interpreted without the participant providing context to the data. In (Hyrskykari et. al, 2008), the authors show that longer fixations can either mean a user found a particular area interesting or that a user found the area difficult to interpret.

One of the most convenient methods for gathering data relating to usability problems are think aloud methods (Van den Haak, 2003) which can primarily be split into two types: concurrent think aloud methods (CTA) or retrospective think aloud methods (RTA). In concurrent think aloud users are asked to actively verbalize their thoughts while performing a specific task. In retrospective think aloud users describe their actions and experiences after the task is completed.

Although both types of methods are useful for gaining insight into the participants' thought processes regarding task completion, CTA methods have several limitations when eye tracking is involved. As cognitive processes are quicker than verbal processes, participants might be thinking more than they are able to verbally express (Eger et. al, 2007). Also CTA methods are more easily affected by reactivity where participants

perform better or worse in completing tasks due to the nature of the task. Many participants forget to express their thought processes aloud when encountering difficulties interacting with the user interface if CTA is used (Guan et. al, 2006). Using concurrent think aloud methods in eye tracking studies is less suitable as participants produce eye movements which would not be present if the task was performed in a regular environment (Kim et. al, 2007). Hence, retrospective think aloud methods are recommended for conducting usability tests that include analysis of objective eye movement data, even though RTA methods have their own drawbacks such as: users forgetting first impressions or task-related steps and intentional or unintentional fabrication of information (Van Gog et.al, 2005).

Since fallible memory and potential for fabrication can be problems when performing traditional RTA usability tests, a variety of cued RTA methods have emerged. In a cued RTA the user is presented with a form of replay of the interactions they previously performed in order to help cue their memory (Eger et. al, 2007). Replays could be:

- video cue (screen video)
- gaze plot cue (superimposed eye movements on still images)
- gaze video cue (superimposed eye movements on a screen video)

This integrated approach to usability testing has proven to be a way to gain richer data from participants. Presenting these visual stimuli serves as a way to get more detailed information, but also allows the participants to reflect upon their actions in a way they might not have been able to do otherwise. The information gathered from the eye tracker accounts for much of the quantitative data needed, whereas the cued RTA provides qualitative data input from the participants.

The post-experience eye tracked protocol (PEEP) method that utilizes playbacks of people's eye movements during RTA has shown to be potentially better than a video without eye movements when exploring new or complex environments (Ball et. al, 2006). Using a video cue that features eye movements (a gaze video replay) has been demonstrated as more effective at eliciting comments from users than an uncued RTA (Kim et. al, 2007). Showing a playback of participants' eye movements overlaid on a video showing the steps they took while completing a task has proven to be a successful way to elicit information from the participants and, in addition, allows for an accurate measure of other variables, such as task time (Van Gog et. al, 2005). One study showed that even if the participants stated that they mostly relied on their memory when talking about a recently conducted task, they did find the video helpful as a reminder (Van den Haak et. al, 2003). In addition, when using video as stimuli for cued RTA, recollections of the task turned out to be very accurate according to actual task performance, i.e. the video almost eliminated the risk of fabrication.

5.2.2 Research Questions and Hypothesis

Implicitly, as ImagePass is a novel authentication mechanism that uses an unfamiliar interface, it might be difficult for potential users to grasp the intricacies of the system and locate the right options to achieve their tasks and goals. Since the system is highly visual, an eye tracking analysis was performed to understand how users would perceive the ImagePass graphical authentication mechanism. The purpose of the experiment was two-fold: to test the usability of the generic components that define the visual interface and to gain initial knowledge on how users behave when they encounter graphical authentication. The goal was to validate the interface components by expecting emergent gazing patterns to coincide with findings from other eye tracking studies:

- users' attention is primarily drawn towards text rather than graphics and photos (Stanford-Poytner Project, 2009)
- headers are visited before the body (Goldberg et al., 2002), and
- the natural top-left to bottom-right gaze path is influenced by content features (Josephson & Holmes, 2002a).

As this was a preliminary exploratory study with a few participants no formal hypothesis were stated.

5.2.3 Participants and Equipment

As suggested by Virzi (1992), observing four to five participants will allow a usability practitioner to discover 80% of a product's usability problems with an average problem detection ranging between 0.32 and 0.42. Consequently, for the purpose of this experiment 5 participants, 3 male and 2 female, were recruited through direct contact. The ages of the participants were between 17 and 28 with a mean age of 24.3 years. All of the participants had normal vision and were experienced computer users who use the Internet for professional and leisure purposes on a daily basis. None of the participants had any previous experience with graphical authentication.

The eye tracking experiment of the ImagePass graphical authentication mechanism was conducted in the Laboratory for Open and Network Systems at the Jožef Stefan Institute in Ljubljana, Slovenia. The ImagePass software was deployed on a remote web server and accessed through a web browser. Eye movements were collected using the Tobii T60 eye tracker hardware from Tobii Technology AB. The system samples eye position at a rate of 60Hz. Raw eye coordinates are converted into fixations using an algorithm that assumes a fixation time of 200ms with 33ms latency and a drift of 0.1 degrees. The Tobii Studio 1.5.4 software was used to display stimulus pages and to define functionally distinct areas, called regions of interest, which could be analyzed separately. The content was viewed on a 17 TFT monitor with a 1024x768 resolution on a personal computer running Windows XP Professional and Internet Explorer 7. The eye tracker was successfully calibrated for all participants, thus complete eye movement data was recorded for all subjects.

5.2.4 Experiment Design

A predefined variant of the ImagePass application was deployed in an online environment and was accessible through a web browser. To give an indication of the spatial distribution of screen fixations, the screen areas of ImagePass screens were categorized according to their functionality and content in areas of interest (AOI). As the first screen, create username, has only one area of interest, the username field, it was not subjected to this analysis. The second screen, password selection, was divided into four areas of interest: Info, Passimage, Selected password and Load New Images. The third screen, password confirmation, was similarly divided into four areas of interest: Info, Passimage, Select password and Finish. The screen object categorization in areas of interest for this experiment is presented in Figure 5.1.

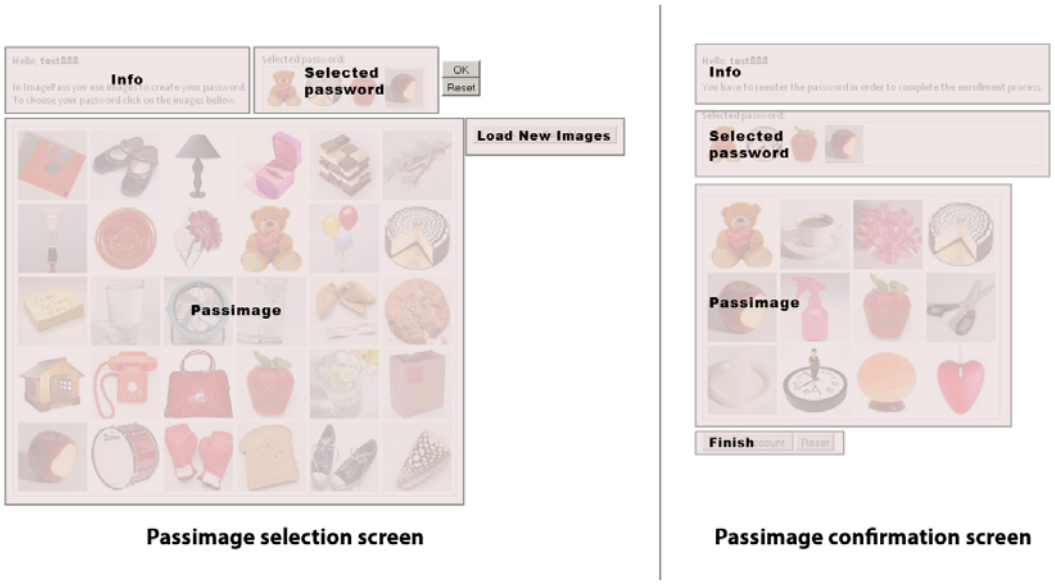


Figure 5.1: *Categorizations of visual areas for ImagePass screens.*

With this categorization it was possible to extrapolate how screen fixations proceed by defining a measure for Attraction (A) of each distinct area of interest. This measure was calculated by dividing the glances in the target area (G_{ta}) with the glances across all areas (G_{aa}). This represents the proportion of glances on a screen area, where Attraction value varies between 0 and 1 (McCarthy et al., 2003).

5.2.5 Procedure

The eye tracking behavior of the participants was analyzed through individual 10-15 minute sessions over 5 days. During each session, the outline of the experiment was explained to the participant. Regarding ImagePass, the participants were only given information that they were going to use a new type of authentication mechanism based on images. To get familiarized with the eye tracking hardware a practice session was set-up in which the eye tracker was calibrated to the participant's eye movements and a photographic test image was shown to the participant for two seconds. Afterwards the visual representation of the result in a form of a heat map was displayed and explained to the participant. The eye tracker was then recalibrated to the participant's eye movements before proceeding with the specific tasks.

During the eye tracking experiment, the participants had to perform two tasks in order. For the purpose of the evaluation, each task was divided in several subtasks with each subtask taking place on a different screen interface. The first task was to complete the enrollment procedure to the ImagePass system by following three subtasks: create username, choose password and confirm password choice. The second task was to login to the ImagePass system by following two subtasks: enter username and enter password. Although the eye tracking behavior of the participants was recorded for each subtask only data from three subtasks of the two tasks was used for further analysis: choose password (1.2), confirm password choice (1.3) and enter password (2.2). Eye tracking data from the first subtask of both tasks was discarded and was not subjected to further analysis, as the create username and enter username screens consisted of only one element.

5.2.6 Results and Discussion

All of the participants completed the given tasks successfully. The Tobii Studio software

was used to summarize the data visually through gaze plots and heat maps for each visited screen. The software determined whether an eye position in a given region is a fixation, generating both text and image data as outputs. The first observation that could be made from the initial data was a potential difference in user behavior between male and female users. Therefore, the collected data was segregated and analyzed by gender.

The enrollment task took the user through three subtask-related screens. An average of 127.75 fixations (101.5 for male, 154 for female) was needed to complete the task in 63.75 seconds (55.5s for male, 72s for female). There is a potential difference between genders for fixations with females observing the screen 51.7% more than males. There is also a potential difference between genders for the average time spent using the application, with females needing 30% more time to complete the tasks.

Since the fixations occur over a differing time interval a measure of Attention was assigned to the experiment as gazes per second. The results suggests that there is a potential difference between genders on how much attention is devoted to the application, with females paying 17% more attention to ImagePass during enrollment than males. To evaluate the perception of ImagePass in more detail the location information from the eye tracker was further analyzed for two screens: password selection and password confirmation. The resulting data for the Password selection screen is presented in Table 5.5., while the resulting data for the Password confirmation screen is presented in Table 5.6. Figure 5.2 displays a visual comparison chart for screen attraction between genders for both screens.

Table 5.5: *Summary of data gathered during password selection*

		Info	**Password**	**Selected password**	**Load a new set**
Male	Gazes	10.5	27	17	7
	Attraction	0.18	0.42	0.28	0.12
Female	Gazes	6	65	22	2
	Attraction	0.06	0.68	0.23	0.02
Group	Gazes	8.25	46	19.5	4.5
	Attraction	0.12	0.55	0.25	0.07

Table 5.6: *Summary of data gathered during password confirmation*

		Info	**Password**	**Selected password**	**Complete**
Male	Fixations	6	17.5	2	2.5
	Attraction	0.22	0.63	0.07	0.08
Female	Fixations	5	35	5	7
	Attraction	0.10	0.67	0.10	0.13
Average	Fixations	5.5	26.25	3.5	4.75
	Attraction	0.16	0.65	0.08	0.11

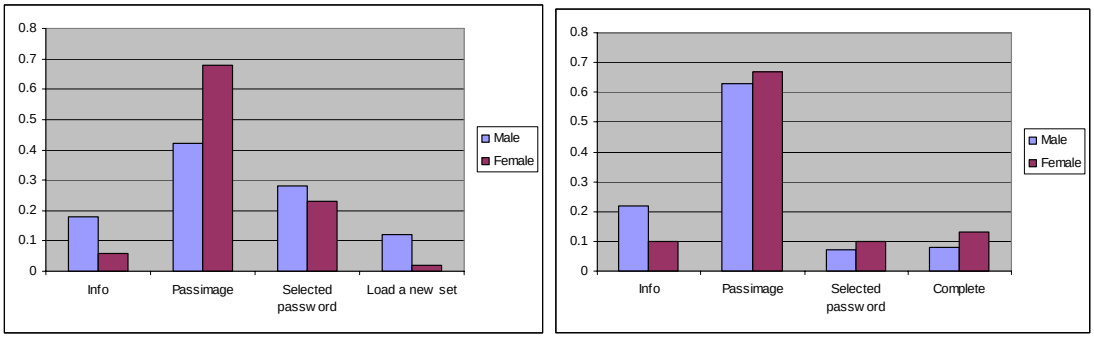


Figure 5.2: *Attraction plots for the Password selection screen (left) and the Password confirmation screen (right).*

Expectantly, during password selection the Passimage area received the most consideration with 55% attraction, followed by the Selected password box with 25% attraction, which could indicate a difference between genders for fixations on the Password area of the screen, and the Passimage area of the screen. A feature visibility analysis shows a lack of prominence of the Load a new set of images button with female participants. Although none of the participants were specifically asked to use this option, the existence of the feature was acknowledged only by the male participants. All of the participants were notably attracted to the Info region thus validating its existence as well as the selected position for enrollment instructions.

During the password confirmation task, the Passimage area received most consideration with 65% attraction. This is an indication for difference between genders for fixations on the Password section of the screen. A feature visibility analysis shows that the Selected password area received less attraction, 8% overall, which is just 32% of the attraction of the same area in the password selection task. There is a variance for attraction to the Selected password area between screens, which suggests the possibility for removal of this interface element in future iterations of the application. The Passimage area as well as the Info area received similar scores for attraction as in the preceding screen, thus confirming the previous results.

To further analyze the divergence between male and female patterns, the respective heat maps during password selection can be observed to identify potential differences. In the sample representative heat maps presented in Figure 5.3, the male participant focuses only on several images while choosing the password, while the female attempts to consider all images before making a choice. Fixation durations in the Passimage area for females are longer thus indicating an increase in cognitive function.

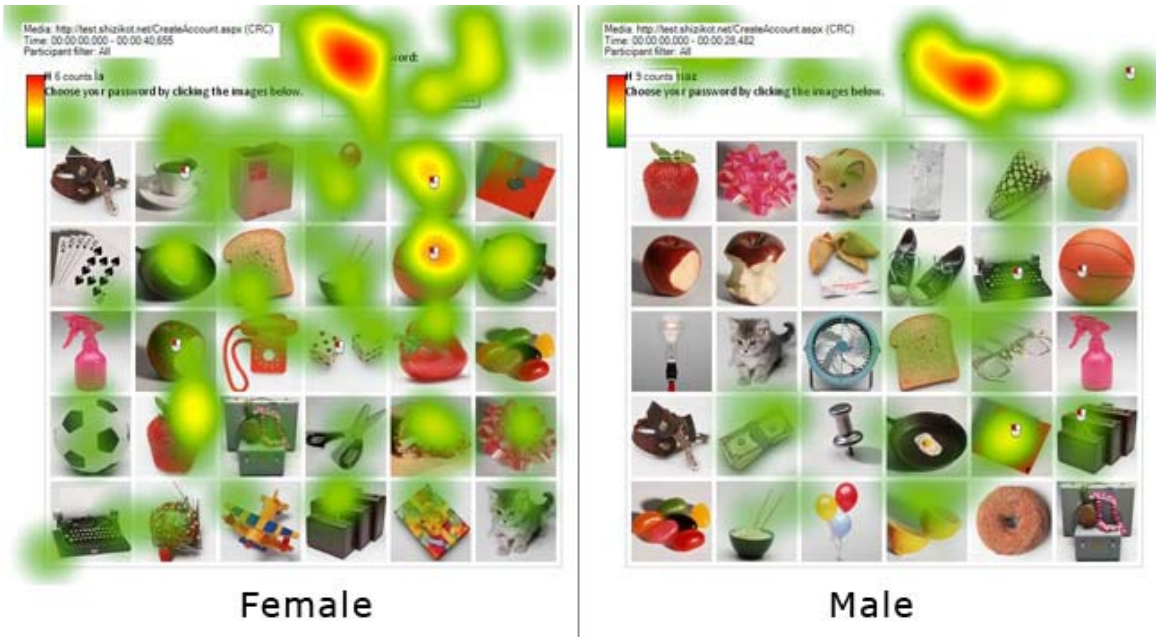


Figure 5.3: *Comparison of eye tracking heat maps between genders*

Insofar, the only explanation for this potential gender-specific result can be found in (Kimura, 1987) and (Humphrey, 1992). In these papers, the authors present differing male and female brain structures which can be evolutionary traced to hunter-gatherer roles. These established brain patters makes them process visual information differently. This difference is not consequential in traditional text-based authentication as the input is not sensory. However, when it comes to the cognitive effort necessary for visual processing of graphical password this distinction might be significant.

Considering gaze paths, in general, most of the participants followed the same pattern. As shown in Figure 5.4, they initially look at the Info area in the upper left corner and scan through the provided instructions. Their focus then drops to the top center of the Passimage area, followed by a successive shift between Passimage and Selected password areas as the password is entered. During the Password selection objective of the enrollment task the clicking on images starts after more than half the time has passed.

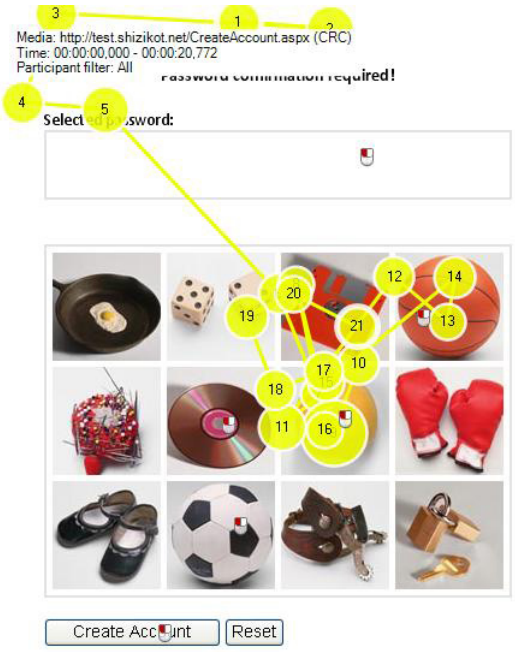


Figure 5.4: *Sample scanpath for ImagePass*

Another interesting observation is that during password selection all participants almost completely ignored the images in the first column, and paid little attention to the images in the last row. This would imply a necessity for an interface change regarding the display of the password selection grid to improve the effectiveness of the system. This could be achieved by either moving its starting position further to the right to align the grid with the gaze path, or adapt the dimensions of the grid by eliminating rows and columns.

5.3 Summary

This chapter presented the first part of the evaluation study preformed on the initial ImagePass working prototype. The usability study was a long-term analysis of ImagePass deployed in a real environment. It focused on evaluating user performance based on frequency of use and training as well as the memorability of the graphical password in general.

The eye tracking analysis offered insight into the initial perception and user behavior during interaction with an object-based graphical authentication mechanism. With some design changes the ImagePass concept was shown as potentially feasible as enrollment to the system was completed relatively fast, in roughly a minute. None of the participants had difficulties with the tasks and all of them completed the task objectives easily and efficiently. There are also potential differences in how male and female users observe the system, which, however, should be further researched before any specified claims are made.

The design and functionality of ImagePass was significantly modified after the first evaluation phase in order to address the discovered issues and prepare the system to collect more relevant data. This modification is explained in more detail in section 4.4 from the previous chapter. The next chapter presents a more detailed eye tracking study with a larger and more varied user base. It evaluates the users' gaze patterns and analyzes the password selection process in more detail to evaluate whether image selection is influenced by image position.

The findings of this research complemented with the findings from the second

evaluation phase, presented in the next chapter, help for determining design guidelines and developing a functioning prototype for a ubiquitous ImagePass deployed on mobile devices.

.

6 Evaluating Graphical Authentication, Part 2

The preliminary eye tracking experiment, presented in the previous chapter, was an exploratory study that investigated the user acuity of the ImagePass system. It focused on testing the usability of the generic components and on defining the initial user behavior when interacting with graphical authentication. The main eye tracking experiment, presented in the first section of this chapter, was a more detailed study with a wider goal. A more relevant and larger group of users was eye tracked during subsequent login attempts over different periods of time to analyze the cognitive user perception with continuous use of the system. Emergent scan paths were clustered to determine viewing user profiles. Different areas of the password selection grid were also analyzed to evaluate whether image selection is influenced by image position.

The second experiment, analyzed the properties of the ImagePass graphical password through two phases. First, the general properties of a single-object image were defined through an online task-oriented effort of naming and categorizing different sets of images. Then, through an online study, the specific properties of the graphical password selection process were scrutinized, by tagging images by color, shape and category. In addition, the effects that graphical password length has on memorability were also analyzed. Finally, the gender role of the participants was also observed in order to discover potential differences in this cognitive process.

6.1 Experiment 5: Eye Tracking Graphical Authentication

6.1.1 Introduction

Eye tracking adds another dimension to usability testing. It enables the understanding of finer points about what draws the users' attention and why. By analyzing what people observe during their task performance it is possible to gain insight into how the user works through usable and unusable designs. In addition, it also helps the facilitator to avoid interrupting the user during the experiment, when the user doesn't offer an otherwise recordable account for the performed actions.

6.1.1.1 Scanpath Theory

The scanpath theory, considered as one of the most influential theory of vision and eye movements, was defined David Noton and Lawrence Stark who examined the eye movement of persons repeatedly viewing the same line drawing (Noton and Stark 1971a, 1971b). They discovered that the initial eye movements on first stimulus exposure were often repeated on subsequent exposures. Gaze sequences of a given person for a given stimulus resembled each other while gaze sequences were dissimilar between persons for a specific stimulus and between stimuli for a specific person. The reason for the within-person similarity results comes from the user checking a stored mental model in order to match the currently observed image. A proposal was made that the representation of visual information in memory is an alternating sequence of sensory and motor memory traces. This sequence of memory traces is laid down during initial viewing and run as a

control program on subsequent exposures to facilitate recognition. The person's idiosyncratic scan pattern recurring from these subsequent exposures was termed a scanpath.

Generally, scanpath theory explains the vision process in a top-down fashion by proposing that an internal cognitive representation controls not only the visual perception but also the related mechanism of active looking eye movements. Evidence supporting the scanpath theory comes from experiments showing the repetitive and idiosyncratic nature of eye movements during experiments with ambiguous figures, visual imagery and dynamic scenes (Brandt & Stark, 1997; Pieters et al., 1999). Further support for scanpath theory comes from findings that scanpaths occur even when persons only imagine a previously seen stimulus (Gbadamosi & Zangemeister, 2001; Laeng & Teodorescu, 2002). More recent scanpath experiments performed using different motor read-out systems have served to better understand the structure of the visual image representation in the brain and the presence of several levels of binding (Privitera, 2006).

On the other hand it has been proposed that the dominant role of top-down, memory-guided control of saccadic eye movement is a drawback of the scanpath theory (Walker-Smith et al., 1977). Also, it has been noted that there are additional factors which influence eye movement, in particular, the task in which a person is immersed has shown to influence scanning (Grier et al., 2007), demographics, stimulus familiarity, and individual differences (Wedel & Pieters, 2006).

Although based on tests with simple stimuli such as line drawings, letter grids or irregular checkerboards, the applicability of the scanpath theory has been investigated on more complex and practically relevant material. One such study is a repeated presentation of print advertisements to participants where although the duration of attention paid to ad elements decreased from presentation to presentation, the order of scanned elements remained largely stable (Pieters et al. 1999). In (Foulsham & Underwood, 2008) the authors found evidence for scanpaths when repeatedly viewing photographs of natural scenes. Josephson and Holmes have tested the scanpath theory using web pages as stimuli (Josephson and Holmes, 2002b). They discovered clusters which included pairs of sequences from the same participant, but also observed similar sequences coming from different participants.

6.1.2 Research Questions and Hypothesis

Statistical findings from Experiment 3 showed significant difference between genders only for enrollment time. On the other hand the preliminary eye tracking study on the first ImagePass prototype produced initial observations that pointed towards potential differences between male and female participants when using graphical authentication. It is possible that these differences have a more cognitive nature that affects the users' perception of the system and not the performance. Therefore, this eye tracking study also analyzes the patterns of users based on gender. In addition, as stated previously, graphical passwords have a more cognitive nature than other general online-tasks. The selection process in graphical authentication, especially when single everyday objects are used as parts of the authentication key, will have a personal component. This individuality would be based on the users' everyday habits, familiarity and interactions with the selected objects. Consequently, it would be expected that the users' scan path patterns will not follow the general scanpath theory.

Formally, the hypotheses of this study are stated as follows:

H9: There is an observable difference in perception of the graphical authentication system between male and female users.

H10: There is a difference in general scan path patterns between ImagePass and expected general scan patterns for web-based applications.

6.1.3 Participants and Equipment

According to Pernice & Nielsen (2009), data from at least 30 participants is required in order to have valid eye tracking results. Therefore, for the purpose of the experiment 33 participants, 18 male and 15 female were recruited through direct contact and advertisements. Demographically, the ages of the participants ranged from 19 to 32 with a mean age of 27.2 years. All of the participants had 20/20 vision and had either a medium or a high web experience. Regarding education, 54% had some college education, while 46% were college graduates or higher. None of the participants had any previous experience with graphical authentication or eye tracking. These demographics are representative for the potential target group that could use this type of authentication, participants that belong in either the generation X or the generation Y cohorts. Graphical authentication, like ImagePass, is an unfamiliar approach to authentication for the casual user. It is expected that users from a younger age group would be more receptive that users with long term experience. The number of discovered usability problems might have been greater if the user profiles were more diverse, however the more mature age groups are not considered as probable candidates.

Like the eye tracking experiment presented in the preceding chapter, all the sessions in this experiment were conducted in one location, the Laboratory for Open and Network Systems at the Jožef Stefan Institute in Ljubljana, Slovenia. Eye movements were collected with the Tobii T60 eye tracker sampling eye position at a rate of 60Hz, with a fixation time of 200ms, latency of 33ms and drift of 0.1 degrees. To display stimulus pages and to define regions of interest, the Tobii Studio 1.5.4 software was used. The content was viewed on a 17 TFT monitor with a 1024x768 resolution on a personal computer running Windows XP Professional. The eye tracker was successfully calibrated for all participants, thus complete eye movement data was recorded for all subjects.

6.1.4 Experiment Design

The redesigned graphical authentication mechanism, ImagePass (see Section 4.4.), was deployed on the same remote web server as the previous versions from March to June 2011. The experiment consisted of one enrollment session taking place in a laboratory, up to five authentication sessions performed over different time intervals, and a final authentication session which also took place in a laboratory. The enrollment session and the last authentication session were eye tracked in the laboratory, while the intermittent sessions were performed at the participants' convenience.

To avoid analyzing authentication as a primary task, all of the participants were misdirected, and informed that the purpose of the experiment is an analysis of their Internet search-behavior patterns. For this reason, an additional module was developed and attached to the ImagePass system. The module had superficial functionalities intended to simulate search behavior analysis. Predefined search tasks were loaded in the module in form of a question that could be answered with a multiple choice response. For each session four different search tasks were defined where the participant had to use a specific search service or visit a particular site, find the requested information and answer the multiple choice question. After each task was completed it would automatically disappear from the search tasks list (Figure 6.1.).

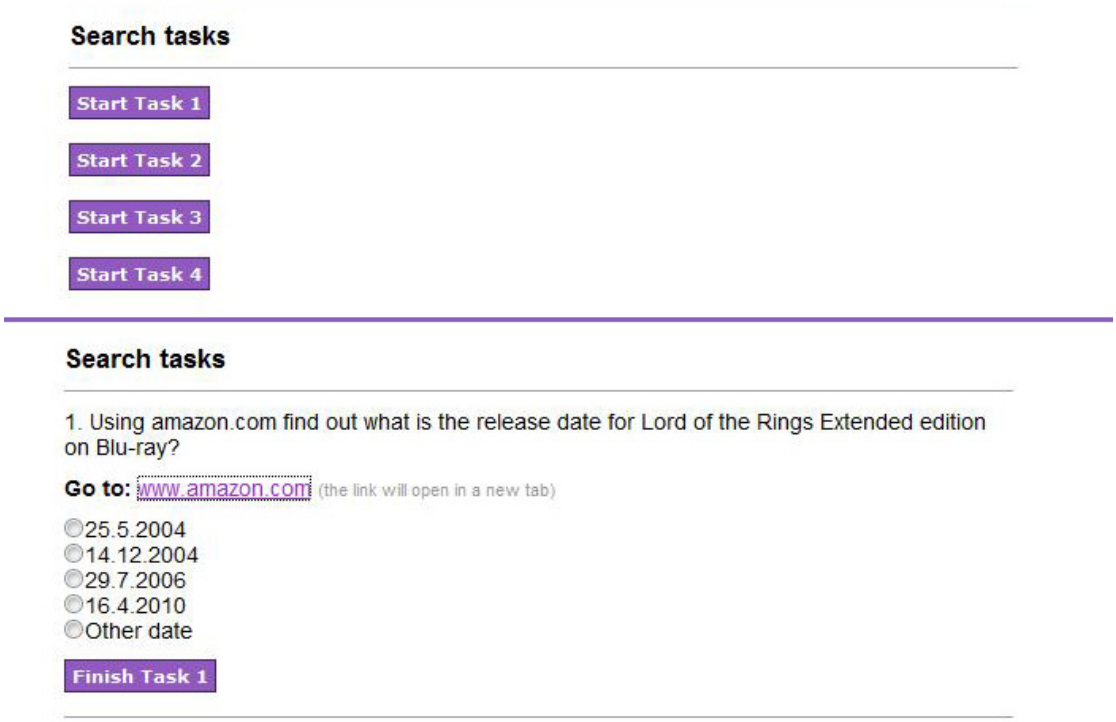


Figure 6.1: *Split screen for faux search tasks*

The data collected for this experiment was solely from the eye tracking sessions. No data was collected during the in-between authentication sessions; the purpose of these sessions was for the participants to continuously use the system. In the first eye tracking session data was collected from the all the Enrollment screens, while in the second eye tracking session (the final session of the experiment) data was collected only for the Authentication screen. Similarly to the preliminary eye tracking experiment the screens were categorized in areas of interest (AOIs) as presented in Figure 6.2. The most observed variable was fixation time on a specific AOI.

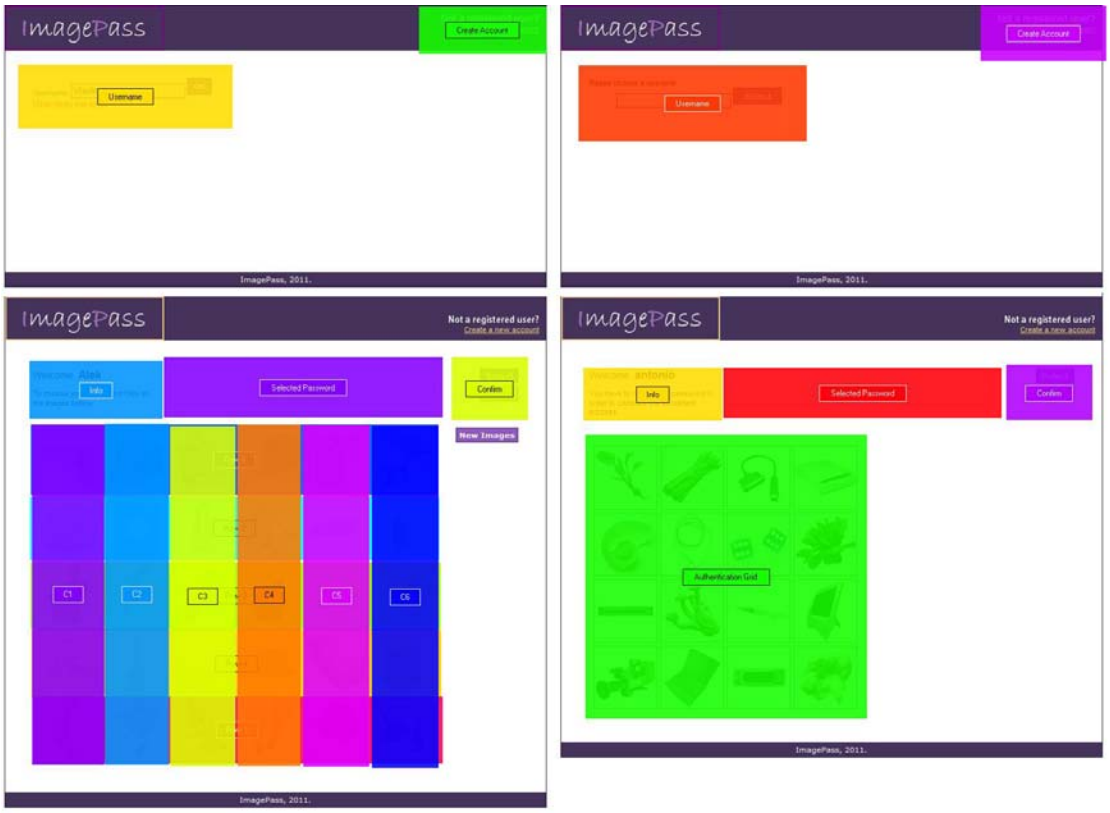


Figure 6.2: *Categorizations of visual areas on ImagePass screens*

The same AOIs were defined for the Login and the Choose Username screens. The area around the username text-field was defined as Username, and the top-right corner of the screen was defined as Create Account. The Select Password screen had the most complex AOI. Initially the AOI were divided as Info, the text area in the top left side of the screen, Selected Password, the area that shows the clicked images, Confirm, the area on the lefts side of the screen around the function buttons and Image Selection the area around the authentication imageset. However, in order to analyze the graphical password selection process in more detail, the Image Selection AOI was further subdivided into five rows and six columns. Finally, the Authentication screen was similarly divided into: Info, Selected Password, Confirm and Authentication Grid. The screen AOI categorization schemes used in the experiment were partially based on the web page object classification schemes used by (Nielsen & Tahir, 2002).

6.1.5 Procedure

The eye tracked enrollment behavior of the participants was analyzed through individual 20-25 minute sessions taking place in a period of 2 weeks. During each session, the participant was introduced to the nature of the experiment before signing a consent form. In addition, the participant filled out an interest questionnaire to determine their demographic characteristics.

In order to familiarize the participant with the eye tracking hardware a practice session was set-up in which the eye tracker was calibrated to the participant's eye movements and a photographic test image was shown to the participant for two seconds. In order to get more relevant results and treat authentication as a secondary task, the participants were intentionally misdirected that the purpose of the experiment is to analyze search behavior. It was explained that they will use an online application which tracks and logs all their activities, and that this application uses a graphical authentication mechanism instead of

traditional text-based passwords.

During the enrollment eye tracking sessions, the participants had to perform two tasks: enroll to the ImagePass system, and Complete Search Tasks. The enrollment task was subdivided into three subtasks with each subtask taking place on a different screen interface: create account, create username and choose graphical password. The second task was to complete 4 search tasks where the participants had to find some specific information on the World Wide Web through a particular search service and answer a multiple choice question regarding that information. For analysis, the data from the "dummy" tasks was discarded as it was irrelevant to the authentication process. Over the following period of 4 weeks, the participants were asked to login to the system remotely on a weekly basis and complete a new set of search-based tasks. After the remote sessions they were invited back to the laboratory for the final task, where essentially their behavior was eye tracked after continuous use of the system. For the final session the participants had to perform two tasks: authenticate to the ImagePass system and Complete Search tasks. The data for the search tasks was again discarded as irrelevant.

To analyze the eye tracking data the individual users' viewing behaviors were analyzed through gaze replays and by observing the respective gaze plots. In addition the fixation times on predefined areas of interest were also evaluated in order to determine specific user scan path profiles and gender specific behavior.

6.1.6 Results and Discussion

To evaluate the authentication process, all of the eye tracking sessions were analyzed separately. The potential differences between male and female participants were noted as observations. No statistical tests were used to evaluate the differences in their performance as a gender divided sample would be too small for relevant results.

The Login screen has 2 defined AOIs. The first area, Create Account, is in the top-right corner of the screen and represents the target area that the participants' focus should follow and use. The second area, Username, is the username box which should be focused only if the participant is a registered user of the system. The descriptive statistics for both areas are given in Table 6.1.

Table 6.1: *Descriptive statistics for fixation lengths on AOIs for Login screen*

	Total			**Male**			**Female**		
	Create User	**User name**	**Not on AOI**	**Create User**	**User name**	**Not on AOI**	**Create User**	**User name**	**Not on AOI**
M	2.67	4.62	4.32	3.47	3.70	1.96	1.66	5.74	7.16
SD	1.47	4.07	3.18	1.18	4.48	0.44	1.52	4.11	3.75

The tested participants spent more time on the Username AOI (M=4.62, SD=4.07) than the Create User AOI (M=2.67, SD=1.47). This difference is observably larger for female users (M=5.74, SD=4.11, and M=1.66, SD=1.52) than male users (M=3.70, SD=4.48, and M=3.47, SD=1.18). By observing a representative gaze plot (Figure 6.3) it can be noticed that the participants start viewing the screen near to the Username AOI, move their focus towards the Create Account AOI and then most of them fall back to the Username AOI and start registering to the system by entering a new username. Expectantly, most of the participants make a brief notice of the logo area some participants notice the footer areas.

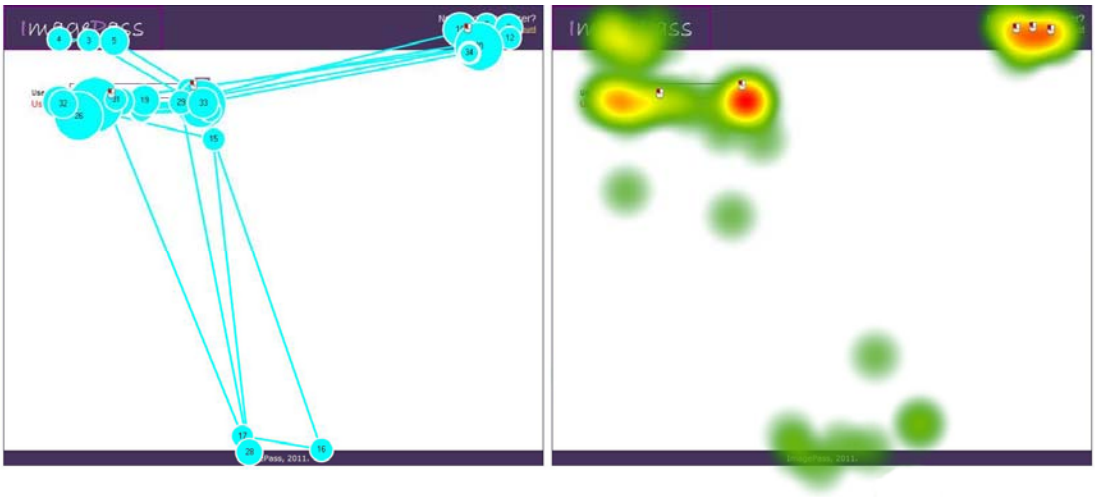


Figure 6.3 *Login screen sample gaze plot & aggregated heat map*

The current positioning of these two areas seems to have a slightly confusing effect on the user. Many of the participants make an attempt to enter their username in the Username field, although they previously notice the Create Account area. Only after they are prompted with the "User does not exist" text appearing bellow the Username text-field the participants' gaze is guided back towards the Create Account AOI. This might be registered as a usability issue as the user is not obviously directed to create a new account. To resolve this issue, the "Not a registered user / Create a new account" text/link could be repositioned bellow the username field and the user could be prompted to create a new username if the entered username is not available to the system.

The observed participant behavior on the Create Username screen is within the expected range. The users' gaze starts from the Create Account AOI as that is where the user was last looking on the previous screen before clicking the Create a New Account link. The gaze quickly shifts to the Username AOI (M=0.379s SD=0.223) and the participant engages in creating a new username (Figure 6.4).

Figure 6.4: *Create username sample gaze plot & aggregated heat map*

The analysis of the data for the Graphical Password Selection screen yielded more complex results. On average participants spent 35.56 seconds (SD=4.67) on selecting their graphical password, without any noticeable differences between male and female participants. Male participants paid more attention to the Selected Password AOI

(M=3.84, SD=1.93) and Info AOI (M=4.95, SD=4.77), than female participants (M=2.99, SD=3.10, and M=2.30, SD=1.11). On the other hand, female participants spent more time observing the Image Selection AOI (M=27.61, SD=2.78), than male participants (M=22.59, SD=3.44). The details of the descriptive statistics are given in Table 6.2.

Table 6.2: *Descriptive statistics for fixation lengths on AOIs for Select Password screen*

		Not on AOI	Row 1	Row 2	Row 3	Row 4	Row 5	Selected Password	Info	Confirm
Total	M	1.39	7.05	5.22	4.36	3.76	3.64	3.59	4.19	2.36
	SD	0.64	4.75	2.04	3.35	3.15	3.09	2.06	4.13	1.45
Male	M	1.45	6.62	5.29	3.96	3.24	3.48	3.84	4.95	2.50
	SD	0.71	3.51	2.40	3.87	3.67	3.77	1.93	4.77	1.00
Female	M	1.24	8.12	5.05	5.35	5.05	4.04	2.99	2.30	2.03
	SD	0.65	9.10	1.34	2.20	1.07	0.20	3.10	1.11	2.87

The Image Selection AOI was subdivided into 5 rows in order to analyze the data more precisely. There is a decrease of attention as the user looks further down in this AOI, with a higher drop between the first and second row and lower drops for the following rows (Figure 6.5). This would imply that users are more likely to select images from the higher than the lower rows.

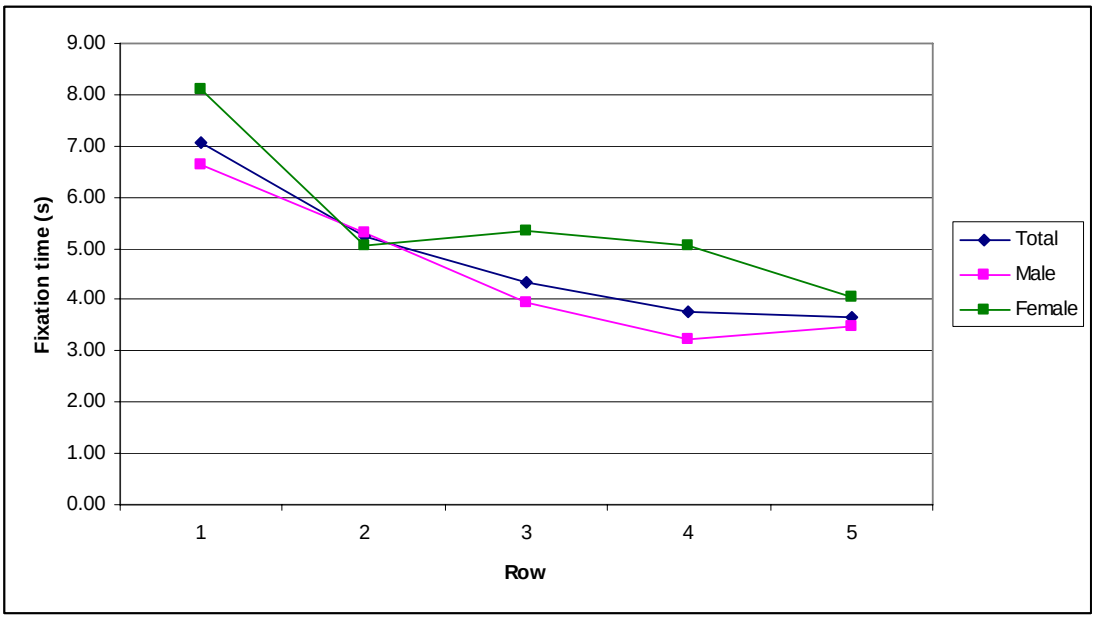


Figure 6.5: *Number of fixations per row*

When the Image Selection AOI is subdivided into 5 columns the results are slightly different. There is a slight up-down variation as the participant observes the first three columns, before the attention start decreasing for the last three columns. When this attention is subdivided by gender groups the results show a different pattern for the first three columns. Female participants pay more attention to the first column and then their attention drops for the remaining columns. However, male participants have an increase in attention from column 1 to column 3 with a sharp drop for the remaining three columns (Figure 6.6).

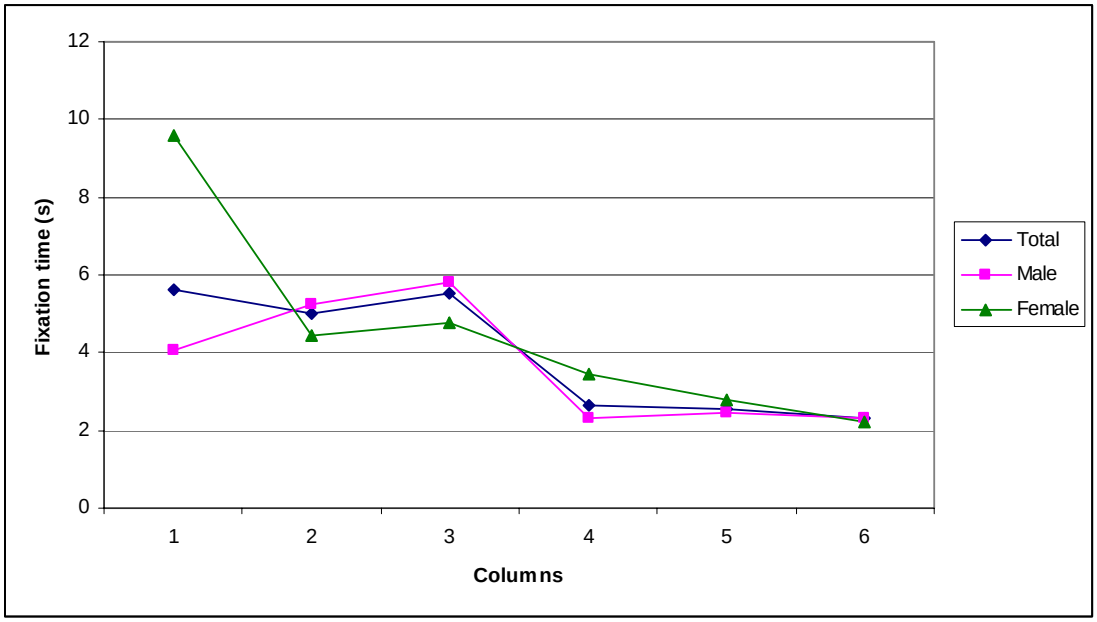


Figure 6.6: *Number of fixations per column*

Generally, the first fixation for all of the participants was within the Info AOI. They would then proceed with selecting the graphical password by fixating on the Selection Grid AOI. While selecting the graphical password, the participants would notice the clicked images appearing in the Selected Password AOI. They would also notice the New Images button however the button remained unclicked for all the testing sessions. Finally, the focus would fall within the Confirm AOI as they click the Select button to finish the selection process. For the selection of the graphical password itself, the following two general user profiles can be determined:

'Get Me Out of Here' Profile. In this profile the user views graphical, and probably all authentication mechanisms as a potential nuisance. The graphical password selection grid gets a perfunctory glance, and the graphical password itself is selected quickly, usually by either repetitively clicking one image, or alternating the clicks between two images. The location of the selected images in the graphical password selection grid is within the first three rows and the first three columns with the other images being almost completely ignored. This profile is more likely to belong to a male rather than a female participant (Figure 6.7).

Figure 6.7: *Sample gaze plot and aggregated heat map for the GMOH profile*

'Let Me Think' Profile. In this profile, the user indulges into graphical password selection more carefully. The images in the graphical password selection grid are viewed longer and more attentively. The selection of the images constituting the graphical password is more varied including more rows and columns from the selection grid, although there is still a preference for higher rows, of left columns. There is either no repetition of images or one repetition of an image in a few cases in the selected authentication key. This profile is more likely to belong to a female rather than a male participant (Figure 6.8).

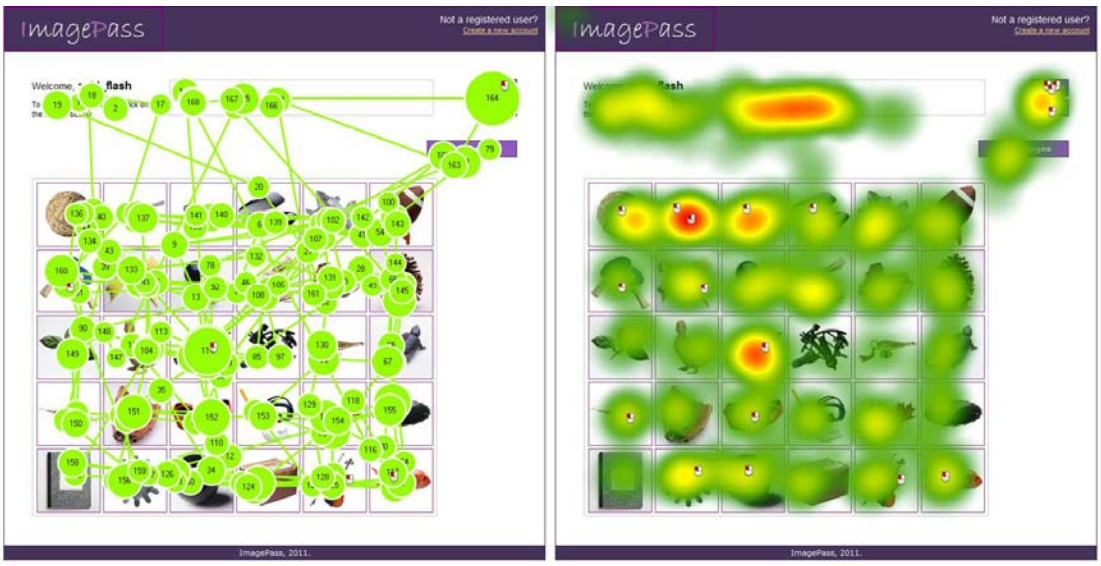


Figure 6.8: *Sample gaze plot and aggregated heat map for the LMT profile*

As mentioned previously, the authentication process was eye tracked 1 month after the enrollment to the ImagePass system. Most of the participants followed the protocol of accessing the system on a weekly basis to perform Internet search task and answer the related questions. A few participants didn't adhere to the procedure and had authenticated either once or zero times to the system and had difficulties remembering the graphical password. There were no observable differences between male and female participants. Based on the data analysis the following two user profiles can be observed during authentication:

'No Problem' Profile. In this profile the user has no difficulty in recognizing the graphical password and uses that graphical password to authenticate to the system. As the page loads they immediately focus on the Authentication Grid AOI and quickly scan the presented images until they focus on the first image from their authentication imageset. Depending on the grid positions of the first image, the user recognizes and clicks the image within 10 to 20 fixations from the initial page load. When the initial image is in the lower part of the identification grid the user temporarily fixates on subsequent images belonging to the authentication imageset before focusing and clicking on the first image. Occasionally, the user fixates on the Selected Password AOI to check how many images have been clicked. With a sporadic glance to the Info AOI the user finalizes the authentication process by focusing on the Confirm AOI and clicking the Login button (Figure 6.9).

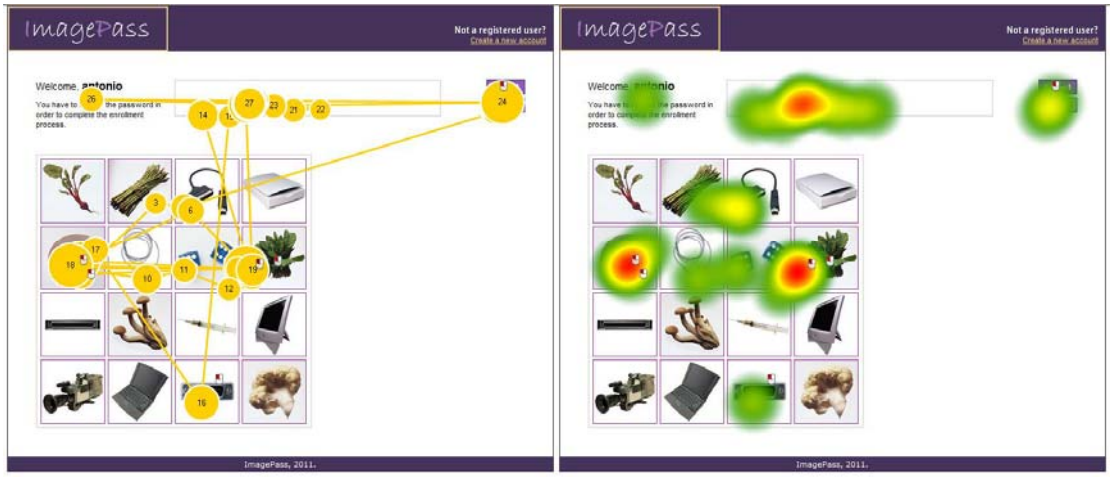


Figure 6.9: Sample gaze plot and aggregated heat map for the NP profile

'What Was It' Profile. In this profile the user doesn't recognize the graphical password immediately and in this case the observed behavior is expectantly more erratic. The user focuses on the Authentication Grid AOI and carefully scrutinizes all the images in order to recognize their corresponding authentication set. After the wrong graphical password is entered, the user starts shifting the focus more chaotically, between the Authentication Grid AOI, the Selected Password AOI and the Info AOI. In the meantime, a variation of the images in the imageset is clicked. In the end the user either recognizes the graphical password after few attempts or in the two instances where the graphical password was not correctly recognized the user opted for the “Create a new account” option as password recovery was not available (Figure 6.10).

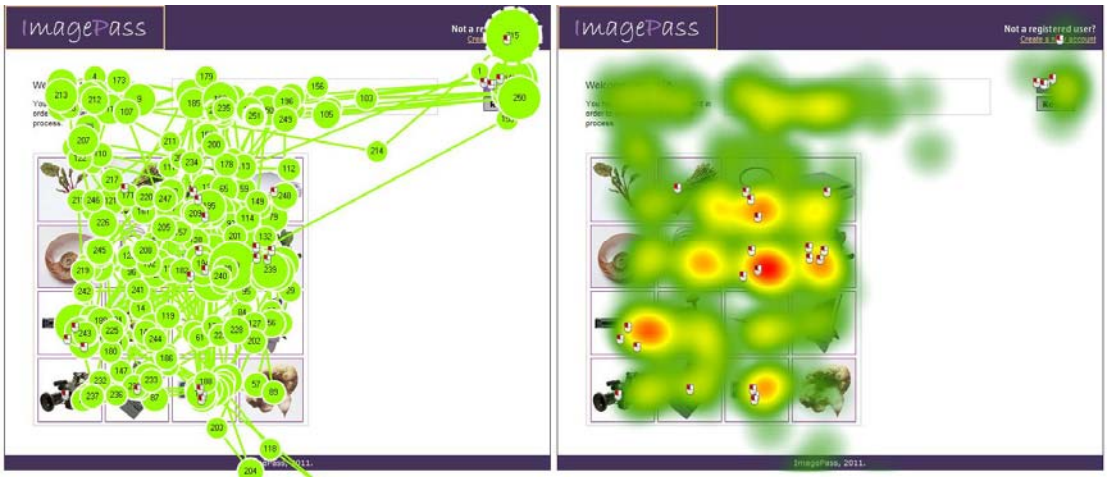


Figure 6.10: *Sample gaze plot and aggregated heat map for the WWI profile*

Generally, the results of the experiment reinforce both proposed hypothesis. The differences in perception of the ImagePass graphical authentication mechanism are observable only during enrollment and are not evident during continuous authentication to the system, which is in support of hypothesis 9. In addition, the existence of different profiles during enrollment and authentication is in support of hypothesis 10, as scanpath theory requires the presence of repetitive patterns which are dissimilar between persons for a specific stimulus and between stimuli for a specific person.

The next and final experiment of this research on graphical passwords, as presented in the next section, will focus on defining the properties of the ImagePass graphical password.

6.2 Experiment 6 – Graphical Password Selection Properties

6.2.1 Introduction

The processing of images with words has been compared in many different tasks and for many different reasons. These tasks can be classified into two broad categories based on memory access:

- **Semantic memory tasks** which include naming images with words, generating images to words, categorizing pictures or their names, and making symbolic comparison judgments about whether X has more of some property than Y.
- **Episodic memory tasks** which include verbal learning and memory paradigms, like free recall, serial recall, paired-associate recall, yes/no recognition, forced-choice recognition, etc.

For this brief review of empirical differences between images and words it is necessary to consider two semantic memory tasks, naming and categorization.

Naming. Certain characteristics of the image name, such as frequency in print, have much larger effects on the naming of images rather than on reading the image name (Oldfield & Wingfield, 1965). However, there is disagreement about which of several variables correlated with frequency in print is primarily responsible for the differences in image-naming. Generally, the age of acquisition of a concept's name is more predictive in naming than frequency (Carroll and White, 1973b). Lachman (1973) statistically measured the degree to which subjects agreed on the name of the image and found that

there is a significant amount of variance in naming. To understand the image-naming process, one must be familiar with at least the following characteristics of each image: the frequency of its potential name as found in print and the age of the conceptual acquisition of that name.

Word frequency is an important variable in cognitive processing. High-frequency words are perceived and produced faster and more efficiently than low-frequency words. At the same time they are easier to recall, but more difficult to recognize in episodic memory tasks. To match stimuli on word frequencies, it is necessary to use estimates on how often words occur in a language. The current norms for word frequencies in the English language are based on Brysbaert & New (2009) and are based on 51 million words as measured on the basis of English subtitles. Although still used in current research, previous word frequencies such as Kucera-Francis (1967) and Celex (Baayen et al., 1993) have been critiqued for being of bad quality (Zevin & Seidenberg, 2002) and not optimal (Balota et al., 2004).

Categorization. Categorization tasks in which the meaning of images or their names must be retrieved from memory have been used to study the question of whether images and their names are accessed through a common semantic representation. In categorization tasks, there is ample evidence that pictures are categorized faster than their names. For example, in a yes/no task in which the category is given ahead of time, images were categorized either faster than words (Potter & Faulconer, 1975) or with the same speed (Smith & Magee, 1980). In a same-different task, when two instances of images or words in a same or different category were presented simultaneously, match or mismatch decisions for images were faster than those for words (Pellegrino, et al., 1977). Comparisons between categories of images and word forms have been made with the deference leaning towards image-image and word-word comparisons. This data can be interpreted as indicating that images and words access a common semantic representation. In essence, images would take longer to categorize than names because naming images takes longer than naming words (Pellegrino et al., 1977).

The image names and categories in this experiment were defined on the basis of the following criteria: they are unambiguous, they include exemplars from widely used naming norms (Brysbaert & New, 2009) and they represent concepts at the basic level (Rosch et al., 1976). The first criterion, unambiguity, refers to the degree to which subjects will show consensus about the name to be given to the image. The second criterion uses the afore mentioned norms in order to assess the naming frequencies. The third criterion refers to the simplest albeit most specific nomenclature that can be chosen for describing a specific image.

6.2.2 Research Questions and Hypothesis

To complete the evaluation process of ImagePass, the structure of the graphical password had to be analyzed. This experiment focuses specifically on the impact of user selection of ImagePass graphical passwords and the security of those passwords. In particular, this experiment defines the properties of the ImagePass graphical password. All the images are tagged based on several properties and the goal of the experiment is to identify factors which could potentially become sources of bias in graphical password selection. A natural question is whether general perceptions and preferences of users influence the graphical password selection process when that process involves images with single everyday objects as content.

By discovering the optimal graphical password properties through preferences in selecting image content, optimal password length (the length above a minimal security threshold that doesn't significantly affect the usability of the system.), and grid selection

hotspots, the results of the experiment show that ImagePass graphical passwords have some potential bias drawbacks similar to alphanumeric passwords. Users have tendencies towards specific types of objects with a specific color and shape, which in certain cases can be gender specific.

The formal statements of the hypothesis are as follows:

H11: Graphical password selection is affected by the category of the object representing the image content of the authentication key.

H12: Graphical password selection is affected by the color of the object representing the image content of the authentication key.

H13: Graphical password selection is affected by the shape of the object representing the image content of the authentication key.

H14: There is a user preference for selection in grid based displays based on image location.

H15: There is no difference in performance for graphical passwords for self-selected passwords of 4, 5 or 6 images.

6.2.3 Participants and Equipment

A total of 302 participants, 163 male and 139 female were recruited for the main part of the study. Their age varied between 18 and 35, with a mean age of 26.2. The participants were either casual or experienced computer users who use the Internet for study and leisure purposes on a daily basis. The recruitment process was coordinated with a computers education centre that teaches different computer skills courses on all levels.

The ImagePass application was deployed on a HP Proliant DL 585 G6 Server with a Microsoft Windows Server 2008 R2 operating system, SQL Server 2008 Developer Edition database engine and IIS 7.5 web server running .NET Framework 4.0. The experiment was completed in two computer labs at the Faculty of Economics in Skopje and 5 computer classrooms at a computer education centre. Approximately 78 PC's running Windows XP Professional or Windows 7 on Pentium dual core processors with 15” and 17” LCD monitors set on a 1024x768 resolution or higher were used. Each PC had Internet Explorer 7, Mozilla Firefox 3 and Google Chrome 8 web browsers installed, with the web browser choice left to the discretion of the participant.

For the preparatory phase of the study, 230 participants participated in completing predesigned Amazon Mechanical Turk tasks. They were paid $0.01 for each processed image and a $0.50 bonus if they completed all the images.

6.2.4 Experiment Design

For the preliminary part of the experiment the subjects had to name the images in order to determine the images' most common name and the degree to which the subjects agreed on that name. Then the subjects had to define a concept that would represent the most common name for a specific image, thus defining the image category. The subjects also had to rate the familiarity and visual complexity of the image as an inappropriate score might affect the memorability of the image. If the average score showed either low familiarity or high complexity the image was discarded. All the tasks were accomplished with the Amazon Mechanical Turk service.

For the main experiment, a web application that used the ImagePass authentication mechanism was deployed online to a public IP address and was made accessible through a specific domain name. Provisionally, this experiment was divided in two phases:

Enrollment and Subsequent use. To test hypothesis 13 and study the effects of graphical password length on memorability, the participants were divided into three similarly-sized groups, based on the minimal password length allowed for system enrolment. Each of the groups had 4, 5 or 6 images as minimal requirements for enrollment. After enrollment, subsequently, the participants accessed the system once a week from varied locations for approximately four additional sessions. To test hypothesis 11 the images were named, classified in 15 different categories and tagged according to its dominant color(s) and shape in accordance with the results from the Mechanical Turk results. To test hypothesis 12 the positions of the graphical password images were recorded during enrolment and the click frequency for each image position was analyzed. The data for the main experiment was collected through preprogrammed system logs that silently collected data in the background by registering left clicks, page loads and time intervals between events. For the analysis each log was classified based on the active screen and the clicked page item.

6.2.5 Procedure

For the preliminary part of this experiment Amazon Mechanical Turk was used to run the image naming procedure. The images were presented as a single object on a white background with each image sized at 100x100 pixels and presented at the center of the screen. The image would be presented on screen for a period of 4 seconds and then subjects had to respond to the image by filling out data in several text fields. They had to respond to every presented image and could leave no blanks. The participants were instructed to identify each image as briefly as possible by typing the first name that came to mind. They were told that the name can consist of no more than two words. If the image was of an object they couldn't recognize or couldn’t think of a name, the participants were instructed to respond with NO. The familiarity of each picture was judged as the degree of how common the object is in the users' experience. The rating was done on a 5 point Likert scale with 1 indicating very unfamiliar and 5 indicating very familiar. The visual complexity of the image was also rated on a 5 point Likert scale with 1 indicating very simple and 5 indicating very complex. Finally, they had to tag two different image properties: color and shape. From the total of 230 workers that performed the tasks on Mechanical Turk, most of the data came from 187 participants.

The main part of this experiment was similar in nature to experiment 3, user evaluation and took place between February and March 2011. The enrollment phase to the ImagePass system was completed in three sessions over two days in a supervised environment with each session lasting on average around 20 minutes. Every participant had to enroll to the ImagePass system by creating a unique username, selecting a graphical password, confirming the graphical password and entering some personal data. Participants from each session received the same instructions about graphical passwords, the only difference being the minimal graphical password length required. The continuous phase of the experiment was completed at different time intervals over a period spanning approximately 6 weeks. To avoid the analysis of authentication as a primary task, study materials for a single course that the participants were following were published online on weekly basis, and made available only through successful authentication through ImagePass. The availability of online materials was announced to all participants via email.

Data from some of the participants in the second part of the experiment had to be discarded as it was unusable for analysis. Most of the discarded data was from participants that did not continue to use the system after enrollment. From the remaining data five participants didn’t define their gender when entering personal data, hence in the

results presented in this section there is a small discrepancy between the values when all of the participants are observed and the values when male and female participants are observed separately.

6.2.6 Results and Discussion

6.2.6.1 Naming and Categorization

The information statistic, H, was calculated for each image by the formula

$$H = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$$

where k is the number of different names given to each image and pi is the number of subjects giving each name. An image that receives the same name from every participant has an H value of 0, indicating perfect name agreement. Conversely, an image which had elicited two different names with equal frequency would have an H value of 1, etc. Increasing H values indicates decreased name agreement as the percentage of the subjects who gave the same name is decreasing.

The criteria for counting the different instances of names were as follows:

- If the name given by the subject was similar to the established name in cases such as misspellings, abbreviations or multiple names, this was included as correct (ex. TV was treated the same as television).
- When two words were used for describing an image this was treated differently than the name using just one of the words (ex. red apple <> apple)
- When two distinct names were written down for an image, only the first name was counted.

Table 6.3 presents the summary statistics for all the measured and relevant variables.

Table 6.3: *Summary statistics for experiment variables*

	M	SD	Skew
Name agreement (H)	0.558	0.526	1.50
Familiarity (F)	3.29	0.956	0.93
Visual complexity (C)	2.96	0.897	1.02
Frequency (B-N)	37.86	88.09	2.00

The H value captures more information about the distribution of names across subjects. The results show a low mean of the H value with a positive skew which can be interpreted that many concepts showed high name agreement. When observed individually, many of the concepts had either a perfect name agreement, an H value of 0, or a high agreement, an H value bellow 0.2. The measures based on ratings, familiarity and complexity, showed a greater range of values reflecting greater consensus among subjects on the extreme of the scale. Complexity ratings are fairly symmetric around the midpoint scale value of 3, while familiarity ratings tend to be negatively skewed. This shows that most images were judged with average complexity and most of the images

were viewed as familiar. The Brysbaert-New frequencies, which were available for most of the images, are positively skewed because of the few high-frequency concepts.

To determine the degree of relationship among these measures, the inter item correlations were computed among all attributes. Table 6.4 presents the matrix of significant correlations for all concepts

Table 6.4: *Significant correlations among the measured variables*

	Familiarity	Complexity	Frequency
Familiarity	1.000	-0.466	0.363
Complexity	-0.466	1.000	-0.180
Frequency	0.363	-0.180	1.000

The correlations among the three measures collected in the present study are all relatively small in magnitude. Familiarity is positively correlated with frequency (r=0.363), but the correlation is modest in size. The reason for this low correlation is because frequency values do not distinguish between different meanings of a word. Visually complex pictures tend to have many names and tend to be rated as unfamiliar. Hence there is a negative correlation between familiarity and visual complexity (r=-0.466). The amount of detail in an image is determined primarily by certain characteristics of the object. Consequently, the two sources of the effects of visual complexity on other variables are the inherent properties in the object itself and the properties inherent in the grammar of its representation.

Based on the performed analysis and the aggregated results, images with a high complexity rating and/or low familiarity were discarded from further use, while all the other images were named and tagged with the appropriate category, color and shape. The final tag names are presented in Table 6.5 and used in the subsequent analysis of graphical password properties.

Table 6.5: *Final categorization and tagging results*

Category	Color	Shape
Animals	Beige	Circle
Appliances & devices	Black	Irregular
Business	Blue	Oval
Clothes & Accessories	Brown	Polygon
Food & Drinks	Green	Rectangle
Games	Grey	Square
Home	Multicolor	Star
Kids & Toys	Orange	Triangle
Music	Pink	
Other	Purple	
Plants & Earth	Red	
Science & Technology	White	
Souvenirs	Yellow	
Sport		
Tools		

6.2.6.2 Category Analysis

The category property of the image was controlled during graphical password selection,

so the expected frequency for a particular image being selected is constant and equal for all categories. The actual results for image selection per category are presented in Table 6.6.

Table 6.6: *Distribution of selected images by category*

Category	Observed frequency			Expected frequency		
	Total	M	F	Total	M	F
Other	16	10	6	16.1	5.8	9.9
Animals	14	5	7	16.1	5.8	9.9
Business	25	8	17	16.1	5.8	9.9
Clothes & Accessories	20	5	15	16.1	5.8	9.9
Food & Drink	29	8	18	16.1	5.8	9.9
Games	32	25	7	16.1	5.8	9.9
Home	8	2	6	16.1	5.8	9.9
Kids & Toys	23	2	21	16.1	5.8	9.9
Music	11	6	5	16.1	5.8	9.9
Plants & Earth	26	1	25	16.1	5.8	9.9
Sport	8	3	5	16.1	5.8	9.9
Tools	7	2	5	16.1	5.8	9.9
Science & Technology	6	2	4	16.1	5.8	9.9
Appliances & Devices	10	5	5	16.1	5.8	9.9
Souvenirs	6	3	3	16.1	5.8	9.9

The chi-square test compared the observed and expected frequencies for each category. The results of the test show that there is a significant difference for the observed number of selected images from each category from an expected distribution when all the participants are observed, $\chi^2(1, N=241) = 71.261$, $p=0.000 < 0.05$, and when the participants are split on gender, $\chi^2(1, N=87) = 85.241$, $p=0.000 < 0.05$ for male, and $\chi^2(1, N=149) = 72.376$, $p=0.000 < 0.05$ for female. The detailed results can be viewed in Table 6.7.

Table 6.7: *Chi-square test statistics for category*

	Total	M	F
X^2	71.261	85.241	72.376
Df	14	14	14
p	0.000	0.000	0.000

To get more precise results for each category separately a binomial test compared the observed frequencies to the expected frequencies under a binomial distribution with an even probability parameter of 0.067. An exact binomial sign test on all the participants indicated that there is a significant positive preference for the categories: Business ($p=0.02$), Food & Drinks ($p= 0.002$), Games ($p=0.000$) and Plants & Earth ($p=0.012$), a significant negative preference for the categories Home ($p=0.017$), Sport ($p=0.017$), Tools ($p=0.008$), Science & Technology ($p=0.003$) and Souvenirs ($p=0.003$).

To analyze the preference of category during graphical password selection based on gender a Mann-Whitney test was performed. The test showed that the selection of images based on category between genders differed significantly for Games, $U = 4923.5$, $z=-5.193$-, $p=0.000 < 0.01$, $r=-0.34$, Kids & Toys, $U = 5717$, $z=-2.941$, $p=0.005 < 0.01$, $r=-0.19$, Plants & Earth, $U = 5468.5$, $z=-3.692$, $p=0.000 < 0.01$, $r=-0.24$, and Other, $U =$

5997.5, z=-2.197, p=0.034 < 0.01, r=-0.14. The detailed results are presented in Table 6.8.

Table 6.8: *Results from the Mann-Whitney test on category*

	Other	Animals	Business	Clothes & Acc.	Food& Drinks	Games	Home	Kids & Toys
Mann-Whitney	5997.5	6413.5	6338.0	6201.5	6294.5	4923.5	6369.5	5717.0
Z	-2.197	-.353	-.532	-1.147	-.681	-5.193	-.706	-2.941
P	.034	.764	.666	.335	.528	.000	.714	.005

	Music	Plants & Earth	Sports	Tools	Science & Tech.	Appliances & Dev.	Souvenirs
Mann-Whitney	6252.0	5468.5	6475.5	6413.0	6456.5	6326.5	6388.5
Z	-1.242	-3.692	-.038	-.461	-.181	-.878	-.674
P	.337	.000	1.000	.717	1.000	.505	.672

If the results are summarized the analysis shows that male participants have a greater affinity towards images from either the Games or Other categories than female participants. Conversely, female participants have a greater affinity towards images from the Kids & Toys and Plants & Earth categories. Both genders showed a higher preference for images from the Business and Food categories, and a lower preference for images from the Home, Sports, Tools, Science & Technology and Souvenirs categories. As there is an abundant difference across categories and gender, these findings are in support of hypothesis 11.

6.2.6.3 Color Analysis

The evaluation of the categorical data was straightforward as the images loaded into the selection grid were carefully controlled and there were always two images per category presented to the user during graphical password selection. On the other hand, the color and shape properties of the images were not controlled during graphical selection. Therefore, it is necessary to realign the observed frequency of a color being selected to the natural expected frequency. To get the expected frequency the observed frequency needs to be multiplied by the probability of a color appearing in the graphical password selection imageset (p_c). The results are presented in Table 6.9.

Table 6.9: *Expected and actual frequency of colors in images*

Color	Observed Frequency			Expected frequency			p_c
	Total	M	F	Total	M	F	
beige	28	8	17	13.4	4.8	8.3	0.056
Black	38	20	18	38.7	14.0	23.9	0.161
Blue	10	8	2	6.3	2.3	3.9	0.026
Brown	24	8	14	37.5	13.5	23.2	0.156
Green	17	6	11	12.6	4.6	7.8	0.052
Grey	27	12	15	32.8	11.8	20.3	0.136
multicolor	25	12	13	27.3	9.8	16.9	0.113
Orange	4	1	3	7.9	2.9	4.9	0.033
Pink	13	0	13	4.7	1.7	2.9	0.020
Purple	2	1	1	2.0	0.7	1.2	0.008
Red	22	4	18	19.8	7.1	12.2	0.082
White	11	3	8	14.6	5.3	9.0	0.061
Yellow	20	4	16	23.3	8.4	14.4	0.097

The chi-square test is performed to compare the observed and expected frequencies for each color. The results of the test are presented in Table 6.10.

Table 6.10: *Chi-square test statistics for colors*

	Total	M	F
χ^2	43.479	27.945	57.700
Df	12	12	12
P	0.000	0.006	0.000

The chi-square test shows that there is a significant difference for the observed number of selected images from each color from an expected distribution when all the participants are observed, $\chi^2(1, N=241) = 43.479$, $p=0.000 < 0.05$, and when the participants are split on gender, $\chi^2(1, N=87) = 27.945$, $p=0.006 < 0.05$ for male, and $\chi^2(1, N=149) = 57.700$, $p=0.000 < 0.05$ for female. To get more precise results for each color separately a binomial test compared the observed frequencies to the expected frequencies under a binomial distribution with the p_c probability parameter. An exact binomial sign test on all the participants indicated that there is a significant positive preference for the colors pink ($p=0.001$) and beige ($p=0.000$), a significant negative preference for the color brown ($p=0.007$), and no significant preference for other colors.

To analyze the selection of color based on gender a Mann-Whitney test was performed. The test showed that the selection of images based on color between genders differed significantly for black, $U = 5774.5$, $z=-2.195$, $p=0.004 < 0.01$, $r=-0.14$, blue, $U = 5972.5$, $z=-2.883$, $p=0.006 < 0.01$, $r=-0.19$, and pink, $U = 5916$, $z=-2.828$, $p=0.005 < 0.01$, $r=-0.18$. The detailed results are presented in Table 6.11.

Table 6.11: *Results from the Mann-Whitney test on color*

	black	blue	brown	green	Grey	orange	pink
Mann-Whitney U	5774.5	5972.5	6468.5	6450.0	6240.0	6425.5	5916.0
Z	-2.195	-2.883	-0.051	-0.139	-0.866	-0.495	-2.828
P	0.042	0.006	1.000	1.000	0.403	1.000	0.005

	purple	red	white	yellow	multicolor	beige
Mann-Whitney U	6450.5	5996.5	6357.0	6083.5	6153.0	6338.0
Z	-0.386	-1.903	-0.674	-1.631	-1.218	-0.532
P	1.000	0.065	0.548	0.145	0.274	0.666

To interpret the results, male participants showed a greater affinity towards images with either a black or blue color than female participants. Conversely, female participants showed a greater affinity towards images with a pink color than male participants. Both genders showed a higher preference for images with a beige color, and a lower preference for images with a brown color. These findings are in support of hypothesis 12 as there are some color-related differences and specifics for image selection in ImagePass.

6.2.6.4 Shape Analysis

When analyzing the shape properties of the selected images, similarly to the analysis of color properties it is first necessary to realign the observed frequency of a shape being selected to the natural expected frequency. To get the expected frequency the observed frequency is multiplied by the probability of a shape appearing in the graphical password selection imageset (p_s). The results are presented in Table 6.12.

Table 6.12: *Expected and actual frequency of shapes in images*

Shape	Observed frequency			Expected frequency			p_s
	Total	M	F	Total	M	F	
Square	26	13	13	12.2	4.4	7.6	0.051
Rectangle	49	15	31	72.7	26.2	44.9	0.302
Circle	35	21	14	20.5	7.4	12.7	0.085
Oval	34	7	25	39.5	14.3	24.4	0.164
Triangle	22	6	16	19.0	6.8	11.7	0.079
Polygon	3	1	2	7.9	2.9	4.9	0.033
Star	5	0	5	3.6	1.3	2.2	0.015
Irregular	67	24	43	65.6	23.7	40.5	0.272

The chi-square test is performed to compare the observed and expected frequencies for each shape. The results of the test are presented in Table 6.13.

Table 6.13: *Chi-square test statistics for shape*

	Total	M	F
χ^2	38.445	51.131	15.292
Df	7	7	7
P	0.000	0.000	0.032

The chi-square test shows that there is a significant difference for the observed number of selected images from each shape from an expected distribution when all the participants are observed, $\chi^2(1, N=241) = 38.445$, $p=0.000 < 0.05$, and when the participants are split on gender, $\chi^2(1, N=87) = 51.131$, $p=0.000 < 0.05$ for male, and $\chi^2(1, N=149) = 15.292$, $p=0.032 < 0.05$ for female. To get more precise results for each shape separately a binomial test compared the observed frequencies to the expected frequencies under a binomial distribution with p_s as a probability parameter. An exact binomial sign test on all the participants indicated that there is a significant positive preference for the shapes square (p=0.000) and circle (p=0.001), a significant negative preference for the shapes rectangle (p=0.000) and polygon (p=0.041), and no significant preference for the other shapes.

To analyze the selection of shape based on gender a Mann-Whitney test was performed. The test showed that the selection of images based on shape between genders differed significantly only for circle, $U = 5526$, $z=-3.068$, $p=0.003 < 0.01$, $r=-0.2$. The detailed results are presented in Table 6.14.

Table 6.14: *Results from the Mann-Whitney test on shape*

	square	rectangle	circle	oval	triangle	polygon	star	irregular
Mann-Whitney	6078.5	6250.5	5526.0	5915.5	6232.5	6469.0	6264.0	6399.0
z	-1.469	-0.665	-3.068	-1.886	-0.977	-0.127	-1.723	-0.209
p	0.195	0.610	0.003	0.076	0.364	1.000	0.161	0.882

By interpreting the results it can be concluded that male participants showed a greater affinity towards images with a circle shape than female participants. In addition, both genders showed a higher preference for images with a square shape, and a lower preference for images with a rectangle or polygon shape. These findings show a limited support for hypothesis 13, as the variance in object preference based on shape is not as diverse as with color and especially category.

6.2.6.5 Password Length & Complexity

Almost all of the participants used the minimal graphical password length allowed in their group. Only 6% of all the participants selected a graphical password that was one image longer than the minimal length, and a negligible number selected a graphical password for two or more additional images. A summary of the results is presented in Table 6.15.

Table 6.15: *Descriptive statistics for password length groups*

	M	SD
4-length group	4.12	0.31
5-length group	5.05	0.22
6-length group	6.02	0.18

For the unsupervised sessions that followed enrollment the password attempt time and login failure rate were compiled as variable from the system logs and were subjected to further analysis. The average time to enter the password for all participants is 9.22s (SD = 22.142), which is a little lower than the time reported in Experiment 3. Dividing password attempts into successful password attempts and unsuccessful password attempts yields average times of 8.56 (SD=21.226) and 12.06 (SD=34.632) respectively. The descriptive statistics for the clustering groups are presented in Table 6.16.

Table 6.16: *Descriptive statistics for login analysis based on clustering groups*

Login		**Group 4L**	**Group 5L**	**Group 6L**
Unsuccessful	Mean	9.101	11.933	18.326
	SD	9.0088	14.1528	18.7658
Successful	Mean	9.6	10.3725	13.2225
	SD	11.358	17.7155	25.004

A one-way ANOVA analysis revealed the following results for successful and unsuccessful logins. There was a significant difference for the completion time in successful logins between the participants from group 4L and the participants from group 6L, $F=6.423$, $p=0.042<0.05$. There was no significant difference for the completion time in successful logins between the participants from group 4L and the participants from group 5L, and there was no significant difference for the completion time in successful logins between group 5L and 6L. There was a significant difference for completion time in unsuccessful logins between participants from group 4L and group 6L, $F=8.362$, $p=0.027<0.05$. There was no significant difference for the completion time in unsuccessful logins between participants from group 4L and group 5L, and there was no significant difference for the completion time in unsuccessful logins between the participants from group 5L and the participants from group 6L.

Regarding login failure rate, a one-way ANOVA revealed a significant difference between groups 4L and 6L, $F=14.634$, $p=0.04<0.05$, and no significant difference between groups 4L and 5L or between groups 5L and 6L. There was no significant difference for any variable between genders in any password length groups.

These results are not in support of hypothesis 15. It can be summarized that a graphical password with a length of 4 or 5 images does not affect authentication performance. However, the effects of a graphical password with a length of 6 images can be observed when compared to the results of a 4-length graphical password. In the current design of the ImagePass authentication mechanism the restriction of a minimal 4 image-length for a graphical password should neither be increased nor decreased. As the results in this experiment show, the users will most likely keep to the minimal number of allowed images with the more conscientious users adding one or possible two images to their graphical password to increase the security of their authentication key.

A final aspect that needs to be considered before concluding this experiment is the complexity of the selected graphical password. Besides graphical password length, differences in graphical password complexity may influence crackability rates. Therefore, it was determined that it would be informative to further examine the complexity of the graphical password. There are no specific measures that consider the factors that determine the complexity of a graphical password, therefore, as a result a complexity index g_x was created for exploratory purposes. The complexity index is a measure that multiplies the graphical password length (g_l) with the number of different images used in the graphical password (n_d). Scores based on this complexity index were computed for

each graphical password. A one-way ANOVA discovered a significant difference between genders for complexity rates, $F=7.612$, $p=0.035<0.05$ with female participants selecting more complex passwords than males.

Regardless of password length all the possible combinations of an image repeating could be observed in the data analysis. Users repetitively selected a single image as a password, alternated the sequence with two images, made a repetition of only one image and used different images for the rest of the authentication key, etc. Clustering the passwords in all of the possible groups yielded small test samples that couldn't be subjected to further statistical analysis.

6.2.6.6 Graphical Password Selection Hotspots

Data gathered during this experiment can be used to evaluate some of the findings in the eye tracking study. During enrollment, when the participants selected their graphical password, the position of the selected image in the enrollment imageset was recorded. When data for all the selected images is aggregated it is possible to get a graphical password selection matrix based on the probability that an image in a particular spot is selected as a part of the authentication key. The summarized results are presented in Table 6.17.

Table 6.17: *Probability matrix for selecting an image in a graphical password*

	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Total
Row 1	0.069	0.048	0.070	0.065	0.042	0.054	0.348
Row 2	0.047	0.038	0.060	0.046	0.061	0.040	0.292
Row 3	0.073	0.055	0.032	0.023	0.018	0.005	0.207
Row 4	0.013	0.012	0.008	0.009	0.011	0.022	0.074
Row 5	0.016	0.011	0.015	0.011	0.014	0.012	0.079
Total	0.218	0.164	0.184	0.153	0.147	0.134	

A visual representation of these results is presented in the Figure 6.11 chart. As initially defined in the eye tracking experiment, users have a selection preference for images in the first three rows. The interest in the first two rows has a slight variance between columns, while the interest in the third row drops sharply as the column number increases. The last two rows share a similar, low, frequency with almost unnoticeable variations between columns.

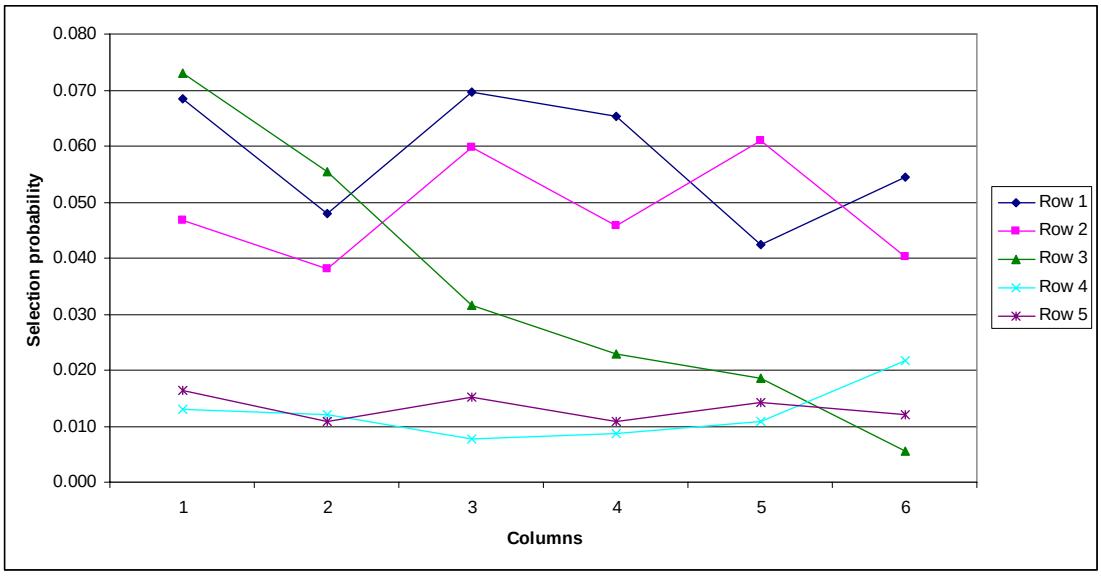


Figure 6.11: *Probability distribution chart for selecting an image in a graphical password*

6.3 Summary

This chapter completes the evaluation process of ImagePass and with that the research on recognition-based graphical authentication as outlined in the introductory section.

The focus was on evaluating the cognitive user process through a large-scale eye tracking study and on defining the graphical password properties for a more memorable authentication key through a user study. The eye tracking analysis helped define cognitive user profiles for enrollment and authentication and further defined the user behavior when interacting with graphical authentication. The user study examined the properties of the graphical password in more details and defined user preferences for single-object images based on category, color and shape.

The findings in this experiment influence the selection of images available in the ImagePass database as well as the algorithm that at this moment randomly selects decoy images to complete the authentication imageset. To lower the probability of a guessing attack based on a matrix that is aligned with the probabilities of a user selecting images from a specific category and/or with a specific color/shape, the decoy-selector algorithm needs to be modified as follows:

- Decoy images that belong to the preferred Food or Business category should be added. A random assignment of images that follow the same categories as selected images could also be considered.
- Decoy images that belong to the preferred beige color should be added. A random addition of images with the same colors as selected images could be considered.
- Decoy images that belong to the preferred square shape should be added. A random assignment of images that follow the same shapes as selected images could also be considered.

The findings of the research presented in this chapter, along with the results from previous experiments lead to finalizing the design guidelines which are presented in the next, concluding section, and complete the requirements for developing a usable and secure recognition-based graphical authentication mechanism that is suitable for ubiquitous environments. The developed mobile prototype is presented at the end of Chapter 4.

7 Conclusion

As of 2011, there are an estimated 2 billion people that use the Internet. Many of these users readily provide personal information in order to perform online tasks, suggesting that they trust that their information is secure on the sites that they visit (Vu, et al., 2007a). The most basic security feature of many online accounts is through user authentication. To gain access to a specific account, a user must supply a username that identifies the account to be accessed and an authentication key which assures the system that the user is who that person claims to be (Schultz, 2005). Authentication keys in the form of text-based passwords are often the only means of authentication for web-based accounts hence they must be kept secure by users if this method is to be successful. Nevertheless, users readily admit to sharing passwords or giving out their passwords to other people (SafeNet, 2005).

There have been many authentication incidents and analysis reported in recent years. In one such incident report, hackers successfully harvested over 40,000 MySpace passwords and published them on several websites (Grimes, 2006). Considering the fact that there is an obvious overlap in the customer base for prominent online businesses, such as banks, e-mail services, top commerce sites, and social networking sites (Ives et al., 2004), there is a high probability that many of the passwords were used across more than one account. The reuse of passwords across multiple accounts leads to "a domino effect," where the access to an important bank account is only as secure as the least secure site for which that same password is used. This illustrates the need for each password to be unique or distinctly different from every other password.

In the analysis of the password habits of approximately half million Microsoft users, Florêncio and Herley (2007) discovered that users managed an average of 25 password-protected accounts, but only used six to seven different passwords each of which was commonly shared across approximately four accounts. In addition, they revealed that the average length of passwords was six characters, only a slight improvement over the four character length found in a 30-year old UNIX study (Morris & Thompson, 1979). Similarly, modern users frequently used passwords consisting of letters typed all in one case, and rarely combined their passwords with numbers and/or special characters. When considering how users managed their passwords for accurate recall, people showed difficulty in password recall and often forgot their passwords. Florêncio and Herley concluded that while computer use and technology has grown by leaps and bounds during the past three decades, user behavior concerning passwords has changed only slightly. People use memory, pieces of paper, and trial and error methods to remember the password for a specific account, and when that fails they result to requesting for their password to be reset.

Vu, et al. (2007a) showed that there are two factors influencing the success of user name and password combination, security and memorability. Brostoff and Sasse (2001) found that people tend to use familiar and easily accessible personal information when creating passwords, such as variations on one's own name, nickname, birthday, or address, because these are memorable. However, Vu et al. (2003) found that the use of such personal information results in the creation of passwords that are not only easy for the user to remember, but are also more easily cracked. One way to prevent users from

selecting easily guessable passwords is to assign strong ones. However, Klein (1990) noted that this approach actually may not produce the desired effect as assigned strong passwords don't have a meaningful personal association for the user. Thus, the user most likely would revert to writing the password down for later use.

Given that the human user is considered to be the weakest link in password security (Armstrong, 2003; Schneier, 2000), research has been devoted to assessing a myriad of password schemes aimed at improving security, and some of the more interesting password technologies include graphical passwords. Sasse (2001; 2003) claims that the perception of the user as the "weakest link" can be altered with appropriate interaction design. The unavoidable usability issue with graphical authentication is the not-so-conventional approach to authentication. In essence, all web users are most comfortable with text-based passwords. To quote Nielsen with his Law of Web user experience: Users spend most of their time on other sites. Thus, anything that is a convention and used on the majority of other sites will be burned into the users' brains and you can only deviate from it on pain of major usability problems. Although intended for standard web interfaces, this statement is explicitly true for usability in security as well. The common web practice for authentication is the text-based password. And even with this familiar concept, the common unsecure user behavior leans towards circumventing the security of the system for his/her personal convenience.

This dissertation presented a contribution to the important topic of graphical authentication by a thorough study and development of a recognition-based system intended for desktop and mobile environments. Through a detailed analysis of current state-of-the art in the field and six experiments, a functioning prototype was developed named ImagePass. Conceptually, ImagePass is a recognition-based graphical authentication mechanism that uses images of single objects as the authentication key.

7.1 Images vs. Words

Episodic memory tasks suggest that visual material is remembered better than verbal material in both recognition (Nickerson, 1965; Snodgrass & McClure, 1975) and recall (Bousfield et al, 1957). There are at least three hypotheses that have been proposed for the superiority of images over words: dual coding, superior sensory codes, and uniqueness of entry in semantic memory.

Dual coding. According to Paivio's dual coding model, images are better remembered than words in most tasks because images are more likely to be dually coded than words (i.e., registered in both the image and verbal stores); and the image code, which is more likely to be stored to an image than to a word, is the more effective code for item memory (Paivio, 1971; Paivio & Csapo, 1973). For example, when subjects are instructed to form visualizations of words and to name images, recall performance is equal for images and words (Paivio & Csapo, 1973), while recognition performance continues to be superior for images over words under the same dual-coding instructions (Snodgrass & McClur e, 1975). Further evidence that images are more likely to be dually encoded than words comes from: the finding that subjects have a difficulty deciding between a studied image and its name, in a forced-choice recognition test (Snodgrass et al., 1974), and from the lack of improvement in item recognition memory for images studied under verbal encoding instructions over those studied under imagery instructions (Snodgrass & McClure, 1975).

Superior sensory code. The superior sensory code hypothesis, proposed by Nelson and his colleagues, attributes the image superiority effect to the more elaborate sensory codes of images when compared to words (Nelson et al., 1977; Nelson et al., 1976). They found that increasing the visual similarity of a set of images in a paired-associate recall

task, while keeping conceptual similarity constant, had the effect of destroying, and even reversing, the superiority of images over words.

Uniqueness. The third hypothesis, that pictures are remembered better than words because of their uniqueness in semantic memory, has been proposed by Durso and Johnson (1979). The argument here is that words, even concrete names, are more polysemous than images, and hence their semantic representation is less likely to be contacted on a recognition test or retrieved during recall than the word's corresponding image. A related hypothesis has been proposed by Potter et al. (1977) to account for the finding that although the relatedness of a probe to a sentence could be decided as quickly for image as for word probes, suggesting an abstract format for sentence meaning, probe pictures were recalled better than probe words.

People remember information best when it is related to them, a phenomenon known as the self-reference effect (Rogers et al., 1977). This effect might be explained through the link between encoding and retrieval, meaning, when individuals generate cues for retrieval, the cues that can be personally associated could be more effective than other cues (Greenwald & Banaji, 1989). Familiarity and relevance of information to one's life is a central feature of password selection (Riddle et al., 1989), but personal alphanumeric passwords are generally weak because they can be easily guessed by others. Using simple images is in adherence with this phenomenon. By using information that is not biographical, but still personally relevant, the user can generate an authentication key that would be both memorable and more resistant to guessing.

7.2 Understanding ImagePass

It is obvious that when the items being processed have a high information load it is likely that fewer items can be processed at a time (Alvarez & Cavangh, 2004). Therefore, ImagePass uses images representing single everyday objects. The selection of this type of images was a direct consequence of the results from the preliminary analysis, specifically Experiment 2. As shown in this experiment, when dealing with image recognition, users have a greater difficulty recognizing images with abstract content or images with human faces. Results in this experiment clearly show that the easiest image content to remember and recall is a single-object. Supported by hypotheses 3 and 4, this experiment showed the superior memorability rates for single object images by demonstrating a higher image recognition success for single-object images than abstract or face images in both short-term and long-term memory tests. The recognition success scored high on measured variables, completion time and users' success perception. This result can be explained by the potential complexity of the proposed images. Single-object items can be described very easily, as shown in the preliminary study of Experiment 6, mostly in a word, sometime two. The words themselves are easily relatable to a mental model the user has about the object, hence the images are remembered more easily. On the other hand, faces have many distinctive features that need to be remembered, eyes, hair, mouth, nose, expression, skin color and are very difficult to describe. Some faces are easy to remember due to the complex mental model the user has built on faces over his/her life experience, but just as easily some faces are just as easily forgotten. Abstract images share in this complexity on a different level, as the user cannot relate to any color or shape present in the image, but can remember the images based on some undefined characteristics. As evident from the focus group sessions of Experiment 3, in real life scenarios users have a preference for objects they can relate to. This finding should be a highly influential factor when the image database for graphical password selection is created. Regardless of the object class, the preference for images included in password selection should lean toward personal associations of the user. For example, in pictures of animals a duck image would

be preferable to a platypus image. Finally, the memorability of the single-object images was further improved in the preparatory phase of Experiment 6. At this instance, images that scored high in the visual complexity rating or low in the familiarity rating were permanently discarded from the system, thus creating a image database that is more memorable, and in turn, more usable for graphical authentication.

Self-generated passwords are more memorable than system-assigned passwords (Vu et al., 2007b), also known as “the generation effect” in behavior research (Slamecka & Graf, 1978). The simple act of generating the password in the mind, consciously entering the password, and successfully confirming the password, forces the user to process this password at several points and reinforces associations for later retrieval. Therefore, ImagePass uses only self-generated graphical passwords with the only imposed restriction being graphical password length. However, like with text-based passwords, this authentication mechanism has the same difficulties of remembering multiple passwords. Furthermore, the graphical password selection process in ImagePass, doesn’t allow the user to easily select repeating graphical passwords. As usually more than one confirmation attempt is needed to improve memory for the password (Vu et al, 2006), when used in multiple environments, ImagePass can have two or three graphical password confirmations in order to improve the memorability of using multiple passwords in different environments.

In ImagePass, users select their graphical password from an archive of images. This ensures that the pictures are suitable for authentication as they are filtered to exclude pictures with poor fidelity, unsuitable dimensions, violations of privacy, obscenity, high complexity, etc. On the other hand, results from Experiment 6 showed that users have some preferences when selecting images representing single objects that in some cases are gender specific which was supported by hypothesis 11. In particular, male users have a strong preference for images that can be related to games while female users have a strong preference for images that can be related to kids & toys, or plants. Interestingly, male users have a low preference for images related to sports or tools, while female users have a low preference for images from the home category. There was a distinct preference in all users for images that can be related to business or food, the later also noted in the user study from Experiment 3 where users made most comments on food images. These preferences are not strong enough in all cases to betray the identity of the user. When compared to text-based password they are analogous to preferences for specific letters, which in some cases depend more on the frequency of use in the natural language than on actual user partiality. Nevertheless, the findings are still viable for improving the guessability of the graphical password for attackers or for improving the usability and security of the authentication mechanism by the system designer. In the design process of a recognition-based graphical authentication mechanism the availability of images in the graphical password selection grid per category, color and/or shape can be distributed in accordance with the frequency findings reported in this dissertation.

The graphical password selection space seems to be adequately sized at 30 images in the web and 16x3 in the mobile version. In all of the ImagePass experiments very few of the participants made an attempt to load a new set of images in the web version. The option itself is an added benefit for the few users that are more meticulous in creating a graphical password. Screen scrolling in the mobile version was not recorded so no specific claims can be made for ubiquitous devices. The size of the challenge imageset during authentication has a direct impact on the amount of time the user will take to locate the target images for the authentication key. For these reasons, once the graphical password is selected, the users' authentication imageset is lowered to 16 rather than the full 30 images during selection. In addition, the image distractors that complete the users' authentication imageset are selected automatically by the system and are fixed for every

user which lowers the guessability of the graphical password. Most traditional text-based authentication mechanisms have a minimal six-character requirement for passwords. The results from experiment 6 showed that using a minimum of six images is rather daunting for the user as it increases the necessary cognitive load and might affect the memorability of the graphical password. The final ImagePass version has a minimal requirement of four images, with no restrictions on whether an image in the imageset can be repeated or not.

The initial deficiencies and advantages of the interface elements were addressed both in the preliminary eye tracking study and in the focus group sessions and were corrected in the system redesign. The first recommendation for the system was not to separate the username and password screens, neither during enrollment, nor during authentication. Although users didn't show strong objections to this issue it was still opposite to their authentication habits. Another interface element that required attention was the selected password box, which revealingly displayed the clicked items. This element predisposes the system to shoulder surfing attacks and proved to be unnecessary beyond the enrollment phase. Instead of removing this box it was sufficient to replace it with blank image entries for each click, and letting the user decide whether the images will be visible or not. The appearance of a blank image would signal the user that an image has been clicked, while the number of clicked images would be indicated by the number of blank display images. This approach is similar to the asterisks of the traditional password field that hides the typed text, but still displays information on type events.

The main eye tracking analysis offered insight into the initial perception and cognitive user behavior during interaction with an object-based graphical authentication mechanism. The study noted potential differences in how male and female users observe the system which was initially related to gender-specific information foraging patterns. Upon more detailed analysis these differences were narrowed and limited to the graphical password selection process and explained by defining separate user profiles which have gender inclinations. In addition, the users' exploration of the system didn’t follow postulation set out by scanpath theory as there were no repetitive patterns dissimilar between persons for a specific stimulus and between stimuli for a specific person

The focus in the evaluation of user performance was to test the system under different working conditions and constraints such as, frequency of use and mnemonic training. Expectantly, the results showed that frequent users adapt to the system, make fewer mistakes, are more efficient and have an improved memorability of the password. On the other hand, the introduction of mnemonic instructions seemingly produced conflicting results. According to the logs and statistical analysis the memorability of the graphical password was similar for all participants. Nevertheless, participants who received mnemonic instructions were under the impression that they memorized the password better, more so, they gave ImagePass a higher usability rating than participants that received no instructions. This can be classified as some kind of HCI “placebo” effect where the users self-expected to have an improved performance because they received specific training.

Compared to other recognition-based graphical authentication mechanism, ImagePass has shown improved usability and security features. For example, unlike Passfaces, a graphical authentication mechanism that uses a sequence of faces as the authentication key, ImagePass does not guide the user through several panel screens in order to enter the graphical password, instead, all the images are presented in a single grid. On the one hand, using multiple panels does guide the password sequence, but on the other it decreases the usability of the system as the user has to wait for the panels to switch. In addition, the security of the system is decreased because the number of panel-switches instantaneously reveals the length of the authentication key, thus increasing the vulnerability of the system. Finally, facial images are not appropriate for small screen

sizes and mobile environments as it would be more difficult to distinguish faces if the image has small dimensions.

7.3 Guidelines

This concluding section will consolidate the findings presented in the previous discussion in the form of a list of guidelines for developing ubiquitous graphical authentication mechanisms. These guidelines will be extended as research continues, since the research field of usable security in mobile graphical authentication is still in its infancy. For the present, this list distils current research and research presented in this dissertation into best practices for mobile graphical authentication design.

7.3.1 Usability Guidelines

Based on research presented in this dissertation the following usability guidelines can be suggested for graphical authentication mechanisms in ubiquitous environments:

Target pictures. The graphical authentication key should consist of at least four images. The authentication image set should be available on one screen without having the user swipe or scroll. Separate pages should be avoided as this increases authentication time and reveals the graphical password length which diminishes both usability and security of the mechanism.

Image content. Image content needs to be simple as mobile screens are relatively small when compared to desktop screens. Larger the challenge set, smaller the images, and simpler the image content needs to be.

Distractors. The distracting images in the challenge set should always be fixed and assigned by the system. Varying distractors will easily reveal the content of the graphical password. The images themselves should be semantically tagged so that the system can choose distractors that are less likely to confuse the user.

Image selection. Users should be allowed to choose the images for their authentication key. However, they should not be allowed to select personal images that they would upload to the system. This is an increased cost for the developer as an appropriate image archive needs to be assembled. The selected images should be easily named, have high picture fidelity, be sufficiently small to avoid storage and bandwidth issues.

Input modality. As almost all new ubiquitous devices are minimal on buttons and trending with touch-screen interfaces, the entry of the graphical password should not require the use of a keyboard. Instead, it should be tapped (or clicked)

7.3.2 Security Guidelines

Based on research presented in this dissertation the following usability guidelines can be suggested for graphical authentication mechanisms in ubiquitous environments:

Breakability. When the mechanism is used in an online environment the pages should expire immediately. In any environment, image names should be randomized and a silent second authentication layer needs to be added as described in Chapter 4.

Challenge set size. The optimal authentication image set size is limited by the screen size and interface design of current mobile devices. The upper limit for the optimal size should be 16 images, as this is the number that can comfortably fit on a mobile screen.

Repositioning images. Touch-screen devices are susceptible to "smudge attacks", where attackers can determine the user's password by the finger smudges left on the smart phone's surface. Therefore it’s allowable to change the positions of the authenticator and

distractor images within the confines of the authentication imageset.

Replacement. Users should be provided with an easy and secure way to replace their authentication key images. At this early stage of deployment, key replacement should be combined with email confirmations as electronic mail systems use alphanumeric passwords and are not threatened when the graphical password is compromised.

Error messages. Error messages should be avoided as much as possible. When necessary they should provide information that is never helpful to the potential intruder.

7.4 Future Work

There is an increasing understanding of the user's role in the security of any system. One way to make the user link stronger is to consider essential factors such as the user's needs, abilities, inclinations and skills in formulating security mechanisms and policies. Even with graphical authentication, improving usability should not focus on educating the users about security management or password selection. Instead, the system should minimize user interaction, meaning that the security mechanism should be unobtrusive, requiring little input or feedback from the user.

Graphical authentication mechanisms are a relatively new innovation and have not yet been implemented in ubiquitous environments. One of the major questions remaining is the extent to which graphical authentication can replace the traditional password. While currently user habits determine the authentication behavior, as we move from the desktop to the ubiquitous environment these established routines are changing. With the expected prevalence of mobile devices, advantages of the text-based password will quickly disappear. The recommendations for text passwords usually note that the authentication key should contain both lowercase and uppercase letters combined with numbers and/or special characters. Touch-screen phones, e-book readers and tablets are already ergonomically diverse requiring interactions different from the traditional typing and clicking making complex passwords more difficult to enter. Hence, graphical passwords are useful for keyboard-less systems. As pervasive computing continues to mature the advantages of graphical passwords might prove as an adequate replacement of traditional passwords.

There are several ImagePass aspects that still need to be researched, especially those that occur when using multiple graphical passwords for accessing different systems. Under the multiple password conditions these aspects include: a systematic examination of frequency of access, the interference resulting from interleaving access, the patterns of access while remembering multiple graphical passwords. All of these factors could significantly impact the ease of authenticating using multiple ImagePass schemes.

Finally, regarding the terminology in graphical authentication, specifically referring to the term "graphical password" as used in this dissertation, it is the author's suggestion that it might be linguistically more correct to use the term "passimage" instead. A password grants access to a system, by letting the user 'pass', with the provision of a string of alphanumeric characters, 'word'. Conversely, the coined term "passimage" alludes to letting the user 'pass' the authentication process by providing a sequence of "images".

8 References

Adams, C., Wiener, M.J., 2002. Multi-factor biometric authenticating device and method. US Patent Office, Classification 713/186.

Alsulaiman F., El Saddik, A., 2006. A novel 3D graphical password schema. In Proceedings of the IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems.

Alvarez, G. A., Cavanagh, P., 2004. The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. Psychological Science 15-2, pp. 106-111.

Anderson J., Bower, G, 1972. Recognition and retrieval processes in free recall. Psychological Review, vol. 79-2, pp. 97-123.

Armstrong, I., 2003. Passwords exposed: Users are the weakest link. SC magazine.

Baayen, R. H., Piepenbrock, R., Rijn, van H., 1993. The CELEX lexical data base on CD-ROM.

Ball, L. J., Eger, N., Stevens, R., Dodd, J., 2006. Applying the PEEP Method in Usability Testing. Interfaces, 67, pp. 15-19.

Balota, D. A., Cortese, M. J. Sergent-Marshall, S. D., Spieler, D. H., Yap, M.J., 2004. Visual Word Recognition of Single-Syllable Words. Journal of Experimental Psychology 133-2, pp. 283-316.

Baron, R.J., 1981. Mechanisms of human facial recognition. International Journal of Man-Machine Studies 15-2, pp. 137-178.

Bergadano, F., Crispo, B., Rufio, G., 1998. High dictionary compression for proactive password checking. In ACM Transactions on Information and System Security, 1-1 pp. 3-25.

Bertini, E., Gabrielli, S., Kimani, S., 2006. Appropriating and Assessing Heuristics for Mobile Computing. In Proceedings of the Workshop on Advised Visual Interfaces (AVI'06), pp. 119-126.

Bicakci, K., Atalay, N. B., Yuceel, M., Gurbaslar, H., Erdeniz, B., 2009. Towards usable solutions to graphical password hotspot problem. In Proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference.

Biddle, R., Chiasson, S., van Oorschot, P.C. 2011. Graphical Passwords: Learning from the First Twelve Years. ACM Computing Surveys 44-4.

Blonder, G.E., 1995. Graphical Password. U.S. Patent 5559961. Lucent Technologies, Inc. New Jersey.

Boroditsky, M. Passlogix password schemes. http://www.passlogix.com, (accessed December, 2010).

Bousfield, W. A., Esterson, J., Whitmarsh, G. A., 1957. The effects of concomitant colored and uncolored pictorial representations on the learning of stimulus words. Journal of Applied Psychology 41, pp. 165-168.

Bower, G.H., Karlin, M.B., Dueck, A., 1975. Comprehension and Memory of Pictures.

Memory and Cognition 2 pp.216-220.

Brandt, S. A., Stark, L. W., 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. Journal of Cognitive Neuroscience 9, pp. 27-38.

Brostoff, S., Sasse, M.A., 2001. "Safe and Sound: A Safety-Critical Approach To Security". Proceedings of the 2001 Workshop on New Security Paradigms, pp. 41-50, New Mexico, 2001.

Brysbaert, M., New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods 41-4, pp. 977-990.

Bucci, W., 1985. Dual Coding: A Cognitive Model for Psychoanalytic Research, Journal of the American Psychoanalitics Association 33, pp. 571-607.

Burnett, M., 2005. Perfect Passwords - Selection, Protection, Authentication". Syngress Publishing.

Buzan, T, 2003. User Your Memory. BBC Active.

Carroll, J. B., White, M. N, 1973. Word frequency and age of acquisition as determiners of picturenaming latency. Quarterly Journal of Experimental Psychology 25, pp. 85-95.

Cave, B. C., 1997. Very long-lasting priming in picture naming. Psychology Science vol. 8, pp. 322-325.

Chiasson, S., Forget, A., Biddle, R., van Oorschot, P.C., 2008. Influencing users towards better passwords: Persuasive Cued Click-Points. In Human Computer Interaction (HCI), British Computer Society, Liverpool, England.

Chiasson, S., Forget, A., Biddle, R., van Oorschot, P.C., 2009. User interface design affects security: Patterns in click-based graphical passwords.

Chiasson, S., van Oorschot, P. C., Biddle, R., 2007. Graphical password authentication using Cued Click Points. In Proceedings of the European Symposium On Research In Computer Security pp. 359-374.

Cowan, C., Pu, C., Maier, D., Walpole, J., Bakke, P., Beattie, S., Grier, A., Wagle, P., Zhang, Q., Hinton, H., 1998. StackGuard: Automatic adaptive detection and prevention of buffer-overflow attacks. In Proceedings of the 7th USENIX Security Conference, pp. 63–78.

Cowen, L., Ball, L. J., Delin, J., 2002. An Eye Movement Analysis of Webpage Usability. In: People and Computers XVI - Memorable yet Invisible: Proceedings of HCI 2002. Springer-Verlag.

Craik, F.I.M., Tulving, E., 1975. Depth of processing and word retention. Journal of Experimental Psychology vol. 104, pp. 268–294.

Dahlback, N., Jonsson, A., Ahrenberg, L., 1993. Wizard of Oz Studies - Why and How. In Proceedings of the International Conference on Intelligent User Interfaces, pp. 193–200.

Davis, D., Monrose, F., Reiter, M.K., 2004. On User Choice In Graphical Password Schemes. Proceedings of the 13th USENIX Security Symposium pp. 151-164.

De Angeli, A., Coventry, L., Johnson, G., Renaud, K., 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. International Journal of Human-Computer Studies Volume 63, Issues 1-2, pp. 128-152.

De Valois, R.L., De Valois, K.K., 1988. Spatial Vision. Oxford University Press. New

York.

Dhamija, R., Perrig, A., 2000. Deja Vu: a user study using images for authentication, Proceedings of the 9th Usenix Security Symposium.

Dirik, A., Menon, N., Birget, J., 2007. Modeling user choice in the Passpoints graphical password scheme. In Proceedings of the 3rd ACM Symposium on Usable Privacy and Security (SOUPS). Pittsburgh, USA.

Dunphy P., Yan, J., 2007. Do background images improve "Draw a Secret" graphical passwords? In Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS).

Durso, F. T., Johnson, M, 1979. Facilitation in naming and categorizing repeated pictures and words. Journal of Experimental Psychology: Human Learning and Memory 5, pp. 449-459.

Ebbinghaus, H., 1964. Memory: A contribution to experimental psychology. Dover Publications.

Eger, N., Ball, L. J., Stevens, R., Dodd, J., 2007. Cueing Retrospective Verbal Reports in Usability Testing Through Eye-Movement Replay. Proceedings of HCI.

Ehmke, C., Wilson, S., 2007. Identifying Web Usability Problems from Eye-Tracking Data. In Proceedings of HCI 2007.

Ehn, P., Kyng, M., 1991. Cardboard Computers: Mocking it up or Hands-on the Future. Design at Work, pp. 169–196. Laurence Erlbaum Associates.

Ericsson, K.A., Kintsch, W., 1995. Long-term working memory. Psychological Review 102 pp. 211–245.

Etezadi-Amolia, J., Farhoomand, A.F., 1996. A structural model of end user computing satisfaction and user performance. Journal of Information & Management Vol. 30-2, pp. 65-73.

Everitt, K., Bragin, T., Fogarty, J., Kohno, T. 2009. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2009).

Eysneck, M., Eysneck C., 1979. Processing Depth, Elaboration of Encoding, Memory Stores, and Expended Processing Capacity. Journal of Experimental Psychology 5, pp. 472-484.

Flechais, I., Sasse, A., Hailes, S., 2003. Bringing Security Home: A Process for Developing Secure and Usable Systems. Proceedings of the New Security Paradigms Workshop.

Florêncio, D., Herley, C., 2007. A large scale study of web password habits. WWW 2007.

Foulsham, T., Underwood, G., 2008. What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. Journal of Vision 8, pp. 1-17.

Friedman, B., Nissenbaum, H., Hurley, D., Howe, D.C., Felten, E., 2002. User's conceptions of risks and harms on the web: A comparative study. Proceedings of CHI 2002, Minneapolis, Minnesota.

Gao, H., Guo, X., Chen, X., Wang, L., Liu, X., 2008. Yagp: Yet another graphical password strategy. In Proceedings of the Annual Computer Security Applications Conference, 2008.

Gbadamosi, J., Zangemeister, W. H., 2001. Visual imagery in hemianopic patients. Journal of Cognitive Neuroscience 13, pp. 855-866.

Going, M., Read, J. D., 1974. Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. Perceptual and Motor Skills 39-1, pp. 109-110.

Goldberg, J.H., Stimson, M. J., Lewenstein, M., Scott, N., Wichansky, A.M., 2002. Eye Tracking in Web Search Tasks: Design Implications. Proceedings of the Eye Tracking Research & Applications Symposium, pp. 51-58.

Gong, L., Lomas, M., Needham, R., Saltzer, J., 1993. Protecting poorly chosen secrets from guessing attacks. IEEE Journal on Selected Areas in Communications 11-5, pp. 648-656.

Greenwald, A., Banaji, M. R., 1989. The self as a memory system: Powerful, but ordinary. Journal of Personality and Social Psychology 57, pp. 41–54.

Gregg, V.H., 1986. Introduction to Human Memory. Routledge & Kegan Paul.

GrIDsure Corporate. GrIDSure http://www.gridsure.com, Last accessed December 2010.

Grier, R. A., Kortum, P., Miller, J. T., 2007. How users view web pages: An exploration of cognititve and perceptual mechanisms. In Human Computer Interaction Research in Web Design and Evaluation, pp. 22-41.

Grimes, R. A., 2006. MySpace password exploit: Crunching the numbers (and letters). http://www.InfoWorld.com/article/06/11/17/47OPsecadvise_1.html (accessed December 2009).

Guan, Z., Lee, S., Cuddihy, E., Ramey, J., 2006. The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye-Tracking. CHI, Montreal, Canada.

Gutmann, P., Grigg, I., 2005. Security Usability. IEEE Security And Privacy. IEEE Computer Society.

Hayashi, E., Christin, N., Dhamija, D., Perrig, A. Use Your Illusion: Secure authentication usable anywhere. In the 4th ACM Symposium on Usable Privacy and Security (SOUPS), Pittsburgh, July 2008.

Heijden, A.H.C., 1992. Selective Attention in Vision. Routledge. London.

Hendry, D.G., Mackenzie, S., Kurth, A., Spielberg, F., Larkin, J., 2005. Evaluating Paper Prototypes on the Street. In Proceedings of Conference on Human Factors in Computing Systems - Extended Abstracts (CHI '05), pp. 1447-1447.

Ho, P., Armington, J., 2003. A Dual Factor Authentication System Featuring Speaker Verification And Token Technology. Proceedings of the 4th International Conference on Audio and Video-Based Biometric.

Humphrey N., 1992. A History of the Mind, Simon and Schuster.

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K-J., Lehtinen, M., 2008. Gaze path stimulation in retrospective think aloud. Journal of Eye Movement Research, 2-4, pp. 1-18.

Isola, P., Xiao, J., Torralba, A., Oliva, A., 2011. What makes an image memorable? In Proceedings of CVPR 2011, pp. 145-152.

Ives, B., Walsh, K. R., Schneider, H., 2004. The domino effect of password reuse. Communications of the ACM, 47-4, pp. 75–78.

Jansen, W., Gavrila, S., Korolev, V., Ayers, R., Swanstrom, R., 2003. Picture Password: A Visual Login Technique for Mobile Devices. National Institute of Standards and Technology Interagency Report NISTIR 7030.

Jermyn, I., Mayer, A., Monrose, F., Reiter, M., Rubin, A., 1999. The Design and Analysis of Graphical Passwords. Proceedings of the 8th USENIX Security Symposium.

Jin, A.T.B., Ling, D.N.C., Goh, A., 2004. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. Pattern Recognition 37-11, pp. 2245-2255, Elsevier.

Jøsang, A., Patton, M., 2001. Authentication for Humans: Proceedings of the 9th International Conference on Telecommunication Systems. Cox School of Business, Southern Methodist University, Dallas.

Jøsang, A., Patton, M.A., 2003. User interface requirements for authentication of communication. In Proceedings of the 4th Australasian user interface conference on User interfaces, pp. 75-80, Adelaide, Australia.

Josephson, S., Holmes, M. E., 2002. Attention to repeated images on the World-Wide Web: Another look at scanpath theory. Behavior Research Methods, Instruments & Computers 34, pp. 539-548.

Josephson, S., Holmes, M.E., 2002. Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. Proceedings of the Eye Tracking Research & Applications Symposium, pp. 43-51.

Ka-Ping, Y., 2003. Secure interaction design and the principle of least authority. In Proceedings of the CHI 2003 Workshop on Human-Computer Interaction and Security Systems.

Ka-Ping, Y., 2004. Aligning Security And Usability. IEEE Security And Privacy. IEEE Computer Society, Los Alamitos.

Karat, C. M., 1989. Iterative usability testing of a security application. In Proceedings of the Human Factors Society 33rd Annual Meeting, pp. 273–277.

Kim, B., Dong, Y., Kim, S., Lee, K-P., 2007. Development of Integrated Analysis System and Tool of Perception, Recognition, and Behavior for Web Usability Test: With Emphasis on Eye-Tracking, Mouse-Tracking, and Retrospective Think Aloud, Usability and Internationalization. HCI and Culture, pp. 113-121

Kimura D., 1987. Are Men's and Women's Brain Really Different? Canadian Psychology 28, pp. 133-147.

Kjeldskov, J., Stage, J., 2004. New Techniques for Usability Evaluation of Mobile Systems, International Journal of Human Computer Studies, vol. 60-6 pp. 599-620.

Klein, D. V., 1990. Foiling the cracker: A survey of, and improvements to, password security. In Proceedings of the 2nd USENIX Workshop on Security, pp. 5–14.

Komanduri, S., Hutchings, D., 2008. Order and entropy in Picture Passwords. In Proceedings of the Graphics Interface Conference.

Koops, B.J., 1999. The Crypto Controversy: A Key Conflict in the Information Society. Kluwer Law International.

Kucera, H., Francis, W. N, 1967. Computational analysis of present-day American English. Brown University Press.

Lachman, R., 1973. Uncertainty effects on time to access the internal lexicon. Journal of Experimental Psychology 99, pp. 199-208.

Laeng, B., Teodorescu, D., 2002. Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cognitive Science 26, pp. 207-231.

LeBlanc, D., Chiasson, S., Forget, A., Biddle. R. 2008. Can eye gaze predict graphical passwords? 4th ACM Symposium on Usable Privacy and Security (SOUPS), Pittsburgh, USA.

LeBlanc, D., Forget, A., Biddle. R, 2010. Guessing Click-Based Graphical Passwords by

Eye Tracking. IEEE Privacy, Security, Trust (PST), Ottawa, Canada.

Lederer, s., Hong, J. I., Dey, A. K., Landay, J. A., 2004. Personal privacy through understanding and action: Five pitfalls for designers. In Personal and Ubiquitious Computing. Springer-Verlag.

Lindgaard, G., 1994. Usability Testing & System Evaluation: A Guide for Designing Useful Computer Systems. London. Chapman & Hall.

Liu L., Khooshabeh P., 2003. Paper or Interactive? A Study of Prototyping Techniques for Ubiquitous Computing Environments. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – Extended Abstracts (CHI '03), pp. 774-775.

Lumsden, J., MacLean, R., 2008. A Comparison of Pseudo-Paper and Paper Prototyping Methods for Mobile Evaluations. In Proceedings of the International Workshop on Mobile and Networking Technologies for social applications (MONET'2008), pp. 538-457.

Madigan, S., 1983. Picture memory. In Imagery, Memory, and Cognition: Essays in Honor of Allan Paivio, pp. 65-89. Lawrence Erlbaum Associates.

Man, S., Hong, D., Mathews, M., 2003. A shoulder-surfing resistant graphical password scheme. In Proceedings of the International Conference on Security and Management.

McCarthy, J. Sasse, A., Riegelsberger J., 2003. Could I have the Menu Please? An Eye Tracking Study of Design Conventions. People and Computers XVII - Designing for Society, Proceedings of HCI 2003, pp. 401-414, Springer-Verlag.

McDougall, P., 2008. Microsoft Turns To Inkblots For Password Generation. http://www.informationweek.com

Mihajlov M., Jerman-Blazic B., 2011. Memorability, Performance and Perception in Graphical Authentication. Interacting with Computers, Elsevier.

Mitnick, K., Simon, W., 2002. The Art of Deception: Controlling the Human Element of Security. John Wiley & Sons.

Monrose, F., Rubin, A.D., 2000. Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems 16, pp. 351-359. Elsevier.

Morris, R., Thompson, K., 1979. Password Security: A Case History. Communications of the ACM 22-11, pp. 594-597.

Narayanan, A., Shmatikov, V., 2005. Fast dictionary attacks on passwords using time-space trade-off. In Proceedings of the ACM Conference on Computer and Communications Security.

Nelson, D. L., Reed, V. S., McEvoy, C. L., 1977. Learning to order pictures and words: A model of sensory and semantic encoding. Journal of Experimental Psychology: Human Learning and Memory 3, pp. 485-497.

Nelson, D. L., Reed, V. S., Walling, J. R., 1976. Pictorial superiority effect. Journal of Experimental Psychology: Human Learning and Memory 2, pp. 523-528.

Nickerson, R. S., 1965. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology 19, pp. 155-160.

Nielsen, J., 1993. Iterative user-interface design. Computer, 26-11, pp. 32–41.

Nielsen, J., 2003. Paper Prototyping: Getting User Data Before You Code. http://www.useit.com/alertbox/20030414.html (accessed October 2010).

Nielsen, J., Molich, R., 1990. Heuristic evaluation of user interfaces. In Proceedings of the ACM Conference on Human Factors in Computing Systems, pp 249–256.

Nielsen, J., Tahir, M., 2002. Hompage Usability – 50 Web Sites Deconstructed, New Rider Publishing.

Norman, D. A., 1969. Memory and Attention. An introduction to human information processing. John Wiley & Sons.

Norman, D. A., 1983. Design rules based on analyses of human error. Communications of the ACM 26-4.

Norman, D. A., 1998. The Psychology Of Everyday Things. Basic Books, New York.

Noton, D., Stark, L., 1971. Scanpaths in eye movements during pattern perception. Science 171, pp. 308-311.

Noton, D., Stark, L., 1971. Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision Research 11, pp. 929-942.

Oechslin, P., 2003. Making a faster cryptanalytic time-memory trade-off. In Crypto'03.

Oldfield, R. C., Wingfield, A., 1965. Response latencies in naming objects. Quarterly Journal of Experimental Psychology 17, pp. 273-281.

O'Neill, E., Johnson, P., Johnson, H., 1999. Representations and user-developer interaction in cooperative analysis and design, Human-Computer Interaction 14, pp. 43-91.

Oorschot, P.C., Thorpe, J., 2008. On predictive models and user-drawn graphical passwords. In ACM Transactions on Information and System Security, 10-4 pp. 1-33.

Paivio, A, 1971. Imagery and verbal processes. Holt, Rinehart, & Winston.

Paivio, A., Csapo, K., 1973. Picture superiority in free recall: Imagery or dual coding? Cognitive Psychology 5, pp. 176-206.

Paivio, A., Rogers, T., Smythe, P.C., 1968. Why are pictures easier to recall than words? Psychonomic Science, vol. 11-4, pp. 137-138.

Parkin, A.J., 1993. Memory: Phenomena, Experiment and Theory. Blackwell.

Pellegrino, J. W., Rosinski, R. R., Chiesi, H. L., Siegel, A., 1975. Picture-word differences in decision latency: An analysis of single and dual memory models. Memory & Cognition 5, pp. 383-396.

Pering, T., Sundar, M., Light, J., Want, R., 2003. Photographic authentication through untrusted terminals. Pervasive Computing, pp. 30-36,

Pernice, K., Nielsen, J., 2009. Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking. Nielsen Norman Group.

Perruchet, P., Pacteau, C., 1990. Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General 119, 264-275.

Peyrichoux, I. & Robillard-Bastien, A. (2006). Maximize Usability Testing Benefits with Eye Tracking. SIGCHI Conference Paper.

Piazzalunga U., 2007. Fundamental Concepts of the Software Quality Engineer. ASQ Quality Press.

Pieters, R., Rosbergen, E., Wedel, M., 1999. Visual attention to repeated print advertising: A test of scanpath theory. Journal of Marketing Research 36, pp. 424-438.

Polson, P., Lewis, C., Rieman, J., Wharton, C., 1992. Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. International Journal of Man Machine Studies vol. 36-5, pp. 741–773.

Potter, M. C., Faulconer, B. A., 1975. Time to understand pictures and words. Nature 253,

pp. 437-438.

Potter, M. C., Valian, V. V., Faulconer, B. A., 1977. Representation of a sentence and its pragmatic implications: Verbal, imagistic, or abstract? Journal of Verbal Learning and Verbal Behavior 16, pp. 1-12.

Preece, J., Sharp, H., Rogers, Y., 2002. Interaction Design. John Wiley & Sons.

Privitera, C. M. (2006). The scanpath theory: its definition and later developments. In Proceedings of Human Vision and Electronic Imaging XI, pp. 87-91.

Real User Corporation. Two Factor Authentication, Graphical Passwords - Passfaces, http://www.passfaces.com (accessed November 2010).

Reber, A. S., 1967. Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior 6-6, pp. 855-863.

Renaud K., 2004. Quantifying the quality of web authentication mechanisms: a usability perspective. Journal of Web Engineering.

Rettig, M., 1994. Prototyping for tiny fingers. In Communications of the ACM vol. 37-4, pp. 21–27.

Riddle, B. L., Miron, M. S., Semo, J. A., 1989. Passwords in use in a university timesharing environment. Computers and Security 8, pp. 569–578.

Rogers, T. B., Kuiper, N. A., Kirker, W. S., 1977. Self-reference and the encoding of personal information. Journal of Personality and Social Psychology 35, pp. 677–688.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., Boyes-Braem, P., 1976. Basic objects in natural categories. Cognitive Psychology 8, pp. 382-439.

Roth, V., Richter, K., Freidinger, R., 2004. A PIN-entry method resiliant against shoulder surffing. In Proceedings of the 11th ACM Conference on Computer and Communications Security.

Rudd, J., Stern, K., Isensee, S., 1996. Low vs. high-fidelity prototyping debate. ACM Magazine vol. 3-1, pp. 76–85.

Sá, M., Carriço, L., 2006. Low-Fi Prototyping for Mobile Devices. In Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '06), pp. 694-699.

SafeNet, 2005. 2004 Annual Password Survey Results. http://www.safenet-inc.com. (accessed January 2011).

Saltzer, J., Schroeder, M., 1975. The Protection Of Information In Computer Systems. In Proceedings of the IEEE 63-9, pp. 1278-1308. IEEE Press.

SANS Institute, 2004. The twenty most critical internet security vulnerabilities — the experts consensus. http://www.sans.org/top20 (accessed November 2007).

Sasse, M. A., 2003. Computer security: Anatomy of a usability disaster, and a plan for recovery. In CHI 2003 Workshop on Human-Computer Interaction and Security Systems.

Sasse, M.A., Brostoff, S., Weirich, D., 2001. Transforming the 'weakest link': A human-computer interaction approach to usable and effective security. BT Technology Journal 19.

Schneier, B., 2000. Secrets and Lies. Wiley.

Schultz, E. E., 2005. Web security and privacy. In Handbook of Human Factors in Web Design, pp. 613–628.

Schultz, E., Proctor, R.W., Lien, M.C., Salvendy, M.C., 2001. Usability and Security – An Appraisal of Usability Issues in Information Security Methods. Computers &

Security 20-7, pp. 620-634, Elsevier.

Sefelin, R., Tscheligi, M., Giller, V., 2003. Paper Prototyping - What is it good for?: A Comparison of Paper- and Computer-based Low-fidelity Prototyping. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – Extended Abstracts (CHI '03), pp. 778-779.

Shelfer, K.M., Procaccino, J.D., 2002. Smart Card Evolution. Communications of the ACM 45-7, pp. 83-88.

Shepard, R.M., 1967. Recognition Memory for Words, Sentences and Pictures. Journal of Verbal Learnings and Verbal Behavior 6 pp. 156-163.

Slamecka, N. J., Graf, P., 1978. The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory 4, pp. 592–604.

Smetters, D.K., Grinter, R. E., 2001. Moving from the design of usable security technologies to the design of useful secure applications. In Proceedings of the ACM Workshop on New security paradigms, pp. 82–89.

Smith, M. C., Magee, L. E., 1980. Tracing the time course of picture-word processing. Journal of Experimental Psychology 109-4, pp. 373-392.

Snodgrass, J. G., McClure, P, 1975. Storage and retrieval properties of dual codes for pictures and words in recognition memory. Journal of Experimental Psychology: Human Learning and Memory 1, pp. 521-529.

Snodgrass, J. G., Wasser, B., Finkelstein, M., Goldberg, L. B., 1974. On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. Journal of Verbal Learning and Verbal Behavior 13, pp. 27-37.

Snyder, C., 2003. Paper prototyping: the fast and easy way to design and refine user interfaces. Morgan Kaufmann Publishers, San Francisco.

Sobrado L., Birget, J.C., 2002. Graphical passwords. The Rutgers Scholar, An Electronic Bulletin for Undergraduate Research, vol. 4, 2002.

Stanford-Poynter Project. http://www.poynterextra.org/et/i.htm (accessed March 2010).

Stobert, E., Forget, A., Chiasson, S., van Oorschot, P. C., Biddle, R., 2010. Exploring usability effects of increasing security in click-based graphical passwords. In Proceedings of the 26th Annual Computer Security Applications Conference.

Stubblefield, A., Simon, D., 2004. Inkblot Authentication. Microsoft Technical Report 85.

Suo, X., Zhu, Y., Owen, G., 2005. Graphical passwords: A survey. In Proceedings of the Annual Computer Security Applications Conference.

Tafasa. Patternlock. http://www.tafasa.com/patternlock.html, last accessed January 2011.

Takada T., Koike, H., 2003. Awase-E: Image-based Authentication for Mobile Phones using User's Favorite Images. Human-Computer Interaction with Mobile Devices and Services, vol. 2795 / 2003, pp. 347-351. Springer-Verlag.

Tao H., Adams., C, 2008. Pass-Go: A proposal to improve the usability of graphical passwords. International Journal of Network Security, vol. 7-2, pp. 273-292.

Tari, F., Ozok, A., Holden, S., 2006. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In Proceedings of the 2nd ACM Symposium on Usable Privacy and Security (SOUPS), pp. 1-5 Pittsburgh, USA.

Thorpe J., van Oorschott, P.C., 2004. Graphical Dictionaries and the Memorable Space of Graphical Passwords. Proceedings of the 13th USENIX Security Symposium.

Thorpe J., van Oorschott, P.C., 2004. Towards Secure Design Choices for Implementing Graphical Passwords. In Proceedings of the 20th Annual Computer Security Applications Conference, Tuscon.

Thorpe J., van Oorschott, P.C., 2007. Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords. In Proceedings of the 16th USENIX Security Symposium.

Tobii, 2009. Retrospective Think Aloud and Eye Tracking Comparing the value of different cues when using the retrospective think aloud method in web usability. Tobii Technology.

Tognazzini, B., 1995. Tog on Software Design. Addison-Wesley Professional.

Treisman, A., Gelade, G., 1980. A Feature Integration Theory of Attention. Cognitive Psychology 12, pp. 97–136, Princeton University Press.

Tulving E., Pearlstone, Z., 1966. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior. vol. 5, pp. 381-391.

Tulving, E., Schacter, D. L., Stark, H. A., 1982. Priming effects in word-fragment completion are independent of recognition memory. Journal of Experimental Psychology: Learning, Memory & Cognition vol. 8-4, pp. 336-342.

Tulving, E., Watkins, M, 1973. Continuity between recall and recognition. American Journal of Psychology, vol. 86-4, pp. 739-748.

Vaidyanathan, J., Robbins, J. E., Redmiles, D. F., 1999. Using HTML to Create Early Prototypes. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – Extended Abstracts (CHI '99), pp. 232-233.

Van den Haak, M. J., de Jong, M. D. T., Schellens, J., 2003. Retrospective vs. Concurrent Think-Aloud Protocols: Testing the Usability of an Online Library Catalogue. Behavior & Information Technology, 22-5, pp. 339-351.

Van Gog, T., Paas, F., van Merrienboer, J. J. G. & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. Journal of Experimental Psychology: Applied, 11(4), 237-244.

Varenhorst, C., 2004. Passdoodles: A lightweight authentication method. MIT Research Science Institute.

Virzi, R.A., 1992. Refining the test phase of usability evaluation: How many subjects is enough?" Human Factors 34 pp. 457–468, Human Factors and Ergonomic Society.

Virzi, R.A., Sokolov, J.L., Karis, D., 1996. Usability problem identification using both low- and high-fidelity prototypes. Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '96) pp. 236-243.

Vu, K. P. L., Bhargav, A., Proctor, R. W., 2003. Imposing password restrictions for multiple accounts: Impact on generation and recall of passwords. In Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society pp. 1331–1335.

Vu, K. P. L., Cook, J., Bhargav, A., & Proctor, R. W., 2006. Short-term and longterm retention of passwords generated by first-letter and entire-word mnemonic methods. In Proceedings of the 5th Annual Security Conference.

Vu, K. P. L., Proctor, R. W., Bhargav-Spantzel, A., Tai, B.-L., Cook, J., Schultz, E. E., 2007. Improving password security and memorability to protect personal and organizational information. International Journal of Human–Computer Studies 65, pp. 744–757.

Vu, K.P.L., Garcia, F., Nelson, D., Sulatis, J., Creekmur, B., Chambers, V., 2007. Examining user privacy policies while shopping online: What are users looking for? In Proceedings of the 12th international conference on human–computer interaction, pp. 792–801.

Walker, M., Takayama, L., Landay, J.A., 2002. High-Fidelity or Low-Fidelity, Paper or Computer? Choosing Attributes When Testing Web Prototypes. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting Proceedings, pp. 661-665.

Walker-Smith, G. J., Gale, A. G., Findlay, J. M., 1977. Eye movement strategies involved in face perception. Perception 6, pp. 313-326.

Wedel, M., Pieters, R., 2006. Eye tracking for visual marketing. In Foundations and trends in marketing Vol. 1, Issue 4. Now Publishers.

Weinshall, D., Kirkpatrick, S., 2004. Passwords you'll never forget, but can't recall. In Proceedings of CHI 2004 extended abstracts on Human factors in computing systems.

Weiss, R., De Luca, A., 2008. PassShapes - utilizing stroke based authentication to increase password memorability. In Proceedings of NordiCHI, pp. 383-392.

Whitten, A., Tygar, J. D., 1999. Why Johnny Can't Encrypt. A Usability Evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium.

Whitten, A., Tygar, J. D., 2003. Safe staging for computer security. In Proceedings of the CHI 2003 Workshop on Human-Computer Interaction and Security Systems.

Wiedenbeck, S., Waters, J., Birget, J. C., Brodski, A., Memon, N., 2005. Authentication using graphical passwords: Effects of tolerance and image choice. In Proceedings of the 1st ACM Symposium on Usable Privacy and Security (SOUPS). Carnegie-Mellon University.

Wiedenbeck, S., Waters, J., Birget, J. C., Brodski, A., Memon, N., 2005. PassPoints: Design and longitudinal evaluation of a graphical password system, International Journal of Human-Computer Studies 63, pp. 102-127.

Wu, M., Garfinkel, S., Miller, R., 2003. Secure Web authentication with mobile phones. In Proceedings of Student Oxygen Workshop, Cambridge, England.

Ylonen, T., 1996. SSH - secure login connections over the Internet. In Proceedings of the 6th Security Symposium. pp. 37.

Zevina, J. D., Seidenbergb, M. S., 2002. Age of Acquisition Effects in Word Reading and Other Tasks. Journal of Memory and Language 47-1 pp. 1-29.

Zurko, M. E., Simon, R. T., 1997. User-Centered Security. New Security Paradigms Workshop.

Index of Figures

Figure 2.1: *AEGIS activity diagram* 9
Figure 2.2: *Authentication process* 13
Figure 2.3: *Recall-based authentication* 16
Figure 2.4: *Recognition-based authentication examples* 17
Figure 2.5: *Rorschach-style inkblot* 18
Figure 3.1: *ImagO concept example* 26
Figure 3.2: *ImagePass concept example* 26
Figure 3.3: *Example for the ImagePass mobile prototype.* 28
Figure 3.4: *Total number of comments grouped according to category and prototype.* 30
Figure 3.5: *Human information processing* 32
Figure 3.6: *Sample grid presenting abstract images from the experiment* 35
Figure 4.1: *ImagePass grid* 44
Figure 4.2: *Username selection* 44
Figure 4.3: *Graphical password selection* 45
Figure 4.4: *Current Selection panel* 45
Figure 4.5: *Login screen* 46
Figure 4.6: *Authentication screen* 47
Figure 4.7: *Database structure* 52
Figure 4.8: *Final web version* 55
Figure 4.9: *ImagePass mobile application on a Samsung Galaxy S* 56
Figure 5.1: *Categorizations of visual areas for ImagePass screens.* 71
Figure 5.2: *Attraction plots for the Password selection screen (left) and the Password confirmation screen (right).* 73
Figure 5.3: *Comparison of eye tracking heat maps between genders* 74
Figure 5.4: *Sample scanpath for ImagePass* 75
Figure 6.1: *Split screen for faux search tasks* 80
Figure 6.2: *Categorizations of visual areas on ImagePass screens* 81
Figure 6.3 *Login screen sample gaze plot & aggregated heat map* 83
Figure 6.4: *Create username sample gaze plot & aggregated heat map* 83
Figure 6.5: *Number of fixations per row* 84
Figure 6.6: *Number of fixations per column* 85
Figure 6.7: *Sample gaze plot and aggregated heat map for the GMOH profile* 86
Figure 6.8: *Sample gaze plot and aggregated heat map for the LMT profile* 86
Figure 6.9: Sample gaze plot and aggregated heat map for the NP profile 87
Figure 6.10: *Sample gaze plot and aggregated heat map for the WWI profile* 88

Figure 6.11: *Probability distribution chart for selecting an image in a graphical password* 101

Index of Tables

Table 3.1: *Number of comments for both prototype concepts across different categories.* 29
Table 3.2: *Descriptive statistics for recognized images in the STM test* 36
Table 3.3: *Kolmogorov-Smirnov for normal distribution of recognized images in the STM test* 37
Table 3.4: *Results of Friedman's ANOVA for the STM test* 37
Table 3.5: *Wilcoxon signed-ranks test for all pair variables in the STM test* 38
Table 3.6: *Significance of Wilcoxon signed-rank for the STM test* 38
Table 3.7: *Descriptive statistics for recognized images in the LTM test* 39
Table 3.8: *Kolmogorov-Smirnov for normal distribution of recognized images in the LTM test* 39
Table 3.9: *Results of Friedman's ANOVA for the LTM test* 39
Table 3.10: *Wilcoxon signed-ranks test for all pair variables in LTM test* 40
Table 3.11: *Significance of Wilcoxon signed-rank test for LTM test* 40
Table 4.1: *Graphical password space comparison* 49
Table 4.2: *Sample table with random ID's* 50
Table 5.1: *Gender distribution for participants per clustering group* 63
Table 5.2: *System log summary* 64
Table 5.3: *Descriptive statistics for login analysis based on clustering groups* 65
Table 5.4: *Focus group session structure* 66
Table 5.5: *Summary of data gathered during password selection* 72
Table 5.6: *Summary of data gathered during password confirmation* 72
Table 6.1: *Descriptive statistics for fixation lengths on AOIs for Login screen* 82
Table 6.2: *Descriptive statistics for fixation lengths on AOIs for Select Password screen* 84
Table 6.3: *Summary statistics for experiment variables* 92
Table 6.4: *Significant correlations among the measured variables* 93
Table 6.5: *Final categorization and tagging results* 93
Table 6.6: *Distribution of selected images by category* 94
Table 6.7: *Chi-square test statistics for category* 94
Table 6.8: *Results from the Mann-Whitney test on category* 95
Table 6.9: *Expected and actual frequency of colors in images* 96
Table 6.10: *Chi-square test statistics for colors* 96
Table 6.11: *Results from the Mann-Whitney test on color* 97
Table 6.12: *Expected and actual frequency of shapes in images* 97
Table 6.13: *Chi-square test statistics for shape* 98

Table 6.14: *Results from the Mann-Whitney test on shape* 98
Table 6.15: *Descriptive statistics for password length groups* 98
Table 6.16: *Descriptive statistics for login analysis based on clustering groups* 99
Table 6.17: *Probability matrix for selecting an image in a graphical password* 100

Appendix A – Mnemonic Story

This is a copy of the mnemonic story given to participants in the user study (Buzan, 2000).

1. Mnemonic techniques

The Greeks so worshipped memory that they made a goddess out of her - Mnemosyne. It was her name from which was derived the current word mnemonics, used to describe memory techniques such as the one you are about to learn. The Greeks had intuitively realized that there are two underlying principles that ensure perfect memory: imagination and association. Quite simply, if you want to remember anything, all you have to do is to associate (link) it with some known or fixed item, calling upon your imagination throughout.

In order to remember well, you have to use every aspect of your mind. To do so, you must include in your associated and linked mental landscape the following:

1. **Color.** The more colors you use, and the more vivid they are, the better. Using color alone can improve your memory by as much as 50 per cent.
2. **Imagination.** Your imagination is the powerhouse of your memory. The more vividly you can imagine, the more easily you will remember. Sub-areas within imagination include the following:
   - Expansion: the more gigantic and enormous you can make your mental images, the better.
   - Contraction: if you can clearly imagine your picture as extremely tiny, you will remember it well,
   - Absurdity: the more ridiculous, zany and absurd your mental images are, the more they will be outstanding and thus the more they will be remembered.
3. **Rhythm.** The more rhythm and variation of rhythm in your mental picture, the more that picture will weave itself into your memory.
4. **Movement.** As often as possible, try to make your mental images move. Moving objects are usually remembered better than still ones.
5. **The Senses:** tasting, touching, smelling, seeing, hearing. The more you can involve all your senses in your memory image, the more you will remember it. For example, if you have to remember that you have to buy bananas, you stand a far better chance of not forgetting your task if you can actually imagine smelling a banana as you touch it with your hands, bite into it with your mouth and taste it, see it as it is approaching your face, and hear yourself munching it.
6. **Sex.** Sex is one of our strongest drives, and if you apply this aspect of yourself to your magnificent daydreaming ability, your memory will improve.
7. **Sequencing and Ordering.** Imagination alone is not enough for memory. In order to function well, your mind needs order and sequence. This helps it to categorize and structure things in such a way as to make them more easily accessible, much in the same way than if that same information were simply dumped randomly on the floor.

8. **Number.** To make ordering and sequencing easier, it is often advisable to use numbers. Many of the memory systems throughout this book will teach you simple and advanced methods for memorizing using number aids in different ways.
9. **Dimension.** Use your right-brain ability to see your memory images in 3D.

In each memory system there is a Key Word. This word is the 'Key Memory Word' in that it is the constant peg on which the reader will hang other items he or she wishes to remember. This Key Memory Word is specifically designed to be an 'Image Word' in that it must produce a picture or image in the mind of the person using the memory system. Thus the phrase 'Key Memory Image Word'.

You will soon realize the importance of being sure that the pictures you build in your mind contain only the items you want to remember and those items must be associated with or connected to Key Memory Images. The connections between your basic Memory System Images and the things you wish to remember should be as fundamental and uncomplicated as possible:

- Crashing things together
- Sticking things together
- Placing things on top of each other
- Placing things underneath each other
- Placing things inside each other
- Substituting things for each other
- Placing things in new situations

2. The Link System

The link system is a basic memory system used for memorizing a short list of items, such as shopping lists, in which each item is linked to or associated with the next. Imagine, for example, that you have been asked to shop for the following items:

- a silver serving spoon
- six drinking glasses
- bananas
- pure soap
- eggs
- biological washing powder
- dental floss
- whole-wheat bread
- tomatoes

Instead of scrambling around for little bits of paper or trying to remember all the items by simple repetition and consequently forgetting half of them, use the following memory principles.

Imagine yourself walking out of your front door perfecting the most amazing balancing trick: in your mouth is the most enormous silver-colored serving spoon, the handle-end of which you are holding between your teeth as you taste and feel the metal in your mouth.

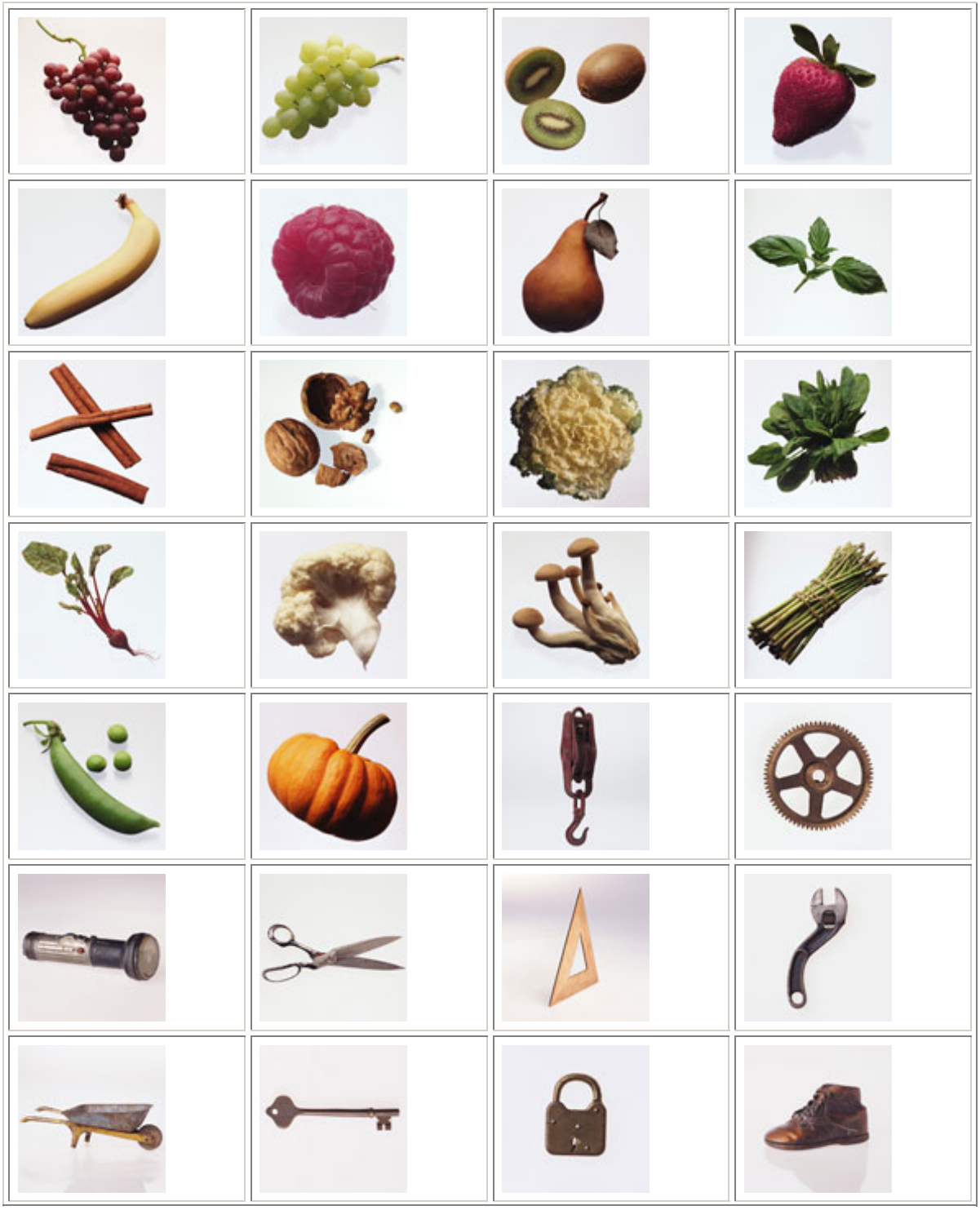
Carefully balanced in the ladle-end of the spoon are six exaggeratedly beautiful crystal glasses, through which the sunlight reflects brilliantly into your bedazzled eyes. Going outside into the street, you step on the most gigantic yellow banana, slip, and start falling. Being a fantastic balancer, you manage not to fall and confidently place your other foot ground ward only to find that you have stepped on a shimmering white bar of pure soap.

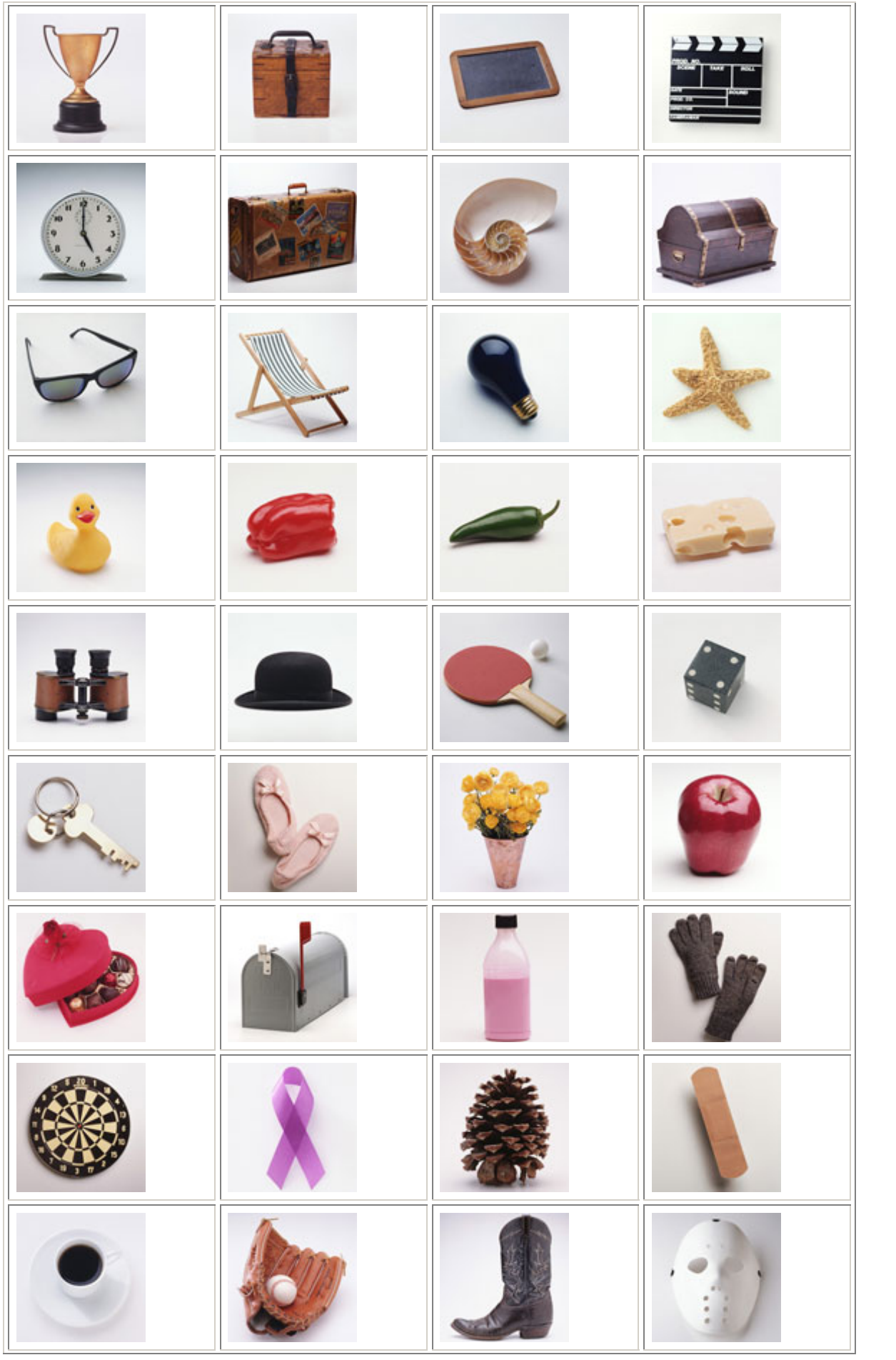
This being too much for even a master like yourself, you fall flat on your back and land down on a mound of eggs. As you sink into them you can hear the cracking of the shells and feel the yellow of the yoke and the white of the albumen soak into your clothes.

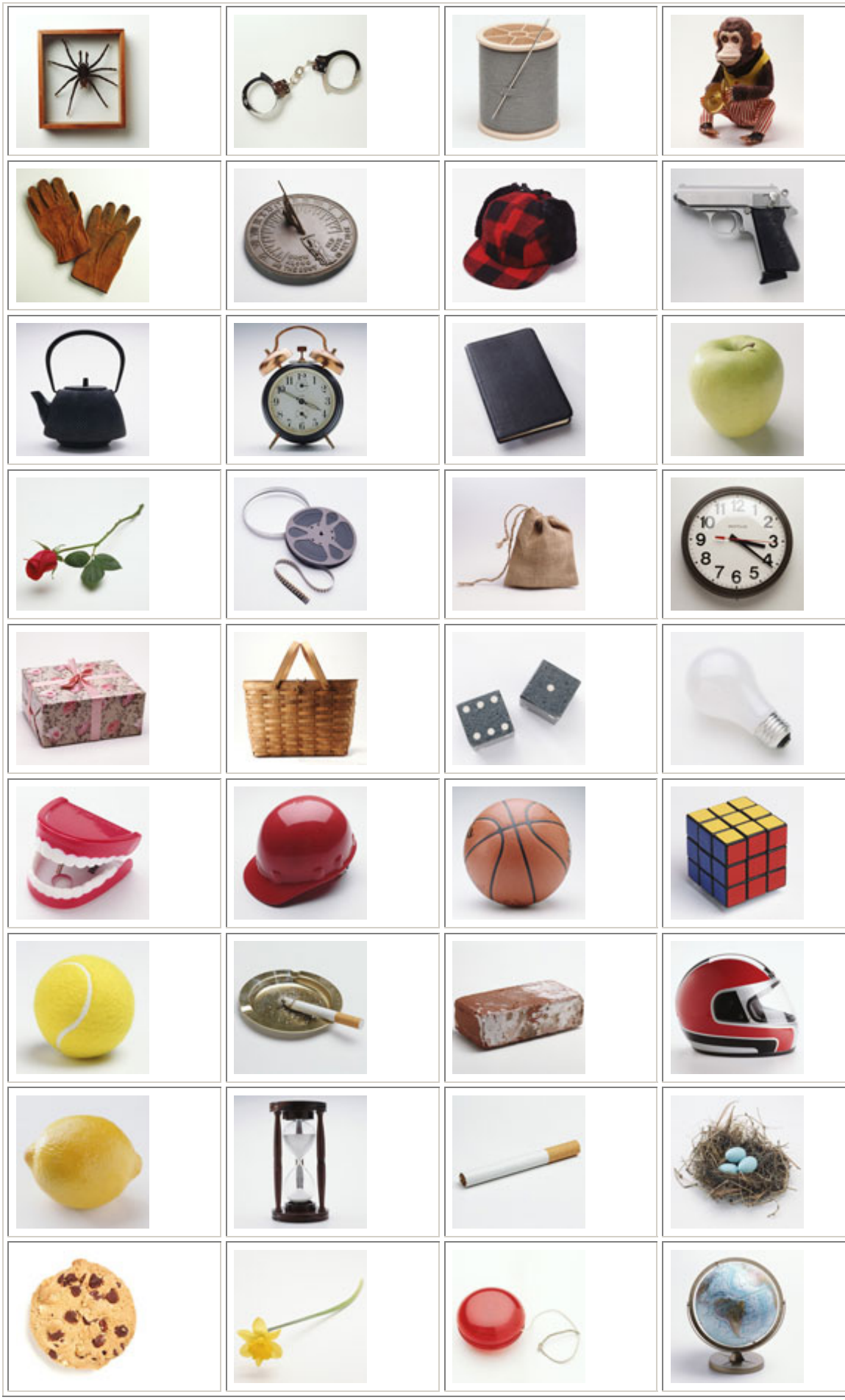
You get up, go back inside, undress and wash your soiled clothes with a super biological washing powder and get back out with your clean clothes. Being tired from the accident you are pulling yourself towards the shop on a gigantic rope made of dental floss.

Just as all this exertion makes you feel hungry, wafting comes an incredibly strong aroma of freshly baked whole-wheat bread. As you enter the baker's shop you notice that every loaf on the shelves if filled with pulsating red tomato bread. You walk out of the baker's shop, noisily munching on your tomato and whole-wheat loaf and go back home.
When you have finished reading this fantasy, close your eyes and run back through the image-story you have just completed. You should remember all nine items on the shopping list

Appendix B – Sample Image Catalog







Appendix C – Author's Publications

Mihajlov M., Jerman-Blazic B., 2011 (in press). Memorability, Performance and Perception in Graphical Authentication. Interacting with Computers, Elsevier. (2010 impact factor: 1.192).

http://dx.doi.org/10.1016/j.intcom.2011.09.001

Mihajlov M., Jerman-Blazic B., Ilievski M., 2011 (in press). Recognition-based Graphical Authentication with Single-Object Images. In Proceedings of the 4th International Conference on Developments in eSystems Engineering.

Mihajlov M., Jerman-Blazic B., Ilievski M., 2011 (in press). ImagePass – Designing Graphical Authentication for Security", In Proceedings of the 7th International Conference on Next Generation Web Services Practices. (ERA Rank: C).

Mihajlov M., Jerman-Blazic B., Josimovski S., 2011. A Conceptual Framework for Evaluating Usable Security in Authentication Mechanisms – Usability Perspectives. In Proceedings of the 5th International Conference on Network and System Security, pp. 332-336 (ERA Rank: B).

Mihajlov M., Jerman-Blazic B., Josimovski S., 2011. Quantifying Usability and Security in Authentication. In Proceedings of the 35th Annual IEEE Computer Software and Applications Conference, pp. 626-629 (ERA Rank: B).

http://dx.doi.org/10.1109/COMPSAC.2011.87

Mihajlov M., Jerman-Blazic B., Saikayasit R., 2010. ImagePass - Developing A Graphical Authentication Mechanism Based on Usable Security. Human Factors in Information Security Inagural Conference Poster Session.

Mihajlov M., Josimovski S., Trenevska-Blagoeva K., 2009. Evaluation of an ePP Solution. In Proceedings of the 9th International Symposium on Communication and Information Technology pp. 124-127 (ERA Rank: B).

http://dx.doi.org/10.1109/ISCIT.2009.5341274

Chorbev I., Mihajlov M., 2009. Building a Wireless Telemedicine Network within WiMAX-based Public Networking Infrastructures. In Proceedings of the 11th IEEE International Workshop on Multimedia Signal Processing, pp. 1-6 (ERA Rank: B).

http://dx.doi.org/10.1109/MMSP.2009.5293305

Chorbev I., Mihajlov M., Joleski I., 2008. WiMAX Supported Telemedicine. In Proceedings of the 30th International Conference On Information Technology Interfaces, pp. 589-594 (ERA Rank: C).

http://dx.doi.org/10.1109/ITI.2008.4588476

Kakasevski G., Mihajlov M., Arsenovski S., 2008. Evaluating Usability In Moodle Learning Management System. In Proceedings of the 30th International Conference On Information Technology Interfaces, pp. 613-618 (ERA Rank: C).

http://dx.doi.org/10.1109/ITI.2008.4588480

Chorbev I., Mihajlov M., 2008. Wireless Telemedicine Services As Part Of An Integrated

System For E-Medicine. In Proceedings of the 14th IEEE Mediterranean Electrotechnical Conference, pp. 264-269.

http://dx.doi.org/10.1109/MELCON.2008.4618445

Chungurski S., Arsenovski, S. Mihajlov M., 2006. SCORM Based System for Building Learning Objects. In Proceedings of the 15th IEEE International Electrotechnical and Computer Science Conference.

http://www.ieee.si/erk06/program.pdf

Kraljevski I., Gacovski Z., Arsenovski S., Mihajlov M., 2005. Choosing Optimal Parameter Values for ANN Classificator in Voice Recognition System. In Poster Session Abstracts of the 27th International Conference on Information Technology Interfaces (ERA Rank: C).

Kraljevski I., Gacovski Z., Arsenovski S., Mihajlov M., 2005. Performance of DTW Speech Recognizer on Packet Switched Network. In Proceedings of the 7th National Conference ETAI.

Arsenovski S., Kraljevski I., Mihajlov M., Jolevski I., 2003. Evaluation of SCORM Based Learning System. In Proceedings of the 4th Conference on Informatics and Information Technologies.

Arsenovski S., Stojchev V., Mihajlov M., 2003. Ilinden Uprising in a Electronic Learning System. Science Assembly "100 Years of Ilinden", MANU.