Title
Protecting the login session from camera based shoulder surfing attacks

Permalink
https://escholarship.org/uc/item/5d91n0jn

Author
Almaula, Varun Kartik

Publication Date
2008

Peer reviewed|Thesis/dissertation
Protecting the Login Session from Camera Based Shoulder Surfing Attacks

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science
in
Computer Science

by

Varun Kartik Almaula

Committee in charge:
Professor Stefan Savage, Chair
Professor Jim Hollan
Professor Hovav Shacham

2008
The thesis of Varun Kartik Almula is approved, and it is acceptable in quality and form for publication on microfilm:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Chair

University of California, San Diego

2008
TABLE OF CONTENTS

Signature Page........................................................................................................................................... iii

Table of Contents ........................................................................................................................................ iv

List of Figures ............................................................................................................................................. vi

Acknowledgements .................................................................................................................................... vii

Abstract ................................................................................................................................................ viii

1. Chapter 1 - Problem Statement ......................................................................................................... 1

2. Chapter 2 - Background and Related Work ......................................................................................... 3
   2.1 Traditional Passwords ..................................................................................................................... 3
   2.2 Graphical Passwords ....................................................................................................................... 6
   2.3 Graphical Password Examples ....................................................................................................... 9
      2.3.1 Pure recall systems ................................................................................................................... 9
      2.3.2 Cued recall systems ................................................................................................................ 13
      2.3.3 Recognition based systems .................................................................................................... 18
      2.3.4 Graphical Password Classifications ....................................................................................... 24
      2.3.5 Biometrics ............................................................................................................................. 25

3. Chapter 3 - Establishing the Security Context for Graphical Passwords ........................................ 27
   3.1 Graphical Password Attack Vectors – Establishing the Shoulder Surfing Threat ....................... 27
   3.2 Defense Mechanisms – Shoulder Surfing ...................................................................................... 30
   3.3 Goals for Login Mechanisms – Ideal vs. Practical ...................................................................... 37
   3.4 Establishing Metrics – Quantifying the Shoulder Surfing Defense .............................................. 39

4. Chapter 4 - Initial Project .................................................................................................................... 40
   4.1 Design ........................................................................................................................................... 40
   4.2 Implementation ............................................................................................................................... 41
   4.3 Results & Analysis ......................................................................................................................... 42

5. Chapter 5 - Research Project .............................................................................................................. 44
   5.1 Design and Incorporation of Contrast Inversion Approach .......................................................... 44
   5.2 Design Goals and Context ............................................................................................................. 47

6. Chapter 6 - Experiment Methodology ............................................................................................... 49
   6.1 Technology Study .......................................................................................................................... 50
      6.1.1 Objective ................................................................................................................................ 50
      6.1.2 Format .................................................................................................................................... 50
   6.2 User Study ..................................................................................................................................... 52
      6.2.1 Objective ................................................................................................................................ 52
      6.2.2 Format .................................................................................................................................... 52

7. Chapter 7 - Results ............................................................................................................................... 53
   7.1 Technology Study .......................................................................................................................... 53
      7.1.1 Mobile Phone: Sony Ericsson Z525a ...................................................................................... 53
      7.1.2 Mobile Phone: Sony Ericsson W580i ...................................................................................... 55
      7.1.3 Point and Shoot Camera: Canon A610 ................................................................................... 56
      7.1.4 ‘Prosumer’ Point and Shoot Camera: Kodak Z812IS ............................................................... 58
LIST OF FIGURES

Figure 1: Entering password doodles in DAS. ................................................................. 10
Figure 2: Components of the PassGo system. ................................................................. 11
Figure 3: Passgraph for Haptic based password ........................................................... 12
Figure 4: Predefined tap regions in Blonder’ graphical password scheme ................. 13
Figure 5: Example password click points in PassPoints scheme ............................... 14
Figure 6: Icon permutations and possible strings in the scheme proposed by Hong et al. 16
Figure 7: Example login window in scheme presented by Hong et al. ....................... 17
Figure 8: Creating a portfolio in the Déjà vu system ................................................... 18
Figure 9: Examples of VIP1/2 and VIP3 login screens ............................................... 20
Figure 10: Sample Passfaces login screen ................................................................. 21
Figure 11: Enclosed password click area formed by bounds of pass icons in Convex Hull Scheme .... 22
Figure 12: Aligning pass icons to the border frame in movable frame scheme ........ 23
Figure 13: Classifications for various graphical password systems .......................... 24
Figure 14: Four challenge rounds to enter a digit in the cognitive trapdoor game ........ 32
Figure 15: Two separate images form a hybrid image which changes interpretation based on viewing distance ........................................................................................................... 35
Figure 16: Disabling the recording capabilities of a camera from over a football field away with a laser pointer ................................................................................................................................. 36
Figure 17: Screenshots of the black screen, the white screen with icons, and the white screen .......... 41
Figure 18: Progression of steps (1), (2), and (3). All screens are displayed as full screen images (the entire display is filled) ........................................................................................................... 46
Figure 19: Sony Z525a, Sony W580i, Canon A610, Kodak Z812IS, Sony HDR-FX1 .......... 51
Figure 20: Determining the contrast inversion delay threshold for the Sony Z525a (** indicates that a post-processing filter was required to obtain results) ................................................ 54
Figure 21: Determining the contrast inversion delay threshold for the Sony W580i (** indicates that a post-processing filter was required to obtain results) ................................................ 55
Figure 22: Determining the contrast inversion delay threshold for the Canon A610 (** indicates that a post-processing filter was required to obtain results) ................................................ 57
Figure 23: Post-processing the original Canon A610 video frame (left) confirms that the arrow grid has been compromised (right) .................................................................................................................. 58
Figure 24: Determining the contrast inversion delay threshold for the Kodak Z812IS (** indicates that a post-processing filter was required to obtain results) ................................................ 59
Figure 25: Post-processing the original Kodak Z812IS video frame (left) confirms that the arrow grid has been compromised (right) .................................................................................................................. 60
Figure 26: Determining the contrast inversion delay threshold for the Sony HDR-FX1 (** indicates that a post-processing filter was required to obtain results) ................................................ 61
Figure 27: Login success rates at various contrast inversion delay settings .............. 63
Figure 28: Time elapsed for a successful login at the 30ms contrast inversion setting ........ 64
Figure 29: Aggregate graph of percentage of challenge screens compromised during a login session ................................. 65
Figure 30: Aggregate graph of percentage of pass icons compromised during a login session .... 66
ACKNOWLEDGEMENTS

I would like to thank Professor Stefan Savage, Professor Jim Hollan, and Professor Hovav Shacham of my advisory committee. They provided valuable insight and the freedom to develop a solution on my own, yet gave excellent advice and guidance on how to focus my thesis contributions. Much of the inspiration for my research can be attributed to attending fascinating lectures in their CSE 227 – Computer Security and COGS 120 – Human Computer Interaction courses.

I would also like to thank my friends and classmates who have made my undergraduate and graduate years at UCSD fly before my eyes and left me with many fond memories. I’m also thankful for having worked with Taurin Tan-Atichat on the original project and his insight throughout my thesis.

Finally, I’d like to thank my parents who have given me every form of support I could ask for and only asked that I enjoy myself and find happiness in return.
ABSTRACT OF THE THESIS

Protecting the Login Session from Camera Based Shoulder Surfing Attacks

by

Varun Kartik Almaula

Master of Science in Computer Science

University of California, San Diego, 2008

Professor Stefan Savage, Chair

Fraud and identity theft account for large financial loses each year. One critical component present at virtually all login terminals used today is the traditional text password or numerical PIN. Compromising this authentication component is a critical, yet increasingly feasible, step for criminals looking to gain access to sensitive information. To increase usability and security, graphical password schemes have been proposed that rely on users’ innate ability to recognize pictures for authentication.

Sparked by the general ubiquity of cell phones with video recording technology, our research aims to identify the susceptibility of the Passfaces graphical password system to camera based shoulder surfing attacks while introducing a novel defense to combat recording devices. First, we survey several graphical password schemes in the context of the shoulder surfing attack vector. We then propose a method of subverting video recording capabilities by using high frequency contrast inversion. By preventing cameras
from successfully correcting elementary settings in response to contrast inversion, little useful information is extracted from the system.

We mate the contrast inversion defense to an already proven user input method, a ‘cognitive trapdoor game’, and place these mechanisms within a graphical password system. A technological study evaluates the effectiveness of these mechanisms against a variety of recording devices, while a user study evaluates how these mechanisms impact the usability benefits of a graphical password system. We believe our system increases the resource investment for a camera based shoulder surfing attack, yet can be used in most authentication terminals without extensive modifications.
1. **Chapter 1 - Problem Statement**

The stakes are high and the consequences are clear. A 2007 Identity Fraud Survey Report uncovered nearly $50 billion in reported identity theft losses [1]. Phishing attacks alone claimed a staggering $3.2 billion in 2007 [2]. Those who fell prey to such attacks were not let off easy either. Another survey found that the median loss per identity fraud case was over $31,000 [3]. And each year brings fresh victims: 2006 saw 8.4 million Americans succumb to identity fraud, bringing the 2004-2006 total to nearly 27 million [1]. Identity fraud is a profitable business that welcomes any and all patrons.

Beyond the impersonal aggregate statistics, the effects of identity fraud on corporations and individual consumers are even more disturbing. Simple password resets, let alone cases of identity fraud, can require extensive support and cost up to £40 per incident. Extrapolating this figure for a company with more than 10,000 employees, each requiring a password reset once a year, represents a significant productivity and monetary loss [4]. For victims of identity fraud, the costs extended far beyond the per incident median loss of $31,000. The median resolution time per victim was as high as 40 hours in 2005 [1]. Even worse, identity fraud can inflict severe pain on victims: the emotional impact has been “found to parallel that of victims of violent crimes” [1]. Other impalpable indicators such as ‘consumer confidence’ can be adversely affected too.

What drives this rampant criminal activity? The aforementioned statistics provide a simple answer: identity fraud supports a lucrative business for criminals, where the low risk of facing justice is easily overshadowed by the prospects of quick money. This is not to oversimplify the complexities of this underground economy. It is an intricate ecosystem with criminals establishing a clear division of labor to reward all those involved. An inside look at a past identity theft ring that utilized the Gozi malware exposes that it is not the technology used to obtain victim information that is sophisticated. Instead, advanced development has been directed towards creating a software portal to distribute this information to support a “service-based economy” that has evolved electronic crime from an “episodic problem” into something akin to “drug trafficking run by syndicates” [5]. Tracking identity thieves is next to impossible, especially when they are sheltered across international borders. Even after investigating significant time and resources
in tracking a US identity theft case all the way to Benin, Africa and ultimately to a face-to-face meeting with the criminal in Switzerland, Chris Hansen, the host of the popular “To Catch a Predator” series, was unable to bring the criminal to justice on the show “To Catch an ID Thief”[6]. The wild-goose chase was mildly amusing, even admirable, but it brought to light the general futility in bringing these criminals to justice after the crime has already been committed.

There are a variety of attack methods that criminals utilize to obtain sensitive information from their victims. Such attacks may include social subversion/engineering, physical information security attacks, or technological attacks. One common thread among many attacks is that they exploit the reliance of a security system on an easily obtainable text password or numerical PIN to authenticate a particular user. The focus of this research attempts to improve upon these traditional authentication mechanisms to make it far more resource intensive for a criminal to recover a sensitive password. By using a graphical password system in tandem with a cognitive trapdoor game and contrast inversion, sharing, observing, or understanding the meaning of a user’s password will require a significant investment of time and resources by a criminal.
2. Chapter 2 - Background and Related Work

Typical interactions with sensitive information on a computer system are guarded by a three step process [7]. The first step is identification, which establishes who is attempting to interact with the system. Generally, identification is accomplished by providing a unique username to the computer system. The next step, authentication, attempts to verify that the current user is actually the user identified in the identification step. This step is often achieved by presenting a challenge to the user and requiring the user to answer the challenge with a secret known only to the system and the correct user. Once the user has been authenticated, access and actions on resources are granted or denied based on whether or not the authenticated user is authorized to interact with these resources. The primary focus of this research is concerned with authentication, although cutting edge approaches such as biometrics, blur the line between identification and authentication.

Authentication mechanisms fall into three categories: knowledge-based mechanisms, token-based mechanisms, and biometric-based mechanisms [4, 8]. Knowledge-based mechanisms authenticate a user by something the user knows. Token-based mechanisms authenticate a user by something the user possesses. Finally, biometric-based mechanisms authenticate a user by something the user is, e.g., the uniqueness of a user’s anatomy or behavior. Most security systems currently utilize knowledge-based mechanisms to authenticate a user: a user is authenticated upon providing a secret password to the system. Token-based systems are predominantly used for banking services where the user presents a magnetic card during the authentication process. However, since tokens can be forged or stolen, banking services employ a secondary level of authentication by requiring a knowledge-based mechanism in addition to the token. Simply presenting a token is insufficient to authenticate a user in these instances and thus, the token is used mainly for identification purposes, which is analogous to a username in the identification process [4].

2.1 Traditional Passwords

The most widespread authentication mechanism used presently is the familiar textual password or numerical PIN. Not surprisingly, criminals focus a large amount of effort in obtaining this key component
of the authentication process to gain unauthorized access to sensitive information. Unfortunately, these traditional password mechanisms have both inherent and practical drawbacks when used in real-world situations. The crux of the problem is the seemingly impossible constraints forced upon the user. Birget eloquently illustrated the two conflicting requirements [9, 10] for a password system:

1. Passwords should be easy to remember, and the user authentication protocol should be executable quickly and easily by humans.
2. Passwords should be secure, i.e., they should look random and should be hard to guess; they should be changed frequently, and should be different on different accounts of the same user; they should not be written down or stored in plain text.

Ignoring other crucial components of these requirements, a seemingly impossible paradox is presented to the user: all accounts should have a unique lengthy and random password, yet recalling any single password should be simple. Traditional passwords have shown acute shortcomings in trying to satisfy both requirements.

Criminals use four primary methods, sometimes in combination, to obtain user passwords [11]:

1. Information harvesting to be able to guess a user password
2. Social engineering to manipulate users to reveal confidential information
3. Direct observation of a user password as it is entered
4. Cracking a user password using a brute-force method

A main susceptibility of traditional passwords is that they are easy to share, either via speech or writing, because they are simply a collection of text characters. Users are by nature vulnerable to carefully planned social engineering attacks [12-14], but the ease by which passwords can be shared makes it trivial to launch a successful pretexting [14] or phishing attack. Likewise, since almost all traditional passwords are entered ‘in the clear’ by the input device, such as a keyboard or number pad, spyware and shoulder surfing attacks can easily observe a user’s password. Shoulder surfing is a direct observation attack that utilizes either a human or remote recording device to literally look over the user’s shoulder as secret password information is inputted into the system, while spyware tends to be installed remotely onto a user’s computer to monitor any data or action on the user’s computer that is deemed valuable. Note, while passwords are often
encrypted once entered into the system, rarely are there any attempts to obfuscate the actual physical key input: it is a direct mapping between password characters received by the system and the characters presented on the keyboard.

Further complicating matters, the reality of account management in the personalized internet age has made it nearly impossible for users to keep multiple secure passwords. It is hard enough to remember one or two random strings or PIN passwords, but remembering 20 or 30 secure passwords for multiple websites, bank accounts, and other services can only be expected of those few with exceptional memory. These types of passwords do not contain meaningful information and require rote memorization, a poor method of remembering [10]. This problem is exacerbated when websites use password authentication just to conduct simple identification and bookkeeping of user activity, e.g., reading public news articles or browsing an online store can require a username and password [15]. Faced with such a daunting task, users cope by actively decreasing the complexity of the security mechanism. Password creation guidelines are ignored and passwords are reused among many accounts, written down, and shared with others. A study attempting to quantify the prevalence of password reuse in response to the scaling problem for online accounts finds a positive correlation between the password reuse ratio and the number of accounts per user [15]. Even users with a small number of online accounts tended to reuse passwords. Users were willing to accumulate large numbers of online accounts, but did not attempt to produce the same number of unique passwords, leaving the door open for Trojan site attacks on dormant accounts to gain access to multiple accounts. Another interesting observation from the study revealed that users made up their own risk management schemes by reusing passwords for accounts with little personal information linked to them while using stronger passwords for information sensitive accounts.

There have been attempts to bridge the gap between the two conflicting requirements on passwords. One approach is to use mnemonics, deriving a password from a permutation on a memorable phrase, to simultaneously increase password strength and improve recall rates [11, 16]. However, two studies [11, 16] have shown that even mnemonic passwords are susceptible to brute force attacks. By creating mnemonic password dictionaries based on popular quotes from media along with simple permutations, user chosen mnemonic passwords were compromised at a surprisingly high rate of 4% [11].
While the shortcomings of traditional passwords have been repeatedly exploited by criminals, the
advantages of traditional passwords have helped them remain entrenched as the primary authentication
mechanism. Traditional passwords require no additional hardware than what is used with virtually all login
terminals including older computers, physical security systems, and ATM machines: a keyboard or PIN pad
and a basic display. Furthermore, traditional passwords are relatively simple to revoke and administer [17,
18]. Despite best efforts, passwords can be compromised or forgotten; thus, it is vital that the administrative
duties of password revocation and creation be painless and, ideally, automated. Successful new
authentication mechanisms will need to simultaneously seek to address the shortcomings of traditional
passwords and strive to retain low implementation and administrative costs.

2.2 Graphical Passwords

Authentication mechanisms which rely on images or icons to verify a user’s identity have garnered
much interest in the security community for their potential to provide solutions to many of the weaknesses
exhibited by traditional passwords. The most intriguing aspect of graphical passwords is that they utilize a
strength of human memory, visual memory, as opposed to a weakness of human memory, alphanumeric
memory.

Human memory is governed by long-term memory (LTM) limitations and the Power Law of
Forgetting. Specifically, the Power Law of Forgetting states that humans go through a phase of rapid
forgetting after learning, but experience a slower decay over the long-term [10]. Information is more likely
to be remembered accurately if it is recalled, or refreshed, within the decay period of the memory type: for
complex alphanumeric strings, this decay window is much smaller than the decay window for images.
Humans have an innate ability, according to the picture superiority effect, to remember “a vast, almost
limitless” amount of visual information [19]. Placed in the perspective of today’s real world landscape,
where having 20 infrequently used accounts with unique, and often, interfering passwords is common
practice, graphical password systems seem far better suited as an authentication mechanism than traditional
alphanumeric passwords.
LTM enhances visual memory because it allows humans to store the meaning of an image, not simply a replica of it [10]. This allows graphical password systems to only require imprecise recall of a secret image in comparison to precise recall of a secret alphanumeric string: a user can forget specific details of a complex image and still log on successfully, so long as they are able to retain the ability to recognize the meaning of an image [8]. With traditional alphanumeric passwords, this sort of imprecise recall often leads to a failed log on attempt.

Another benefit of graphical password systems is that graphical passwords are not easily shared between two entities. Sharing the password at work, home, or even under the pressures of a sophisticated social engineering attack is a difficult endeavor since it is nontrivial to recreate the image by writing it down or verbally describing it [20, 21]. A successful attack would require a direct breach of the system by either the user directly copying an image screenshot and sharing it, or an outside attacker gaining control of the system to grab an image screenshot. While both of these examples are still very real threats that would need to be addressed in a commercial system, they do demonstrate the added threshold of control required to successfully compromise a graphical password. Likewise, there is potential for graphical password systems to be less observable and recordable by attackers, but the level of susceptibility greatly depends on how the system is implemented. If the system does not try to obfuscate the input of the user’s response to a challenge image, the system would be just as or more vulnerable to an observation or recording attack. However, if the system does try to obfuscate the user’s input, the mouse and/or keyboard events being sent to the system may not have any direct utility to an attacker. In comparison, keyboard strokes for entering an alphanumeric password are easily recordable and observable because there is no attempt to obfuscate the mapping between the keystroke and the secret password characters.

Often, the strength of a password mechanism is defined by its password space, or the amount of unique passwords that can be generated within the system. This size of this password space is governed by three factors:

1. The theoretical number of unique passwords that can be generated in the system
2. The constraints imposed by the system design that limits this theoretical password space
3. The constraints imposed by human tendencies and memory limitations that limits this theoretical password space

While the size of the theoretical password space of a system, factor (1), is often touted as a strong security metric, a stronger indicator of system strength is the size of the usable password space, the remaining password space after factors (2) and (3) are taken into account. In comparison to a simple 6-8 character alphanumeric password, most graphical password systems offer a significant increase in the theoretical size of the password space. Images contain a wealth of information that can be permuted on several factors such as image size, pixels, meaning, color, and content, affording very large password spaces [22]. Additionally, because visual information is generally more memorable than alphanumeric information, the constraints of factor (3) in a graphical password system do not shrink the password space as sharply as they would in a traditional password system. Human tendencies and image preferences can still be manifested in a graphical password system, so care in determining how passwords are administered and how user choice impacts password selection is necessary. Despite this, assigning random images from an image database to a user is more likely to be successful than assigning a random password because humans have a stronger visual memory. Factor (2) will also define the practical use of the theoretical password space. In graphical password systems, such constraints may include the actual number of images in the database used to present a challenge image set to the user or how the system defines the user interaction with a set of images.

The usable password space of an authentication mechanism determines the vulnerability of that mechanism to an automated cracking or guessing of the secret password. Creating a dictionary for a graphical password system is resource intensive or altogether infeasible: there are not any well known sortable or searchable graphical dictionaries for raw image content and automated processing of an image for interpretation requires a non-trivial compute time [22]. Since graphical passwords have a generally higher usable, or memorable, password space, it is harder to create sophisticated guessing attacks that exploit human tendencies and known weak passwords. It is when graphical password systems are placed in real world settings that potential problems may occur. The principal threat to most graphical password systems occurs during the password creation phase when a user is allowed to choose an image, a portfolio
of images, or click points within an image as their secret [4, 8, 10, 19, 20, 23]. To alleviate the dangers of self-selected passwords, it is best if password administration and creation is automated.

2.3 Graphical Password Examples

Password systems can be classified as pure recall, cued recall, or recognition based systems [4, 23]. In a pure recall system, a user is expected to reproduce the secret password without any hints or contextual information. These systems tend to put increased cognitive stress on the user since it requires a precise remembering of the secret information. A cued recall system still requires the user to reproduce the secret password, but also provides the user a framework of cues and contextual information to aid in remembering the secret information. Recognition based systems only require the user to locate and recognize images from a larger set that includes decoys to authenticate themselves: because the user need not precisely reproduce the shared secret, memory retention and authentication success rates tend to be much higher.

2.3.1 Pure recall systems

The Draw-A-Secret (DAS) scheme introduced by Jermyn et al. allows users to input a graphical ‘doodle’ via a stylus on a grid for authentication [24].
While allowing for an arbitrary stroke length for a password theoretically produces a seemingly infinite password space, the scheme does suffer from a small grid size and a difficulty of inputting diagonal lines. Combined with user tendencies in inputting doodles, the usable password space of the system is far more constrained.; a study by Thorpe and van Oorschot found that users inherently chose symmetric shaped doodles for their password, which would potentially impact the security of the mechanism [19].

PassGo is another grid-based scheme which aims to encourage users to remember more complex passwords in a larger password space [25]. This is achieved by increasing the grid size, allowing users to select intersection points instead of simply cells, and allowing for diagonal lines. To aid users in reproducing their doodle in a larger grid, specific reference cells are shaded to give context and dot indicators are used to indicate stroke completion.
Figure 2: Components of the PassGo system.

The hope is that by introducing shaded cells, the user will have a better ability to precisely reproduce the password. However, real world results from a class login webpage showed that login rates did not stabilize until 8 weeks into the course, indicating that there exists an extended learning and trial phase for this mechanism to be successful.

Yet another grid based scheme is QDAS, which improves on DAS by annotating each cell with an integer to increase memorability and using qualitative spatial relations to store passwords [26]. Storage of QDAS passwords only consists of the starting cell and the following direction changes indicated by the stylus input to capture a stroke. Additionally, the grid dynamically resizes as the doodle is being entered to create a level of obfuscation against observation attacks.

An innovate grid based scheme incorporates pressure in a doodle, to create a Haptic-based graphical password [17]. Using these pressure signatures in combination with a graphical password increases the possible password space and also introduces a strong level of obfuscation against observation.
attacks: observers will not be able to extract any pressure information from a user login attempt by merely visually recording the input.

![Example of passgraph; bold lines indicate high pressure](image)

Figure 3: Passgraph for Haptic based password.

The system does not give pressure feedback to the user to avoid giving attackers any hints about the user’s password. Unfortunately, this means the user will have to go through an extensive trial period to establish pressure levels and will also require the user to remember and reproduce correct pressure levels at each login attempt. Furthermore, the system will require specialized and standardized hardware to be effective. While it seems well suited for PDA and laptop computer applications with a standardized pressure input method, the system cannot be used on desktop computers or at ATMs without specialized peripheral hardware.

While pure recall graphical systems do utilize visual memory in the form of a unique doodle, they do little to change the cognitive demands imposed on the user. Users tend to be able to visualize an imprecise representation of their doodle, but are less likely to be able to precisely reproduce the original doodle at a login terminal [19, 21]. Reproducing the correct cell or intersection positions, stroke order, and stroke length requires focused dexterity and extensive practice, especially for the more secure and complex doodles encouraged by these systems. From a feasibility perspective, not all login terminals are equipped to handle freehand input, e.g. an ATM, and would require non-trivial retrofit designs to incorporate these hardware-based graphical password solutions. Even terminals that do support freehand input, e.g. PDAs
with styluses, laptops with touchpads, and desktops with mice, would require users to adapt to non-standardized input methods, each with varying levels of quality, for inputting their doodles.

### 2.3.2 Cued recall systems

Blonder introduced an initial graphical password scheme that presented an image with predefined 'tap regions' on the picture [27]. A user clicks on an ordered subset of these tap regions to indicate their graphical password.

![Predefined tap regions in Blonder’s graphical password scheme.](image)

Since users cannot choose arbitrary locations on the image, the usable password space is not large enough to warrant a definitive break from the traditional alphanumeric password scheme.

The PassPoints system is also a click-based scheme, but addresses the usable password space by using more intricate images and making any pixel within an image an eligible click point [10]. Both the theoretical password space and the usable password space greatly benefit from these modifications, albeit the usable password space does not increase as sharply as the theoretical password space.
Figure 5: Example password click points in PassPoints scheme.

The system experience is largely defined by two parameters, the click point tolerance and the image content. The user cannot be expected to click on the exact same pixel at every login, due to memory constraints and input imprecision, and thus the system must allow for an error tolerance region around the secret click point. This creates a tradeoff between improving security, using a smaller error tolerance, against improving usability, using a larger error tolerance. Studies have shown this to be a delicate, but achievable, balance to strike: a 10x10 vs. a 14x14 pixel error tolerance can produce significantly different login success rates [10, 28]. Another study found that images with meaningful content in only a few areas produce ‘hotspots’ that users tend to choose as part of their password and thus reduce the usable password space [23]. Such images would need to be avoided in order to confidently prevent dictionary attacks in a commercial version of this system.

Another method of cued recall is to create associations between items that are known only by the system and the user. The system presents some information as a challenge to the user, the user then performs a secret function, the association, on the challenge information, and finally the user responds to the system with the correct information resulting from the secret function [29, 30]. The idea is relatively abstract and can have many implementations of varying effectiveness and complexity. Some simple PIN pad tricks involve creating new associations between the labeling on the keypad and the information inputted by the key. For example, the user may only need to remember a path on the PIN pad as the secret
information, such as upper right, middle, lower left, lower middle. The labeling of the keys on the path, e.g., using a subset of numbers or alphabet characters, can be changed on every login and thus the user input will be dynamic and obfuscated to an observer each session. More advanced schemes can create completely new associations in which a user may be required to remember an association between images and numbers or characters: the system may label PIN pad keys on screen with images and the user will need to remember the association between each image and the appropriate number or character to successfully authenticate.

Some complex associative schemes have been developed to make it extremely difficult to extract any useful information during an observation attack on a graphical password system. One scheme, WIW, is resistant against shoulder surfing attacks by utilizing complex associations to create time-varying passwords [31]. WIW requires the user to create and memorize a new alphabet derived from perturbations on a set of icons. At every login attempt, the user is presented with $h$ different scenes, that each contain $k$ pass objects hidden among $n$ objects, with each pass object having $m$ possible perturbations for a given scene instance: the user must memorize $h \times k \times m$ strings for the alphabet. The system presents these scenes to the user as a challenge, after which the user computes a response function to create a string vector composed of alphabet strings based on the specific scene instance. A similar system proposed by Hong et al. allows for users to confidently authenticate themselves on a machine that is already compromised by spyware [32]. A user is required to memorize a set of portfolio images and a whole set of permutations on each image, e.g., an image of a face can be happy, sad, etc. For each permutation for a given image, the user must create and memorize a string associated with the image.
Thus, at a login session, the user is presented with a large set of decoy and pass images in multiple grids. For each grid, the user will locate his or her pass icon, determine its permutation, recall its associated string, and enter one component of their response to the system.
In this scheme, the user will be required to memorize $k \times m$ strings, where $k$ is the number of pass icons and $m$ is the number of permutations per icon. While both these complex associative schemes provide high levels of obfuscation and resistance to observation attacks, the systems may not pass a threshold of practicality to be effective. There is a high memory and cognitive load imposed on the user for every authentication attempt. Furthermore, creating and memorizing a large number of associations is a potentially long and tedious endeavor, which, using traditional passwords schemes as real world evidence, often forces the user to cope by subverting the system security mechanisms: in these schemes, the user may decide to simply create the same string for all permutations on a given icon to reduce the number of unique strings to be memorized.

_Cued recall_ systems are a marked improvement over _pure recall_ systems, in that they attempt to jog the user’s memory by presenting familiar cues and context during a login session. However, these systems still require the user to precisely reproduce the original secret information: if the mouse click in a click-based authentication system or the string association for a given image in an associative system is not...
accurately entered, the login attempt will fail. A strength of many cued recall systems, such as PassPoints or WIW, is their large theoretical and usable password space. One caveat is that systems with large password spaces may increase the time required for the user to scan through decoy and pass information on screen, leading to longer overall login times at an authentication prompt.

2.3.3 Recognition based systems

Déjà Vu addresses human memory limitations in recalling secure passwords by authenticating users on their ability to recognize previously seen images [8]. The system strives to prevent the use of weak passwords, prevent the sharing or writing of passwords, and eliminate the necessity of precise recall for successful authentication. A user’s password is a collection of images that comprises the user’s portfolio. At the login screen the system will present a subset of the user’s portfolio mixed with a larger set of decoy images and will authenticate the user when the user picks the correct subset of images.

To minimize the risks associated with user selected passwords and to help storage requirements, Déjà Vu utilizes Random Art that can generate random images based on initial seeds. Results from initial studies show an overwhelming improvement for successful login rates in favor of Déjà Vu over traditional PIN or
alphanumeric passwords. While the system has the potential to generate an infinite amount of unique images, the practical password space is constrained because both portfolio and decoy images must be presented as a manageable set on a single authentication screen. The total number of images presented on screen for Déjà Vu is 20, which includes 5 portfolio images, and provides comparable resistance against brute force attacks as a traditional PIN system. While this small practical password space is of concern, it should be considered in context of larger problems: the major drawbacks of PIN and alphanumeric password systems are the lack of memorability and ease of casual password sharing. Déjà Vu elegantly proposes solutions to these drawbacks while maintaining the same level of resistance to brute force attacks as current PIN systems. While not necessarily true for alphanumeric passwords, PIN passwords at ATMs are less likely to be subjected to brute force attacks, especially considering that a simple account lock after three failed login attempts would stem the attack. Other pertinent system design choices for Déjà Vu include how many stages to break up the authentication process into, how many decoy and portfolio images to present in each stage, whether to cycle or re-cycle portfolio and decoy images on failed login attempts and between unique login sessions, whether to have static or random locations for images, and how to obfuscate the user’s input to protect the system against observation attacks.

Visual Identification Protocol (VIP), a system similar to Déjà Vu, is the subject of a study by Angeli et al. seeking to answer some of these outstanding questions for recognition based graphical password systems [19]. The study examines whether graphical passwords conclusively have better memory characteristics than traditional password mechanisms and how poor graphical system design can nullify the picture superiority effect. Three versions of the VIP interface are tested against the traditional PIN interface: static image location and ordered image selection (VIP1), random image location and ordered image selection (VIP2), and random image location and unordered image selection (VIP3).
Figure 9: Examples of VIP1/2 and VIP3 login screens.

The systems are evaluated along three security metrics: guessability, observability, and recordability. Results show that systems with fixed locations, PIN and VIP1, benefit from motor memory caused by repetition and exhibited low login error rates with fast login times. Conversely, systems with random image locations, VIP2 and VIP3, exhibit higher login error rates and longer login times due to the increased visual scanning required for users to locate their password images. VIP3 is considered a pure recognition based system since there are no locations or sequences to memorize and offers better protection against direct observation attacks and password sharing. In this scheme, VIP3 is designed as a single stage login screen, meaning multiple password portfolio images are mixed with decoy images. This results in a smaller password space since the entry of password images is allowed to be an unordered sequence. Instead, using a random image location setup in a multi stage login sequence, with one password image mixed with decoy images at each stage, would help maintain the benefits of VIP3 without incurring such a large password space hit. The picture superiority effect is validated in the study for all three VIP variations, but only after significant time had past since the training phase. This indicates that system design and password administration in a graphical password system are still crucial components that affect memorability, especially in the learning and trial phases: while the picture superiority effect is explicit in the context of long term memory decay, it is not a cure-all for the entire authentication system.

One of the most popular recognition based graphical password systems is Passfaces, which authenticates users based on the inherent human ability to recognize and process human facial information efficiently and accurately [20, 33]. Like the aforementioned recognition based systems, Passfaces users
have an image portfolio with secret images that comprise a password. However, these images are unique human faces which are presented over multiple rounds to users. Each round displays one face from the portfolio amid eight other decoy faces in a 3x3 grid. In each round, users select the correct face from their portfolio from the 3x3 grid until their image portfolio is exhausted or a preset number of rounds have been completed (usually 4).

![Sample Passfaces login screen.](image)

Results show that users of the Passfaces system login successfully at an exceptionally high rate in short-term trials and continue to exhibit high success rates even after long periods without interaction with the Passfaces system [4]. Furthermore, users require 2/3 fewer password reminders, suggesting that even a failed login attempt is productive in that it is sufficient to aid users in jogging their memory to recall an image in their portfolio. One concern from the study is the added resource requirements for storage, CPU power, and network bandwidth associated with Passfaces. While these are legitimate concerns, these results are quite skewed by the decade old hardware used in the experiment: with more modern hardware and network connectivity, a Passfaces authentication service should be able to deployed with minimal resource penalties and few, if any, hardware upgrades. Another concern with the PassFaces system is the potential for human tendencies in the facial image selection phase to influence the security of the system [34]: user preferences for gender or race could limit the usable password space of the system. To avoid this situation, commercial versions of Passfaces prevent user selection of faces for the portfolio.
Two systems that blur the line between cued recall and recognition based systems are the Convex Hull Click (CHC) and movable frame schemes [9, 22]. In addition to requiring users to recognize previously seen pass icons, these schemes utilize an innovative challenge-response function to obfuscate user selection. By obfuscating user selection, these schemes offer a high level of defense against direct observation attacks, including attacks that record user input with a capture device such as a camera. Both schemes present a large array of small icons, all of which are decoys except a small subset, e.g. 3 to 4, of pass icons. In the CHC scheme, users are required to locate their pass icons and click anywhere inside the convex hull formed by their pass icons. This process is repeated for multiple rounds to increase the security of the system.

![Figure 11: Enclosed password click area formed by bounds of pass icons in Convex Hull Scheme.](image)

In the movable frame scheme, users locate three of their pass icons, one of which is located on the outer frame of the screen. The user extends an imaginary line from the two pass icons within the frame and uses the mouse to move the frame so all three pass icons are aligned.
The challenge-response component in both schemes allows users to recognize and respond to their secret pass icons without explicitly disclosing which icons are from their portfolio. This level of obfuscation is essential since most graphical password systems force the user to reveal their pass icons by directly clicking on them with a mouse or stylus. Unfortunately, to achieve a high level of obfuscation and a large enough usable password space to combat brute force attacks, these schemes require large numbers of small icons to be displayed on a large high-resolution screen. The cognitive load demanded of users is non-trivial, with typical login times exceeding one minute as users parse through hundreds of icons and compute a response for each round [22]. Within the correct security context, such as a sensitive military environment, the added time and concentration required to complete the authentication process may be met with greater acceptance.

Recognition based systems provide some of the most compelling results for memory recall rates among graphical password systems. Since these systems only require imprecise recognition as opposed to precise reproduction, users are far more likely to remember their secret passwords, even during long periods of infrequent use. Recognition based systems do suffer from smaller practical password spaces in contrast to cued recall and pure recall systems because there is a finite limit to the amount of image information that can be presented to and analyzed by the user. To reach a threshold of acceptable security, recognition based systems tend to require multiple rounds of authentication [23], and, in the case of the
CHC and movable frame schemes, also require a user to compute a response to a challenge. The result of these additional mechanisms means that users experience far longer login times than with traditional alphanumeric passwords: a secure alphanumeric password may take 5-10 seconds to enter correctly, whereas a secure password in a recognition based graphical password system may take 30 seconds to over a minute to enter correctly.

2.3.4 Graphical Password Classifications

In addition to the aforementioned authentication mechanism classifications, graphical password systems are also grouped into three prevalent techniques: cognometrics, locimetrics, and drawmetrics [19, 21]. Cognometric systems rely on the cognitive abilities of humans to identify ‘target’ images among decoy images and to compute basic functions on information as a response. Locimetric systems require the user to identify targets points within an individual image. Drawmetric systems require the user to reproduce a pre-drawn image on a grid. A synopsis of various authentication systems and their appropriate classifications follows:

<table>
<thead>
<tr>
<th>Pure Recall</th>
<th>Locimetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Haptic DAS variant</td>
<td>1. Blonder</td>
</tr>
<tr>
<td>2. DAS</td>
<td>2. Pass Points</td>
</tr>
<tr>
<td>3. QDAS</td>
<td></td>
</tr>
<tr>
<td>4. PassGo</td>
<td></td>
</tr>
<tr>
<td>Cued recall</td>
<td></td>
</tr>
<tr>
<td>1. Blonder</td>
<td></td>
</tr>
<tr>
<td>2. Pass Points</td>
<td></td>
</tr>
<tr>
<td>3. Associative graphical password scheme</td>
<td>4. PassGo</td>
</tr>
<tr>
<td>4. Spyware resistant password scheme</td>
<td></td>
</tr>
<tr>
<td>5. WIW</td>
<td></td>
</tr>
<tr>
<td>Recognition</td>
<td></td>
</tr>
<tr>
<td>1. Deja vu</td>
<td></td>
</tr>
<tr>
<td>2. VIP</td>
<td></td>
</tr>
<tr>
<td>3. PassFaces</td>
<td></td>
</tr>
<tr>
<td>4. Convex Hull Click CHC/movable frame</td>
<td>5. Convex Hull Click CHC/movable frame</td>
</tr>
<tr>
<td>5. handwing</td>
<td>6. Associative graphical password scheme</td>
</tr>
<tr>
<td></td>
<td>7. Spyware resistant password scheme</td>
</tr>
<tr>
<td></td>
<td>8. WIW</td>
</tr>
</tbody>
</table>

Figure 13: Classifications for various graphical password systems.
2.3.5 Biometrics

Biometric authentication verifies the identity of a user by using the individual’s physical anatomy or behavioral patterns. A digital sample of a user’s unique characteristic, e.g. iris, fingerprint, keystroke patterns, etc., is stored by the security system and is compared to a live digital sample extracted from the user at the login terminal. In general, biometrics removes the memory load on the user during the authentication process: the user simply needs to present the desired personal characteristic for verification. In essence, the system challenge is simply to ask the user to be present, and the user response function is to allow the login terminal access to extract a sample. For behavioral biometrics this user response function might be more complex and require the user to actually perform a set of actions to be authenticated.

Depending on how the system is deployed, biometrics can actually combine both identification and authentication into a single step. The identification of the user is so strong and unique, especially in comparison to a traditional username, that a secondary mechanism to verify identity may be unnecessary. Certainly, visions of the future often include a user walking up to an ATM and only completing a simple iris scan to allow access to the user’s bank accounts. While this idealistic vision of the future depicts biometrics as a panacea to authentication security challenges, there are currently a plethora of unresolved practical considerations that, at the very least, confirm that this vision is still in the distance.

Several issues arise from the storing of a digital template of the user’s unique information. Two scenarios, ‘failure to enroll’ and ‘failure to acquire’, are digital sampling issues [35]. There is some portion of the population that will not be able to enroll in the system because these individuals do not possess the biometric information or are unable to use the system, e.g. a user whose fingerprint has been altered through physical trauma. Likewise, the system will be unable to successfully acquire a sample from a portion of the population because the system-user interface breaks down, e.g. a user whose iris cannot be sampled because of colored contacts or an eye patch.

Assuming the system is able to enroll and acquire a proper digital sample, there are still additional hurdles. Since digital samples are approximations in nature, there is a level of inaccuracy in the system governed by two parameters: the false accept rate (FAR) and false reject rate (FRR) [35]. The FAR refers to the likelihood the system incorrectly authorizes a user with the wrong identity, while the FRR refers to
the likelihood the system incorrectly denies authorization to a legitimate user of the system. These turning parameters are interconnected, and a real world system will have to judge whether it is more appropriate to err on the side of security or usability.

Another concern with current biometric systems is cost and feasibility. Biometric systems require specialized hardware as well as costs associated with inserting the new technology in current terminals, user training, and initial digital sample acquisitions. Behavioral biometrics do not necessarily require specialized hardware, but may not achieve the same level of accuracy and performance as physical biometrics.

The issues relating to sample acquisition, sample accuracy, and hardware costs are challenging, but certainly within the reach of technological innovation within the next decade. More abstract issues relating to social acceptance and personal privacy are less likely to be easily overcome. One study investigating user acceptance of fingerprint biometric authentication finds that users consider application contexts when using the mechanism: the benefits of using biometrics for securing personal information during personal purchases is far more well received than when using biometrics for corporate purchases [36]. Additionally, users overwhelmingly preferred biometrics for its ease of use over its perceived security benefits. Another study reveals that users have strong personal aversions to biometric systems before even using them [35]. Only after an extensive field trial with an iris verification mechanism at an ATM machine, did users overcome their negative perceptions of biometric technology and gain enough confidence to use such systems in the future. Clearly, biometrics will have an increased role in the security of personal information in the future, but it must first overcome by demonstration the deep seated social issues and privacy concerns of users to gain acceptance. Biometrics does not strike a balance between security and usability in itself. While it improves ease of use in terms of user memory load and login speed, the security model of a computer system must change significantly: not only is the user’s personal information relating to the application important, e.g. a credit card, SSN, etc., but now the actual shared secret to authenticate the user, e.g. a biometric digital sample, is of even higher importance than a current traditional password. If this information is breached, the term ‘identity fraud’ takes an even more sinister meaning. What recourse do users have to reclaim their identity if their non-revocable iris or fingerprint sample is stolen?
3. Chapter 3 - Establishing the Security Context for Graphical Passwords

Before delving into design and implementation, it is important to establish the security context of this research project based on attack vectors, application environments, security and usability goals, and the metrics decided upon to evaluate a security system. There is no single security system that is equally immune to all attack vectors, usable in all application contexts, achieves all security and usability goals, and meets all standard evaluation metrics: some systems might exhibit extreme defenses against automated cracking attacks, are geared towards use in a military environment, and sacrifice login speed and personal privacy whereas other systems might exhibit strong shoulder surfing resistance, are geared for financial transactions, and strive to protect personal privacy. Either system can be evaluated and utilized as a successful or unsuccessful security system depending on the context of use. Thus, with that in mind, the following sections explore the threats, applications, and metrics related to this research project.

3.1 Graphical Password Attack Vectors – Establishing the Shoulder Surfing Threat

As described earlier, criminals rely on four primary methods, sometimes in combination, to obtain user passwords [11, 20]:

1. Information harvesting to be able to guess a user password
2. Social engineering to manipulate users to reveal confidential information
3. Direct observation of a user password as it is entered
4. Cracking a user password using a brute-force method

Graphical passwords systems can increase the required resources and significantly diminish the threat for most of these attack vectors. Information harvesting to gain insight on a user’s graphical password selection tendencies is difficult to conduct, especially if the graphical password is administered randomly and not user selected. Generic social engineering attacks are also less likely to reveal a graphical password since the graphical information is difficult to write down or verbally communicate. Of course
social engineering attacks are still possible, e.g. asking a user to send a digital copy of their picture password to a remote location, but the required breach is far more severe and the resources required have been increased significantly. Cracking a graphical password has not necessarily been proven to be an effective attack simply because there has not been any definitive method of aggregating a searchable and orderable graphical image dictionary to launch an attack with. However, depending on the implementation, the usable space of a graphical password scheme can be susceptible to a brute force attack: a cued recall PassPoints style graphical password scheme would have a larger usable password space than a recognition based Passfaces style graphical password scheme. Yet, even Passfaces at least matches PIN password resilience to a brute force attack and arguably improves on the more important facets of authentication, memorability and cognitive load, the main drawbacks of PIN authentication: a random guessing attack can be carried out on a Passfaces style system, but by utilizing automated account freezes after repeated login failures and not unnecessarily revealing feedback on failed login attempts, this attack can be mitigated.

The most glaring attack vector that graphical password schemes tend to be susceptible to are direct observation attacks, specifically shoulder surfing attacks that attempt to record the screen presentation and/or the user input. Graphical passwords are inherently visual by nature, so masking the presentation or the user selection of the secret information is extremely important. Stylus or click based input are especially vulnerable as either a human or electronic method of observation can capture users selecting their passwords in the clear. Relying on users to shield the presentation and input device is not foolproof, as users can easily forget or find themselves unable to adequately shield their actions, especially at busy POS (Point of sale) terminals. Also, more sophisticated electronic attacks with miniature cameras or chemical attacks on input keys can easily extract password information from both traditional and graphical password systems.

Quantifying the prevalence of such observation attacks is extremely difficult because victims are unaware of their sessions being observed or would logically otherwise confront the situation immediately. Fraud reports list attack vector statistics, but do not include shoulder surfing or observation attacks as these statistics are virtually impossible to collect: shoulder surfing statistics might be absorbed into other aggregate statistics such as users who believed their credit card information was fraudulently used,
obtained, forged, collected from a paper or computer record, physically stolen, etc. Instead, most shoulder-surfing schemes are discovered only after the fact, e.g. when video surveillance reveals a human perpetrator or miniature cameras are found near ATM machines. What is apparent is that banks and ATM manufactures do take the shoulder surfing threat seriously. A 2002 Diebold ATM Fraud and Security white paper identifies shoulder surfing as an ATM fraud technique and considered the threat severe enough to warrant advising retrofitting ATMs with mirrors or altogether ergonomically redesigning ATMs to allow for better inherent personal privacy [37]. Many ATMs now utilize polarized screens, but humans or cameras in direct view of the terminal screen can still easily capture sensitive information. Even in 1994, Anderson recognized the threat, describing a direct observation attack at an ATM as early as 1984 in New York as well as a 1994 heist in which criminals in San Jose video recorded ATM machines [38]. Numerous recent reports confirm that common identity fraud does not necessarily utilize high-tech attack vectors. One 2007 report on ID theft reveals that the Internet was the exclusive tool of ID thieves only about 10% of the time [39]. Another 2008 report finds that thieves are increasingly falling back to traditional methods of stealing identity information [40], likely due to the ease of which successful social engineering attacks and information harvesting can be performed. In an informal 3M ‘snooping’ survey, 89% of business travelers admitted to snooping on unsuspected seatmates when traveling and 36% admitted to reading over someone else’s shoulder in a public place [41]. Clearly, both innocent and mischievous eyes wander in public places, increasing the likelihood of sensitive password information being compromised. The proliferation of camera and video enabled cell phones greatly increases this threat as shoulder surfing can be made less invasive, automated, and anonymous. A not too farfetched scenario may include an attacker using a cell phone behind a user at an ATM: the attacker can act as if he is using his cell phone for a benign purpose, e.g. dialing a call or writing a text message, while waiting in line, but is actually photographing or video recording the user’s entire ATM session. An automated attack with a remote camera is equally disconcerting. With these attack vectors in mind, we seek to understand existing graphical password systems’ defenses against shoulder surfing in order to create a novel design that mitigates the threat.
3.2 Defense Mechanisms – Shoulder Surfing

Of the various graphical password schemes discussed in the related work section, a handful address the shoulder surfing threat. These and other schemes rely on two basic methods of deterring a shoulder surfing attack: obfuscating the presentation of information displayed and/or obfuscating the input of information by the user. Many schemes also introduce ‘noise’ in either the output or input of information, making it difficult for an attacker to discern the valid information among the noise, while ensuring the system and user will be able to do so relatively easily.

The Convex Hull Click (CHC) and movable frame schemes obfuscate the output of information by adding noise and obfuscate the input of information by requiring the user to compute a challenge action [9, 22]. The presentation of the secret pass icons is hidden among a large amount of decoy icons. By requiring users to either click within the convex hull of their pass icons or move the frame of the screen to align their pass icons, attackers are unable to disambiguate which specific icons caused the user to perform the action. In essence the user has to imprecisely compute a one-to-many function, recognizing his pass icons and computing the action, whereas the attacker has to reverse engineer a many-to-one function, figuring out every possible set of icons that could have resulted in the user performing that action. By combining these methods, the schemes are able to provide a high level of defense against shoulder surfing attacks by a human or a video recording. Moreover, an attacker would need several recordings of a user’s login sessions to gain enough information to begin to decode their possible pass icons. This level of resource investment in time and computation discourages shoulder surfing attacks. Unfortunately, the cost in average login time for the user is also likely to be too high for widespread acceptance. Hundreds of decoy icons need to be presented on screen to achieve a high level of ambiguity, resulting in a high cognitive load for the user, typical login times of a minute or more, and the need for a high resolution display. While the schemes may be well suited for specific applications, the degree to which usability is sacrificed makes it unlikely that they will be used at ATMs, POS terminals, or consumer PCs.

The WIW scheme and a similar scheme proposed by Hong et al. also claim to be resilient against multiple video recordings and spyware [31, 32]. The user response to the presentation of graphical images and their permutations is obfuscated by a large set of associations for each image permutation. The
drawback of these schemes is that the user is responsible for creating and remembering an extremely large set of image-string pairs, increasing both the cognitive load and time demanded of the user for every login session. In these cued recall schemes, generating the noise and obfuscation introduced to the system is the user’s responsibility. Password administration and the learn/trial phases in these schemes is complex, requiring the user to come up with and memorize a large number of associations.

Recognition based graphical password schemes seem to be able to exhibit a good balance of resilience against information harvesting, social engineering, and brute force attacks while retaining low cognitive loads for users and tolerable login session times. However, recognition based systems such as Passfaces are vulnerable to direct observation attacks because the user’s input directly highlights the secret graphical password image during the selection phase: when Passfaces is used with a mouse, the user hovers the mouse pointer over and clicks the secret image, making it a trivial exercise for a devious onlooker to learn the user’s password. Acknowledging this threat, the Déjà vu system creators suggest “the method for the image selection should be hidden” by either obfuscating the selection of images on screen or the key presses for image selection so an observer is unable to obtain any useful information [8].

This type of obfuscation is further examined by Tari et al. for use with the Passfaces system [20]. In addition to having users input image selection choices with a mouse, the Passfaces system is also configured to accept key input from the number pad. Since the Passfaces system utilizes a 3x3 grid to present images onscreen, the mapping between the image locations in the grid and a standard 3x3 number pad on a keyboard is simple and logical to understand for the user. Results of a user study show that users that used Passfaces with a mouse were far more susceptible to a shoulder surfing attack than those who used a keyboard. In the shoulder surfing attack simulation, the ‘attacker’ was given the freedom to move anywhere around the user to create an optimal viewing angle and a notebook and pencil to record any relevant information. As expected, analysis of the user study shows that attackers had a high success rate of snooping passwords when users used the mouse: the delay induced by moving the mouse pointer over the selected image, gave enough time for the attacker to recognize and record the user’s secret image. However, when a keyboard was used, attackers were unable to observe both the keyboard and the images being presented on screen at the same time: the user input in response to the onscreen images was sped up
to a threshold that prevents direct observation attacks with the naked eye. While these results are promising, the study mentioned but did not test more sophisticated shoulder surfing attacks using cameras or how using a laptop without a number pad would affect system performance.

Roth et al. introduced a ‘cognitive trapdoor game’ applied to a traditional PIN pad to combat shoulder surfing [42, 43]. The system utilizes the existing PIN pad design, and does not require any additional hardware so long as a monochrome screen is present to display a ‘soft’ onscreen PIN pad. The premise of the system is to obfuscate the user’s choice inputs so that an attacker can observe both the presentation onscreen and the user input and be unable to deduce the user’s password. The method of obfuscation is a cognitive trapdoor game, a game which is simple to play for the parties that know the secret PIN, the system and user, and difficult for the party that does not know the secret PIN, the attacker. Users enter a PIN digit by responding to four challenges: for each challenge, the system divides the PIN pad into a set of black colored keys and a set of white colored keys and users respond with a binary answer indicating which set contains their PIN digit.

![Figure 14: Four challenge rounds to enter a digit in the cognitive trapdoor game.](image)

After four challenges, the system can deduce the user’s PIN digit choice. However, an attacker trying to play the game and observe the user’s PIN digit would have a very difficult task at hand: each challenge requires an attacker to record the color of each key and observe the user’s binary response. Combined with the user’s high speed of entry, this will ensure an attacker will unable to successfully extract the PIN with the naked eye. Since any single challenge doesn’t reveal the digit, a breach of a single challenge will only net the attacker minimal information; a successful attack requires four consecutive breaches of the system.
to reveal a PIN digit. Additionally, chemical PIN pad attacks that reveal user input choices by fingerprints would be unable to gain information since both keys are likely to be repeatedly pressed during a login session.

Recognizing the threat of camera and video based shoulder surfing attacks, the creators of the system modified the input method because it is trivial to extract a user’s password with a recording device using the aforementioned input method. The new method introduces a ‘probabilistic cognitive trapdoor game’ in which the system presents a similar challenge sequence to the user. However, instead of limiting the possible correct PIN digits to one entry, the challenge sequence limits it two or more entries so an attacker cannot be sure which PIN digit is the correct one [42, 43]. The system will authenticate a legitimate user with high probability, while an attacker with recording capabilities can guess the correct PIN with low probability. In a user study an optimal shoulder surfing scenario was established, with an attacker’s camera with full view of both the screen and keyboard. Results show that the traditional PIN method was compromised 100% of the time, while the full PIN was never extracted from the cognitive trapdoor scheme. Surprisingly, the simulation did not include the shoulder surfing resistant probabilistic trapdoor scheme and also did not allow the attackers to use frame-by-frame analysis to try to recover the user’s password; without frame-by-frame analysis, this only tested whether the attackers could play the game in real time after the login session was recorded. Without the freedom of video analysis for the attackers in the simulation, the user study did not accurately model the video based shoulder surfing threat and thus cannot provide any assurances against such a threat. Other important findings from the simulation’s survey include a user desire for more security coupled with a willingness to accept an increased level of effort during the login process. The results show an increase in the user concentration required to authenticate successfully and a 10x login speed reduction compared to the PIN process. Users were able to login in less than 30 seconds and the authors believe that this level of required concentration and time is around the threshold of what users find acceptable. While the system shows promise as an excellent obfuscation method, it still does not improve on the memory limitations of the PIN system nor does it offer anything better than a probabilistic defense against camera based shoulder surfing threats: the
probabilistic cognitive trapdoor scheme reduces overall security by possibly authenticating impostors and does not offer sufficient defenses against repeated recording attacks on a given user.

Other solutions take a hardware based approach to defend against shoulder surfing attacks. The Haptic-based graphical password scheme [17] discussed earlier is a grid based pure recall scheme that requires the user to reproduce a doodle, similar to DAS [24], but also incorporates a pressure signature that is undetectable by an observer. While a mouse based pressure input might be detectable by scrutinizing muscle twitches in the hand, this scheme utilizes a stylus for input, making pressure changes extremely difficult to observe. The major drawback of this scheme is that it is a pure recall system that still requires a high cognitive load to precisely reproduce the visual doodle. Furthermore, the benefits of visual memory might not extend to reproducing pressure changes in a complex doodle signature. In fact, user training and trials show that users required an extensive trial phase and required practice at every login session to recalibrate relative pressure levels. Curiously, the results did not include any video based shoulder surfing simulations, but it is relatively safe to assume that pressure variations would be hard to detect with a camera. Logistically, this solution would require equipment standardization, extensive hardware costs, and retrofitting for widespread use. Currently the solution seems better suited for PDA style devices.

EyePassword is another innovative shoulder surfing resistant scheme that utilizes cameras to track the user’s pupil movements as keyboard input for a system with an onscreen keyboard [18]. Users authenticate themselves by focusing their eyes on the keys onscreen to enter the password keystrokes. Calibrating the eye tracking software and hardware for each user incorporates the following parameters: eye tracking resolution, user eye height and angle difference from predefined ‘normal’ viewing distance, error or uncertainty radius for each key, user head movements, trigger selection mechanism, and feedback mechanism (auditory beep or onscreen asterisk). Usability results are positive, showing that the system has similar typing error rates and about a 4x time penalty in comparison to traditional keyboard entry. While the results do not include shoulder surfing simulations, the system undoubtedly increases the resource investment needed to successfully compromise the system. Single camera attacks will not be able to observe the keyboard and eye movements, and even two camera attacks require intricate calibration knowledge of the system and a perseverance to cross-reference the information from both cameras.
Currently, the main drawback of this system is the eye tracking hardware is cost prohibitive for large scale use. However, hardware advances and commoditization can lower costs and possibly incorporate the technology in future advanced ATM surveillance cameras. The cost for training, user education, and calibration are also difficult to quantify and are a cause for concern. Ignoring these cost issues, the system still does not improve upon the inherent cognitive load issues associated with complex alphanumeric passwords and while the usability metrics are not too far off from a traditional keyboard, the results did not use complex and random characters to test input time. The system shows great potential as a discreet input method and could be better suited when used in conjunction with a graphical password scheme.

There are a host of other technologies that can be incorporated as shoulder surfing defense mechanisms or can shed insight when creating original methods for alternative authentication mechanisms. Developed within the realm of computer vision and cognition, hybrid images are static images that can have two interpretations depending on the subject’s viewing distance [44, 45]. By keying on the multiscale processing abilities of the human visual system, two images with different meanings are superimposed one over the other, keeping the high frequencies of one and the low frequencies of the other.

The results are quite remarkable, yielding a single image that has vastly different interpretations for a subject one foot away vs. a subject 10 feet away. Hybrid images could be an excellent component of a shoulder surfing defense in a graphical password system, ensuring that only the user directly in front of the screen can view the intended meaning of the image, whereas observers at more distant locations would only see the decoy meaning of the image. The practicality of such a system would hinge on: the ability to auto-

Figure 15: Two separate images form a hybrid image which changes interpretation based on viewing distance.
generate a hybrid image from two distinct images, being able to fine tune the distance at which the interpretation changes, and preventing a remote camera with a zoom lens from extracting the intended meaning.

The shoulder surfing defenses discussed thus far have been passive ones. Another alternative is to seek out an unwanted recording device and actively disable its ability to record sensitive information [46-48]. Systems that can seek out cameras [46, 48] do so by detecting the retroreflective CCD which sends light directly back to its origin. Once the camera’s CCD has been located, all of these systems exploit inherent disadvantages of camera optics: overexposure, lens flare, and blooming. By sending a beam of concentrated and localized light at the camera, the camera’s CCD is overwhelmed with light and is unable to record any useful footage. A laser beam is the easiest means of generating such a light source, but other bright and localized light sources are possible, such as an LED array.

Figure 16: Disabling the recording capabilities of a camera from over a football field away with a laser pointer.

These active systems have been created in response to the ubiquity of camera phones, which weaken system security and personal privacy boundaries. While directing powerful beams of light at cameras is most likely not practical for much else than specialized military applications, these ideas are valuable and can be incorporated into a shoulder surfing defense mechanism.

One approach to defending against keyloggers, software that intercept all keystrokes from a keyboard, is to add random noise into the input. One technique [49] adds a large amount of random decoy keystrokes embedded with the user’s actual keystrokes. By directing the legitimate keystrokes to the appropriate browser text box and letting decoy keystrokes ‘drop to the floor’, the input is disambiguated,
yet still scrambled for most standard keyloggers. This approach of deliberately adding noise to the input or output of information can be a useful tactic as a shoulder surfing defense mechanism, provided the system and user can easily disambiguate useful information from the added noise.

Adding pressure as another parameter for an input device can also frustrate shoulder surfing attacks. Two approaches, one augmenting the mouse with pressure sensitive input [50] and another using a stylus [51], utilize pressure based input as a user interface productivity enhancement. Pressure based input has no observable visual cues and, without explicit pressure feedback presented on screen, allows for additional hidden information to be inputted into the system.

3.3 Goals for Login Mechanisms – Ideal vs. Practical

It is beneficial to describe the ideal characteristics of the ideal authentication mechanism we wish to create for both design and evaluation motivations. Generally, an authentication mechanism should:

- increase the required user concentration level up to, but not past, an acceptable threshold
- impose a low cognitive load on the user, including a model that allows for a successful login even if imprecise recall only allows the user to remember the meaning of the secret information
- for a given attack vector, push an attacker’s resource investment past a threshold such that the attack is not worthwhile
- obfuscate the input and/or the output of information and obscure the meaning or relationship of the secret information from an attacker
- exploit a human strength of the user and a computational weakness of an attack/attacker; exploit a human weakness for an attacker and a computational strength of the system
- strive for universal user enrollment and operability (anyone can use the system)

Specifically, an authentication mechanism should:

- be designed with a defined process such that users cannot subvert or forego security mechanisms
- allow for easy administration and have revocable secret information in the case such information is compromised
• avoid expensive and specialized hardware
• allow an attacker to observe an entire login session using a chemical PIN pad or shoulder surfing attack and prevent him from being able to understand or reproduce a successful login
• promote fast user logins to increase the information i/o rate that an attacker must successfully capture, analyze, comprehend, and reproduce; minimize or eliminate periods of idle activity during the login period that could potentially leak sensitive information
• introduce some amount of randomness/dynamic information in a login session such that multiple sessions are unlikely to be identical
• be a part of a multi-layered security system such that each component can focus on specific threats while compromising any single component leaks little or no information from the system
• store secret information securely and in an encrypted format

Clearly, no existing system fulfills all these goals, and it would be wishful thinking to believe we will propose a new system that does so. Several of these goals are actually inversely related to each other, in that satisfying one goal will negatively impact satisfying another: in this environment, it is the security tradeoff design decisions that define the viability of a system. The system we wish to design should:
• meet PIN security metrics, and strive to beat PIN security metrics where possible
• increase the user concentration level to ensure proper use of the system, but avoid incurring high memory recall demands
• push the resource investment for shoulder surfing attacks to an unfavorable level for attackers, while not adversely increasing the attractiveness of any other attack vectors
• utilize existing common hardware resources and minimize deployment costs as much as possible
3.4 Establishing Metrics – Quantifying the Shoulder Surfing Defense

To evaluate a graphical password system that is resistant to shoulder surfing attacks, the following set of metrics are of interest:

- overall login success and error rates for users, types of errors exhibited
- time to complete a successful login
- amount of information leaked during a successful or an unsuccessful login
- success rate of shoulder surfing attacks at extracting entire password
- amount of images or information extracted by a shoulder surfing attack
- increase in resource investment required to carry out a successful shoulder surfing attack
- minimum number of repeated (multi session) shoulder surfing attacks required on a single user to successfully deduce the user’s password
- analysis of impact on dictionary and brute force attack vectors
4. Chapter 4 - Initial Project

Sparked by the general ubiquity of cell phones with video recording technology, our initial research aims to identify the susceptibility of a system such as Passfaces to camera based shoulder surfing attacks while introducing a novel defense to combat recording devices. Our initial research has shown that Passfaces performs poorly against a video camera shoulder surfing attack. In response, we proposed a method of subverting video recording capabilities by using high frequency contrast inversion. Preliminary results have shown success in limiting camera based shoulder surfing attacks. Our follow-up research is geared towards incorporating the contrast inversion approach in a more complete Passfaces style graphical password system that offers increased security and usability benefits.

4.1 Design

The initial proposed system exposes the deficiencies that cameras and video recording devices have in adjusting critical settings, e.g. focus, aperture, shutter speed, white balance, ISO, etc., in reaction to a scene change in which the amount of light has been almost instantaneously inverted. In comparison, the human visual system is far more adept at adjusting to such a change. By displaying a graphical password in this context, we show that cameras and video recording devices are not able to accurately capture what is displayed to the user. Additionally, to prevent humans and devices from successfully shoulder surfing, graphical passwords are in the form of a set of small grayscale icons, making it extremely difficult for eavesdroppers or devices to recognize images.

While this project is a classic example of the tradeoff between usability and security, the focus is on a proof of concept for the system: the goal is to demonstrate that cameras and video recording devices can be successfully defeated by using inverting images, i.e. the devices should not be able to change camera settings fast enough to capture visual information from the screen. In any practical authentication system, usability is of high concern, but this prototype may not strike the perfect balance.
4.2 Implementation

The prototype system does not use specialized hardware and is a small standalone Java GUI software application. It has been developed and tested on a Dell Latitude laptop with a high resolution LCD. To keep implementation relatively simple, the password length is three grayscale icons. Icons are displayed in sets of four, a 2x2 grid, which map to user input with the directional keys of a keyboard. The prototype repeatedly displays an icon set until the user is able to distinguish the password icon from the decoys and input it with the keyboard. To allow tuning of the system, the amount of time the black and white screens are displayed is configurable as is the ‘darkness’ of the grayscale icons.

![Figure 17: Screenshots of the black screen, the white screen with icons, and the white screen.](image)

The original prototype easily defeats the camera and video recording functions of the Motorola V710 and Sony Ericsson Z525a cell phones. However, the video recording function of standard digital cameras, e.g. Canon PowerShot A610 and SD400, are more difficult to defeat. These cameras can adjust their settings ‘on-the-fly’ quite rapidly. The initial focus and settings lock, activated when half depressing the shutter release button, does not seem to hinder the camera’s ability to quickly change settings during a video capture. Empirically we find that displaying the white screen with the icon set for around 300-400ms can defeat the camera in most cases. However, frame-by-frame analysis shows that a single frame of the white screen with icons is usually captured by the camera. Specifically, we find that the Java GUI application does not draw the entire screen in one frame when switching from the black screen to the white
screen: there is usually a single video frame that captures the upper half of the screen drawn black, and the lower half of the screen drawn white with half of the grayscale icons showing. Because this transition is not quite as dramatic as an inversion from an entirely black screen to an entirely white screen, the camera is able to adjust its settings fast enough to discern the images on the lower half of the screen.

We remedy this situation as follows: in between displaying the black screen and the white screen with icons, we insert a short amount of time where an entirely white screen without icons is displayed. However, this entirely white screen cannot be displayed for too long because the camera will be able to adjust its settings to the bright white screen and capture the grayscale icons in subsequent frames. After tweaking the timings, we are able to adjust the settings such that the entirely white screen need only be displayed for about one frame, or less than 50 milliseconds. In combination with adjusting the darkness of grayscale icons, we are now able to consistently defeat the video recording feature of our test cameras from a variety of zoom settings, shoulder surfing locations, and levels of ambient light.

4.3 Results & Analysis

To test our prototype, we used a Dell Latitude laptop as the authentication terminal. The aforementioned Canon A610 and SD400 cameras were used for video recording. These cameras have a relatively high quality 640x480, 30fps video recording mode and both feature the Digic II processor used in Canon’s higher end pro digital SLR cameras [52, 53]. To allow attackers to analyze video, frame-by-frame analysis was completed in VirtualDub [54]. Three test subjects, all new graduates (22-23 years old) with backgrounds in Computer Science, Mechanical Engineering, and Management Science, completed our study. The subjects alternated being users of the authentication system and potential shoulder surfers. A verbal description of the authentication system and their password was given to users, and a detailed hands-on walkthrough of the digital camera’s video recording function was given to shoulder surfers. Shoulder surfers were told to orient themselves and the camera in whatever way they found most advantageous to steal the password.

The first trial presented the grayscale icons onscreen without the use of contrast inversion: the user would indicate their password selection by entering the appropriate keyboard directional key. This was set
up to crudely model the Passfaces interface. We found that with frame-by-frame analysis of the video, the mock shoulder surfing attacks achieved a 100% defeat rate of this version of the authentication system. Attackers were easily able to capture both the visual information onscreen and the user input on the keyboard with a single camera and analyze the video to discern the password.

The subsequent trials used the same login mechanism augmented with our contrast inversion defense. We found that our users authenticated themselves without error in every trial (100% input accuracy). When the UI was tuned to an ‘easy setting’, i.e. when the white screen with icons was displayed well over 400ms and the grayscale icons were quite dark, we found shoulder surfers were able to pick up one out of every three icon sets displayed. However, the attackers had to fill the entire camera view with the screen to capture the icons, leaving nothing to capture the users input on the keyboard. On a slightly more difficult UI setting, i.e. when the white screen with icons was displayed less than 300ms and the grayscale icons were lightened, we found that shoulder surfers had a 0% defeat rate of the authentication system. Frame-by-frame analysis showed the camera was only able to capture a blank white screen. For this UI setting, our users were still able to successfully authenticate themselves on every trial.

The results show that our system easily defeats cell phone recording devices and still picture captures of cell phones and cameras. Furthermore, when the system is tweaked correctly, it is also possible to consistently defeat the video recording function of most consumer digital cameras. Our users are still able to authenticate themselves with high accuracy and relative ease. For our tests, the camera angle, initial camera settings, zoom, ambient light, and camera view of the system do not seem to affect the deterrence capabilities of the system.
5. **Chapter 5 - Research Project**

5.1 **Design and Incorporation of Contrast Inversion Approach**

The results from the initial project are quite promising, but for the contrast inversion defense to be implemented in a viable system the shortcomings of the prototype need to be identified and addressed. The major drawback of the initial project is that it obfuscates the output of information and the input of information in a single stage. By combining all the defenses in a single stage, the usability and security of the system suffers. Since the graphical information is presented to the user for less than 500ms, the user has a daunting task of ‘digesting’ the information in this short window: the user must identify the four icons, analyze the content of each icon, recognize the secret pass icon, remember its position, and finally enter the corresponding keystroke to indicate a selection. While any one of these tasks is quite simple, accurately completing all of them in a single step imposes a nontrivial amount of ‘mental’ pressure.

To minimize the cognitive load imposed on the user, the system only displayed four elementary grayscale icons during each round. This reduces both the theoretical and practical password space of the system. Since only small, two color icons are being used, the theoretical space of graphical information is smaller than a system using complex color images. Additionally, using only four icons per round compromises the practical password space because an attacker can randomly guess the correct pass icon with a $\frac{1}{4}$ probability for each round. These metrics do not even match a standard PIN pad’s resistance to a brute force attack.

Clearly, the design revisions required to incorporate contrast inversion into a graphical password system hinge on the ability to separate what components of the login process the contrast inversion protects. In hindsight, it is difficult to protect the presentation of graphical information with contrast inversion: system security and usability benefit by using rich and meaningful graphical information. Conversely, the input of information is a much better candidate to use with contrast inversion, provided that the input method also utilizes the terminal screen. In the keyboard version of Passfaces, users indicate the correct location of their secret image by using the keyboard [20]: this input method does not involve the
screen since it uses a direct mapping between onscreen image locations and the number pad locations. Ideally, we would like a different method of input selection that utilizes the screen to create an indirect mapping between keystrokes and position selection.

Numerous existing input obfuscation techniques have been discussed earlier, including the Convex Hull Click [9] and Haptic-based schemes [17]. However, the most amenable scheme to incorporate the contrast inversion defense is the ‘cognitive trapdoor game’ [42, 43]: instead of directly hitting a number pad location to indicate a position, the scheme breaks input selection into a series of onscreen challenges with only two possible keystroke responses by the user. The main drawback of the system is its susceptibility to a camera based shoulder surfing attack because playback of a login session allows an attacker to reverse engineer the user’s input choice. By combining the contrast inversion approach with a cognitive trapdoor scheme, a camera based shoulder surfing attack can be effectively mitigated.

The proposed scheme works as follows:

1. The system presents a black screen with a 3x3 grid consisting of eight decoy images and one password image. This screen is presented until the user recognizes the location of their password image and responds with a spacebar keystroke. The user should make a mental note of this position.

2. The system then presents a black screen with a 3x3 light gray grid with no graphical content for a relatively long period of time (2 seconds). In this round, the user focuses their gaze upon the position of the password image from step (1). The user should maintain their gaze on this location into the next step.

3. The system presents a white screen with a 3x3 light gray grid of arrows for a relatively short time (100 milliseconds). The grid positions are randomly divided into two sets of light gray directional arrows: one set of right arrows and one set of left arrows. This grid

---

1 Similarly to the initial prototype project, upon entering and exiting the white screen with the 3x3 grid of arrows, a white screen with an empty grid will be presented for an extremely short amount of time (20-30ms). Omitting this blank white screen can lead to a situation in which the top half of the screen is black and the bottom half of the screen is white with the grid of arrows because the redrawing of the screen is not instantaneous. This would lessen the effect of the contrast inversion, allow the camera to capture a portion of the arrow grid, and possibly allow the attacker to deduce some information about which location contained the user’s password.
is the first of four challenges in the cognitive trapdoor game to deduce the user’s image selection. As the user continues to focus his gaze on the password image location from step (2), he will see a directional arrow ‘flash’ in this location. The user enters the corresponding keystroke, either the left directional arrow or the right directional arrow, as a response to the system challenge.

(4) The system will repeat steps (2) and (3) three more times to determine if the user can successfully play the cognitive trapdoor game to enter the correct location of the password image in step (1). Each successive challenge will manipulate the two random sets of directional arrow icons so that the system will be able to deduce which location the user is entering.

(5) Steps (1) – (4) are repeated for each of the user’s portfolio of password images (4 images).

Figure 18: Progression of steps (1), (2), and (3). All screens are displayed as full screen images (the entire display is filled).

If the user fails to play any step of the cognitive trapdoor game correctly, the system can be configured to continue through the entire process to avoid ‘leaking’ information to a brute force attacker on failed login attempts.

By utilizing the ‘cognitive trapdoor game’ as an input obfuscation method, the system retains the benefits of a Passfaces style recognition based graphical password system while adding a stronger shoulder
surfing defense mechanism. It also prevents the contrast inversion defense from interfering with the presentation of graphical information, allowing users ample time to discern their graphical password image among decoys. The cognitive trapdoor game is simple to play for users because they must only respond with binary input when presented with a challenge. After recognizing a password image in step (1), the only demand on users is to remember and monitor the position of their password image within the grid for steps (2) – (4). The game presents binary information using just the black and white colors, ensuring users have little information to digest and can make quick decisions to produce the correct response. Additionally, the use of the cognitive trapdoor game with a Passfaces style graphical password system reduces the hardware requirements. No specialized hardware is needed for a general ATM, POS terminal, or personal computer so long as it has a color display and two input values. Examples of possible input methods include a PIN pad, a two button mouse, or a touchpad with two tap regions. This binary input method is advantageous over Passfaces configured with a numerical PIN pad because a number pad is not always available, e.g. laptop keyboards, and it is resistant to chemical keypad attacks since both input values will be used repeatedly.

5.1 Design Goals and Context

This system is designed to augment a Passfaces style recognition based/cognometric graphical password system with a bolstered shoulder surfing defense mechanism. It is not focused on ‘re-inventing the wheel’ on recognition based graphical passwords, since multiple studies have shown the memory retention benefits of such systems [4, 8, 19, 25]. Instead, the focus is on accurately modeling, defending, and testing against a real world shoulder surfing threat model: one can expect such an attack will utilize at least one, maybe two, consumer level recording devices, e.g. cell phones and point and shoot compact cameras. We wish to raise the minimum threshold of a successful shoulder surfing attack beyond this common equipment to an unappealing level that requires investing in specialized equipment and spending unbearable amounts of time. Perhaps an SLR camera or HD video camera could extract some amount of information, but attackers wielding such equipment are far less likely to ‘blend in’ with the surroundings. In practical use, our system would likely be used as part of a layered security system. An ATM would
utilize a token, e.g. an issued card; an online bank site would utilize an id number or login string; a POS or computer terminal might use biometrics or a login name.
6. Chapter 6 - Experiment Methodology

To evaluate our proposed system we choose to separate the experimentation into two separate components. One component, the technology study, focuses on methodically trying to ‘break’ the contrast inversion defense over a variety of parameters. The other component, the user study, focuses on determining how the contrast inversion and cognitive trapdoor game interact with a recognition based graphical password system.

Previous works that claim a level of shoulder surfing defense either omitted a formal real world test of the introduced mechanism [9, 17, 18], used test subjects as attackers and did not equip them with cameras [20], or used cameras but did not allow test subjects to complete frame-by-frame analysis to break the mechanism [42]. The main drawback of studies that utilize subjects as model attackers is that the subjects are not familiar with the system: expecting them to recreate a successful shoulder surfing attack against a system they have just been introduced to does not model the real world threat. Furthermore, studies that include cameras to model a shoulder surfing attack, but do not allow for frame-by-frame analysis, fail to accurately recreate the attack vector. In real world attacks, cameras are utilized to replay and edit video information offline, making it far easier to successfully compromise a system. By separating the technology and user studies, we hope to accurately model the shoulder surfing threat and systematically defeat the defense mechanism of our proposed system.

The system is developed as a standalone Java GUI software application and is used for both the technology and user studies. The test hardware includes a standard Dell Dimension desktop PC, a 1280x1024 resolution 19” LCD monitor, and a standard full size keyboard.
6.1 Technology Study

6.1.1 Objective

The technology study aims to accurately recreate a camera based shoulder surfing attack and to identify the threshold points at which the defense mechanisms are no longer reliable. Instead of using untrained user subjects as model attackers, we launch our own attacks to break the defense mechanisms of the system. By acknowledging upfront that the system will fail at some threshold point, we utilize our intimate knowledge of the system design and a methodical approach to empirically find the conditions under which the system succeeds and fails. Hardware and software resources are utilized to mount the attack and correctly model the real world threat.

6.1.2 Format

To reduce the complexity of the technology study and improve the strength of the security claims of the system, it is assumed that all keystrokes entered during the login session are captured accurately by either a keylogging mechanism or a secondary video camera. Thus, when attempting to subvert the defense mechanisms of the system during the study, the camera based attack can focus solely on capturing information from the screen. Furthermore, a human based shoulder surfing attack, i.e. an attack that relies only on human cognition and does not use cameras, is not modeled in this study: a camera based attack with offline analysis is inclusive of the attack vector posed by a human and it has been previously analyzed in [42, 43].

While the same software GUI is used for the user and technology studies, the technology study does not require a model user to be logging in during the attack. The camera attempts to grab any information from the screen and it is assumed that the user’s keystroke has been captured. Any single challenge screen with a grid of arrows that is successfully captured can be considered to eliminate half of the possible image locations.

To identify the resource threshold required to break the contrast inversion defense mechanism of the system, a variety of video recording devices are used. To model a cell phone attack, two cell phones are tested: a Sony Ericsson Z525a cell phone with a 640x480 pixel camera capable of 176x144 QCIF video
resolution at 10fps [55], and a Sony Ericsson W580i cell phone with a 1600x1200 pixel camera capable of 176x144 QCIF video resolution at 15fps [56]. To model a handheld camera attack, a Canon Powershot A610 camera [52] capable of 640x480 VGA video resolution at 30fps and a ‘prosumer’ Kodak Z812IS [57] camera capable of 1280x720 HD video resolution at 30fps are tested. Finally, a Sony HDR-FX1 HDV [58] Camcorder capable of 1440x1080 HD video resolution at an interlaced 60fps is tested to model the highest hardware resource investment available for an attacker. In addition to the recording devices, a tripod is used to stabilize the recording when applicable.

Figure 19: Sony Z525a, Sony W580i, Canon A610, Kodak Z812IS, Sony HDR-FX1

For each video recording device, the speed of the contrast inversion, i.e. the length of time the white screen with the grid of arrows is displayed, is varied to determine at what point the system is compromised. Both the recording device’s accessible video recording settings, e.g. zoom, exposure compensation, etc., and physical location are varied to optimize each test. To complete the attack model, frame-by-frame analysis and post-processing, e.g. contrast and brightness filters, of the video is completed in Virtual Dub [54] to determine if any information can be extracted from the recording. When the system is compromised, the number of challenges and pass icons extracted are noted to determine the severity of the breach.
6.2 User Study

6.2.1 Objective

The user study aims to determine how the cognitive trapdoor game and the contrast inversion defense influence login success rates, error types, and completion time. The study establishes a lower bound for optimal contrast inversion delay settings and records the typical duration for a successful login.

6.2.2 Format

Test subjects are given a brief written introduction to the system and are then presented with the four images of their password portfolio. A simple demonstration phase is conducted to allow users to acquaint themselves with the system and complete a successful login. The user study consists of eight subjects and varies the contrast inversion delay to determine the fastest speed that still allows for high login success rates. Three different contrast inversion speeds are tested, 100ms, 200ms, and 300ms, with two trials for each test. The software will track login success rate, error types, and completion time for successful logins for each user test.
7. Chapter 7 - Results

7.1 Technology Study

7.1.1 Mobile Phone: Sony Ericsson Z525a

The Sony Ericsson Z525a represents an entry level mobile phone for our tests. To set up a worst-case scenario for stronger conclusions about its recording capabilities, the phone was held approximately one foot away from the monitor so that the login application filled the entire video capture view. To help the phone optics adjust faster to the contrast inversion, the brightness compensation setting on the phone was set to the lowest possible value, -2, and the ‘Night Mode’ setting was switched off. The ‘Night Mode’ setting increases light sensitivity and would make capturing information during the contrast inversion almost impossible. In addition, the ‘High Quality Video’ 176x144 resolution was selected to limit the information loss due to high codec compression.
A post-processing filter reduced the brightness and increased the contrast of the recorded video frames to extract any information hidden from the naked eye.

The results show that up until 1000–1250ms, the contrast inversion defense is able to adequately disrupt the video recording capability of the Sony Z525a. At the 1250ms setting, the phone begins to extract a significant percentage of information from the screen until the 1750ms setting, at which point the entire password is extracted in one recording session. The results indicate a contrast inversion setting of 1000ms would reliably defend against this class of recording device.

During testing we observed that the phone’s recording capability was limited by its slow response in changing camera optics to a scene change, lack of optical zoom, low 10fps video frame rate, and low 176x144 video resolution. To be able to discern information during video playback, the phone had to be placed about one foot away from the login screen because of its low recording resolution.
7.1.2 Mobile Phone: Sony Ericsson W580i

The Sony Ericsson W580i represents a more recent mobile phone for our tests. To set up a worst-case scenario for stronger conclusions about its recording capabilities, the phone was held approximately one foot away from the monitor so that the login application filled the entire video capture view. To help the phone optics adjust faster to the contrast inversion, the brightness compensation setting on the phone was set to the lowest possible value, -2, and the ‘Night Mode’ setting was switched off. The ‘Night Mode’ setting increases light sensitivity and would make capturing information during the contrast inversion almost impossible. In addition, the ‘High Quality Video’ 176x144 resolution was selected to limit the information loss due to high codec compression.

A post-processing filter applied a grayscale format, reduced the brightness, and increased the contrast of the recorded video frames to extract any information hidden from the naked eye.

Figure 21: Determining the contrast inversion delay threshold for the Sony W580i (‘*’ indicates that a post-processing filter was required to obtain results).
The results show that up until 350-400ms, the contrast inversion defense is able to adequately disrupt the video recording capability of the Sony W580i. At the 450ms setting, the phone begins to extract a significant percentage of information from the screen until the 500ms setting, at which point the entire password is extracted in one recording session. The results indicate a contrast inversion setting of 350ms would reliably defend against this class of recording device.

During testing we observed that the phone’s recording capability was limited by its slow response in changing camera optics to a scene change, lack of optical zoom, and low 176x144 video resolution. To be able to discern information during video playback, the phone had to be placed about one foot away from the login screen because of its low recording resolution. In addition, the codec used to compress the video stream lost enough detail that even post-processing on frames did not uncover information: several video frames show that the phone was able to change its optics adequately, but because the frame was ‘garbled’ arrow directions could not be discerned. In comparison to the Sony Ericsson Z525a, the W580i did show a vast improvement in camera optics, hardware, and software processing in that it was able to begin compromising the system about three times faster than the Z525a according to the contrast inversion delay metric.

7.1.3 Point and Shoot Camera: Canon A610

The Canon A610 represents an entry level point and shoot camera for our tests. To set up a worst-case scenario for stronger conclusions about its recording capabilities, the camera was placed on a tripod approximately three feet away from the monitor so that, in combination with the available 4x optical zoom, the login application filled the entire video capture view. To help the camera optics adjust faster to the contrast inversion, a custom white balance was preset to a white screen of the LCD display. In addition, the highest quality video recording format, 640x480 video resolution at 30fps, was selected.
Figure 22: Determining the contrast inversion delay threshold for the Canon A610 (* indicates that a post-processing filter was required to obtain results).

A post-processing filter applied a grayscale format, inverted the color scheme, increased the brightness, and increased the contrast of the recorded video frames to extract any information hidden from the naked eye.
Figure 23: Post-processing the original Canon A610 video frame (left) confirms that the arrow grid has been compromised (right).

The results show that up until 500-625ms, the contrast inversion defense is able to adequately disrupt the video recording capability of the Canon A610. At the 625ms setting, the camera begins to extract a significant percentage of information from the screen until the 750ms setting, at which point the entire password is extracted in one recording session. The results indicate a contrast inversion setting of 500ms would reliably defend against this class of recording device.

During testing we observed that the camera’s recording capability was limited by its slow response in changing camera optics to a scene change. The recording resolution and optical zoom did not play a major limiting role, but increased resolution and a more powerful optical zoom would allow for capture attacks from a greater distance. The video had low levels of distortion, so the post-processing filters were only used to confirm what the naked eye already had seen.

7.1.4 ‘Prosumer’ Point and Shoot Camera: Kodak Z812IS

The Kodak Z812IS represents an advanced point and shoot camera for our tests. To set up a worst-case scenario for stronger conclusions about its recording capabilities, the camera was placed on a tripod approximately three feet away from the monitor so that, in combination with the available 12x optical zoom, the login application filled the entire video capture view. To help the camera optics adjust faster to the contrast inversion, the ‘Single AF’ setting was selected so that the camera’s focus was locked on the
login screen for the entire recording session. In addition, the highest quality video recording format, 1280x720 video resolution at 30fps, was selected.

![Kodak Z812IS](image)

Figure 24: Determining the contrast inversion delay threshold for the Kodak Z812IS (‘*’ indicates that a post-processing filter was required to obtain results).

A post-processing filter decreased the brightness and increased the contrast of the recorded video frames to extract any information hidden from the naked eye.
The results show that up until 300ms, the contrast inversion defense is able to adequately disrupt the video recording capability of the Kodak Z812IS. At the 350ms setting, the camera begins to extract a significant percentage of information from the screen until the 400ms setting, at which point the entire password is extracted in one recording session. The results indicate a contrast inversion setting of 300ms would reliably defend against this class of recording device.

During testing we observed that the camera’s recording capability was limited by its slow response in changing camera optics to a scene change. The high recording resolution, albeit with grainy video quality, and optical zoom should allow for capture attacks from a greater distance. The camera’s onboard image processing chip was overzealous in removing noise from the video stream, but arrow direction information was still recoverable with a simple post-processing filter. Compared to the older model Canon A610, the Kodak Z812IS offers a faster reaction time to scene changes and allows for attacks to be carried from more remote locations due to its higher video resolution and powerful optical zoom.

7.1.5 HD Video Camcorder: Sony HDR-FX1

The Sony HDR-FX1 represents the most advanced video recording camera and the highest resource investment for an attacker in our tests. To set up a worst-case scenario for stronger conclusions about its recording capabilities, the camcorder was placed on a tripod approximately 10 feet away from the
monitor so that, in combination with the available 12x optical zoom, the login application filled the entire video capture view. To help the camcorder optics adjust faster to the contrast inversion, a custom white balance was set to a white screen displayed on the monitor and the neutral density (ND) filter was set to the most aggressive setting. The ND filter reduces light intensity over all colors equally and was well suited for compromising the contrast inversion mechanism in these tests. In addition, the highest quality video recording format, 1440x1080 video resolution at 60 interlaced frames per second, was selected.

![Sony HDR-FX1](image)

Figure 26: Determining the contrast inversion delay threshold for the Sony HDR-FX1 (‘*’ indicates that a post-processing filter was required to obtain results).

A post-processing filter decreased the brightness and increased the contrast of the recorded video frames to extract any information hidden from the naked eye.

The results show that at any tested delay setting, the contrast inversion defense is unable to adequately disrupt the video recording capability of the Sony HDR-FX1. There is no usable contrast delay setting that would reliably defend against this class of device.
During testing we observed that the camcorder’s ND filter significantly reduced the effectiveness of the contrast inversion. It was so powerful that the camcorder’s advanced settings, such as a lockable aperture and shutter speed, did not need to be employed: the camcorder’s automatic exposure settings were more than adequate to compromise the system when paired with the ND filter. While the tests show the camcorder’s effectiveness degrading around 50ms, we were unable to test lower contrast inversion settings because redrawing the screen in Java became unstable: our Java application did not predictably compute and redraw complex panels in such short periods of time. The camcorder’s 60 interlaced frames per second video format began to show its inability to capture screen information around 50ms, as occasional frames displaying the arrow grid would be missed. The camcorder’s high resolution and powerful optical zoom allow for attacks to be mounted at greater distances from the login terminal.

7.2 User Study

Eight undergraduate university students were recruited to participate in the user study. There were three female subjects and five male subjects. Four of the subjects claimed to use some form of vision correction, either glasses or contacts. Subjects completed trials at 100ms, 200ms, and 300ms contrast inversion delay settings with short breaks in between each setting. The easiest setting, 300ms, was completed last so that fatigue would not adversely affect the user’s ability to login successfully. Sixteen total trials were completed across all user tests for each delay setting.
Overall Login Success Rates

Figure 27: Login success rates at various contrast inversion delay settings.

Results show that users were able complete a successful login at high rates for the 200ms and 300ms delay settings. The 100ms delay setting experienced the most failed login attempts, but still exhibited high overall login success rates. All five errors recorded during the trials were due to a user hitting the incorrect arrow key in response to a challenge.
Figure 28: Time elapsed for a successful login at the 30ms contrast inversion setting.

For successful logins at the 300ms delay setting, users averaged 50.4 seconds to login and the average of each user’s best login time was 49.53 seconds. While the 300ms delay setting is the slowest of all login settings, it is chosen as the best candidate for measuring login times: by the last set of tests the user’s comfort levels with the system had increased and variations in their login behaviors had minimized.
8. Chapter 8 - Analysis

The technology study was designed to provide the most optimal conditions for the various recording devices so that we can make strong conclusions about the defense capabilities of the system. To compensate for the low video resolutions of the mobile phone devices, the tests placed the phones within one foot of the login screen. The point and shoot cameras were also placed within three feet of the screen and mounted on a tripod to eliminate any camera shake that could affect camera optics. The most aggressive neutral density filter was used on the HD camcorder to push the system to its limits. Aggregating the results from these tests aids in determining what level of reliable defense the system provides for different application contexts.

Figure 29: Aggregate graph of percentage of challenge screens compromised during a login session.
When interpreting the results, we consider the system to be compromised when a significant percentage, e.g. 30% or greater, of challenge screens are captured: with this much information, an attacker can make an educated guess or record multiple recording sessions to deduce a user’s password.

While all devices exhibited some amount of reaction delay to the contrast inversion, the mobile phone class of devices also suffered from low video capture resolution and the absence of a high quality optical zoom. The dated Sony Ericsson Z525a reacted extremely slowly to the contrast inversion, requiring over one second to recognize a contrast change and translate it into adjusting the camera’s optics. Surprisingly, the newer Sony Ericsson W580i camera phone exhibited better responsiveness to the contrast inversion than the entry level Canon A610 camera, capturing a significant percentage of challenge screens at the 400ms delay setting. A delay setting of 300-400ms would reliably defend against both of these
mobile phones. The improvement in technology between the release dates of these two camera phones is clearly evident in the performance gap exhibited in the results. While continued technology advancements will improve both its video capture resolution and processing power, the camera phone class still suffers from weak optical zoom technology. Improvements will be made in this aspect as well, but currently the appeal of camera phones for an attacker is their small footprint and benign demeanor when capturing video. If the phone must be placed within five feet of the login terminal to successfully capture video, then, at the very least, our system has reduced the likelihood that a camera phone shoulder surfing attack would go unnoticed by a victim.

Both handheld cameras in our tests were limited mostly by their inability to react to the contrast inversion. The recording resolution and frame rates seemed adequate, especially when paired with a more powerful optical zoom. Once again the technology improvement between the Canon A610 and the Kodak Z812IS is reflected in the performance difference in the results. However, a delay setting of 300ms or less would still reliably defeat both handheld cameras. These cameras pose less of a threat than cell phones at login terminals due to their increased size and obvious malevolent nature when spotted anywhere near a terminal screen.

The Sony HDR-FX1 HD camcorder results do show a limit to the effectiveness of the contrast inversion defense. By utilizing the neutral density filter, the camcorder was able to compromise the system for all testable delay settings, even when using the automatic exposure mode. The benefits of specialized hardware and large, high quality, optics are apparent in these tests. However, some comfort can be taken in knowing that an attacker using such a device would have to invest thousands of dollars and be willing to risk being exposed due to its highly intrusive nature.

Surprisingly, ambient light had little effect on the performance of the cameras. Testing was completed both with daylight and indoor light and showed no discernable differences. However, differences in LCD displays can influence results. There was a noticeable difference in clarity and responsiveness when using a newer 19” LCD display when compared to an older model, due to improvements in the response time, the panel brightness, and the contrast ratio of the LCD panel. Another possible parameter of the system that was not tested is the design, size, and darkness of the arrow images.
Perhaps advanced manipulation of the arrow design could take advantage of compression imperfections or fool camera optics while still allowing for high login success rates.

The results of the user study match up well with the results of the technology study. Users logged in with high success rates, 93.75%, for both the 200ms and 300ms delay settings. Looking at the aggregate statistics for the technology study, a delay setting at or below 300ms would mitigate the threats posed by all the tested recording devices, except for the Sony HDR-FX1 HD camcorder. The failed login attempts at the 100ms delay setting were due to users anticipating and reacting too quickly to the arrow grid, which resulted in them accidentally hitting the wrong arrow key. When the speed of the contrast inversion was changed to 200ms and 300ms, users felt more relaxed and exhibited greater patience in responding with the arrow key. Fatigue did play a role in the user study, as users’ concentration seemed to wander by the last set of 300ms trials. However, the slower contrast inversion was easier on the users’ eyes and fatigue did not produce a noticeable performance hit. During exit interviews, some users expressed concern regarding the longer login times, but nearly all recognized the shoulder surfing threat and were willing to add some amount of time to the login session given the context of the application. The results show that the current format takes anywhere between 45 and 55 seconds to complete a four round image challenge, but the design of the system can be modified to reduce login times and accommodate a variety of applications.

In analyzing the results of the technology and user studies, we wish to revisit the design goals for the system. In order to successfully login to the system, the user must follow the procedural steps and maintain a high level of concentration: self-selected passwords are not allowed and the format of the system does not lend itself to be easily subverted. However, because the system is a recognition based graphical password system and has the same memory requirements as a system like Passfaces, we can expect similar levels of high recall rates. The system also models the four round, 3x3 image grid, format of Passfaces, meeting existing PIN metrics against brute force attacks. The system only uses a keyboard and a color display to obfuscate the output and input of information, ensuring low deployment costs as no specialized hardware is required to defend against camera based shoulder surfing attacks. Most importantly, the results show that the system increases the resource investment required for a successful shoulder surfing attack to unfavorable levels, requiring a large and expensive camera to defeat the system. This is
accomplished without ceding the benefits of graphical password systems, including improved recall rates, resistance to information harvesting, and resistance to social engineering attacks.
9. **Chapter 9 - Conclusions**

With skyrocketing cases of fraud and identity theft, it has become apparent that traditional authentication mechanisms guarding sensitive information are under constant attack. The familiar text password is increasingly trivial to compromise due to its weaknesses in resisting numerous attack vectors. Graphical password systems have been shown to improve upon traditional passwords in mitigating such attacks while also improving recall rates and usability. However, the highly visual nature of these systems makes them prime targets for camera based shoulder surfing attacks. The increasing number of small mobile phones equipped with camera recording capabilities reduces the invasiveness of a shoulder surfing attack.

In this thesis, we first identify the shortcomings of the traditional text password and survey a variety of graphical password systems. We then consider how these systems address the shoulder surfing attack vector and explore various defenses aimed to mitigate this type of attack. After clearly describing the threat, the design goals for a new defense mechanism, and the metrics most applicable to evaluate the new mechanism, we introduce our contrast inversion defense that is designed to significantly increase the resource investment of a shoulder surfing attacker. Our initial project exposes the shortcomings of a Passfaces graphical password system under a camera based shoulder surfing attack and establishes a proof of concept to defend against recording devices. By combining this defense mechanism with an input obfuscation game, we are able to effectively frustrate most entry level consumer recording devices.

To test our system, we conduct both a technology study and a user study. By using a variety of recording devices in an optimal shoulder surfing environment, our comprehensive and systematic testing method allows us to clearly reveal the defense capabilities of the proposed system. When combined with the results of the user study, it is evident that the contrast inversion defense is highly successful at protecting the login session from most entry level recording devices while still allowing for high user login success rates. These results allow us to claim that our system has raised the resource threshold for a successful shoulder surfing attack beyond a level that most attackers are willing to meet.
9.1 Alternate Designs and Future Work

While the proposed system has shown to both defend against various camera recording devices and maintain high login success rates for users, it is still within the context of a large tradeoff space. The new defenses introduced increased system security, but also increased the time required to complete a successful login. Alternate designs may choose to increase the number of icons presented in the icon grid and the associated number of challenges per round, while reducing the total number of rounds during the login process. For example, with four challenge screens per round, 16 different locations can be disambiguated and with five challenge screens per round, 32 locations can be disambiguated. For specific applications, two rounds with a 4x4 grid may prove to provide adequate security with short login times. Other applications, such as single sign-on environments, may be willing to handle longer login times because a single login session can authenticate a user for several accounts. Also, preset timing and exposure lock attacks can be mitigated by separating challenge screens by a random amount of time and alternating the contrast inversion: the transition can be from an empty black grid to a white arrow grid or from an empty white grid to a black arrow grid. It may also be the case that the contrast inversion defense is better suited to be adapted to a different scheme entirely.

Another option would be to eliminate the challenge scheme. Instead, during the contrast inversion, a number can be flashed in each of the grid locations and be entered with the number pad. This would greatly reduce the login time, but a single captured video frame would compromise the pass icon. In another scheme, a moving cursor could be displayed over various grid positions during the contrast inversion and the user would time their response to coincide with the cursor passing over their pass icon location.

Future work that utilizes the contrast inversion defense may conduct a larger user study that encompasses multiple age groups. Research is also required to evaluate how subjects who may be prone to seizures due to photosensitive epilepsy can be affected by the contrast inversion defense. Additional endeavors in defending against shoulder surfing may investigate other areas in which humans possess an innate physical or cognitive ability that automated devices lack or are not as proficient at. While technology trends indicate that cameras will continue to increase resolution and processing power, it should be noted
that even the most advanced HD video camcorder tested in the technology study defeated the system not because of its high resolution and image processing, but because of its effective neutral density filter. The availability of, or the feasibility of retrofitting, neutral density filters for small mobile phones is another opportunity for future research. However, mobile phones still have several years before they can pack the image processing power of larger SLR and HD video cameras. Furthermore, several hardware components, such as advanced lenses and filters, for high end camera devices have not miniaturized in size over the years. Our system is defending against a moving target, a future compact mobile phone that can combine HD video resolution with an advanced optical lens, but we believe our system can provide adequate defenses against this target for the next several years.
Bibliography


[53] I. Canon U.S.A. PowerShot SD400.  

[54] Virtual Dub. [http://www.virtualdub.org/].


[56] Sony Ericsson. Specifications - W580i.  

[57] Eastman Kodak Company. EASYSHARE Z812 IS zoom digital camera.  

[58] Sony Electronics U.S.A. HDR-FX1 HDV handycam® camcorder.  