Human Computation with Perceptive Intelligence

THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE (BY RESEARCH)

IN

COMPUTER SCIENCE

BY

KHOT ROHIT ASHOK

200606006

ROHIT_A@RESEARCH.IIIT.AC.IN

Center for Security, Theory & Algorithmic Research (C-STAR),
International Institute of Information Technology,
Hyderabad, India 500032.

© Khot Rohit Ashok 2009.

International Institute of Information Technology,
Hyderabad, INDIA

CERTIFICATE

This is to certify that the thesis entitled "*Human Computation with Perceptive Intelligence*" submitted by *Khot Rohit Ashok* to the International Institute of Information Technology, Hyderabad, for the award of the *Degree of Master of Science (by Research)* is a record of bona-fide research work carried out by him under my supervision and guidance. The contents of this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Date: Advisor: Dr. Kannan Srinathan

|| ज्ञानेश्वर कृत पसायदान ||

आता विश्वात्मकें देवें । येणे वाग्यज्ञें तोषावें ।
तोषोनिं मज ज्ञावे । पसायदान हें ॥

जें खळांची व्यंकटी सांडो । तया सत्कर्मी- रती वाढो ।
भूतां परस्परे पडो । मैत्र जीवाचें ॥

दुरितांचे तिमिर जावो । विश्व स्वधर्म सूर्यें पाहो ।
जो जे वांच्छिल तो तें लाहो । प्राणिजात ॥

वर्षत सकळ मंगळी । ईश्वरनिष्ठांची मांदियाळी ।
अनवरत भूमंडळी । भेटतु भूतां ॥

चलां कल्पतरूंचे आरव । चेतना चिंतामणींचें गाव ।
बोलते जे अर्णव । पीयूषाचे ॥

चंद्र्मे जे अलांछन । मार्तंड जे तापहीन ।
ते सर्वांही सदा सज्जन । सोयरे होतु ॥

किंबहुना सर्व सुखी । पूर्ण होऊनि तिन्हीं लोकी ।
भजिजो आदिपुरुखी । अखंडित ॥

आणि ग्रंथोपजीविये । विशेषीं लोकीं इयें ।
दृष्टादृष्ट विजयें । होआवे जी ।

येथ म्हणे श्री विश्वेशराओ । हा होईल दान पसावो ।
येणें वरें ज्ञानदेवो । सुखिया जाला ॥

ज्ञानेश्वरी अध्याय १८ ओवी क्रमांक १७९३ ते १८०१

“He is quick thinking in clear images;
I am slow, thinking in broken images.
He becomes dull, trusting to his clear images;
I become sharp, mistrusting my broken images.
Trusting his images, he assumes their relevance;
Mistrusting my images, I question their relevance.
Assuming their relevance, he assumes the fact;
Questioning their relevance, I question their fact.
When the fact fails him, he questions his senses;
When the fact fails me, I approve my senses.
He continues quick and dull in his clear images;
I continue slow and sharp in my broken images.
He in a new confusion of his understanding;
I in a new understanding of my confusion.”

Robert Ranke Graves

English poet, translator and novelist

Abstract

Human visual system is a pattern seeker of enormous power and subtlety. We not only can see things clearly, but are also capable of describing them with precision and remembering them for a long time. Having these capabilities had a major impact on our sustenance, survival and perpetuation as species. Although computers can perform a variety of tasks that are beyond human capability because of speed, complexity, or dangerous environments, attempts to replicate human perceptual abilities have been strikingly inferior, even for the visual tasks that people consider extremely simple.

In this thesis, we advance the research in the field of human computation by leveraging human perceptual abilities to solve problems that computers alone cannot effectively solve. In particular, we address two important problems: user authentication, and image annotation.

User authentication has issues in both security and usability. For example, passwords are either '*secure but difficult to remember*' or '*memorable but not secure*', when by definition, they needs be both secure and memorable. Graphical passwords are viable alternative to text passwords since they are based on proven human ability to recognize and remember images, coupled with the larger password space offered by images. In this thesis, we propose and evaluate, *Marasim*, a novel Jigsaw based graphical authentication scheme, using Tagging. Marasim is aimed at achieving the security of system chosen images with the memorability of self chosen images. Empirical studies of Marasim provide evidence of increased memorability, usability and security.

Additionally, we examine the manual image annotation problem. Recently there have been a number of attempts to lure humans into annotation process. Notable examples are interactive games like ESP, and social tagging like Flickr. However, we found that extant methods in their present form are inadequate to result in annotations of high quality. We therefore, introduce two intelligent system designs for semantic annotation of images in the form of a game and a CAPTCHA. First one is GoFish, a web variant of standard Go-Fish, a popular playing card game. While the other one is image recognition CAPTCHA, named iCAPTCHA. Behind both these designs is a strong emergent semantic theory that ensures superior annotations.

Keywords: Human Computation, Perceptive Intelligence, Human in the loop system, Usable security, User Authentication, Graphical passwords, Tagging, Games with a Purpose, CAPTCHA.

Preamble

Foreword

As with so many printed works in the early twenty-first century, some aspects of this thesis are likely to seem passé or quaint by the time it is printed. While any work is the product of the thoughts, efforts and tolerance of many folks, we take all blame for running rough-shod over concepts and insights of others and for any missteps and misunderstandings. We only hope that the work as a whole will provoke thinking in ways that advance the discourse in the arena.

…

Acknowledgments

"At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us!!"

First and foremost, I am grateful to my supervisor *Kannan Srinathan*. It is due to his excellent guidance and constant backing that this research has been possible. His expertise and unconditional love for science and philosophy has always inspired me and will continue to inspire many…

I Specially thank *Kavita Vemuri* mam, with whom I have worked closely on many projects over the last two years. Her contribution and support have been invaluable in getting the dissertation completed.

Thanks to the collegues in the *C-STAR*, and *Cognitive Science* lab who have helped with the experiments, listened to the presentations, and offered valuable insight and feedback throughout the process. Special thanks to all my professors: *Venkaiyah sir, Bruhadeshwar sir, Kishore sir, Madhu sir, P R K Rao sir* and *Anirudha Joshi sir* for their guidance.

Thanks to the members of my commitee, all ananyomous reviewers of my papers for their guidance, expertise and for offering different perspectives, all of which have helped shape this disseration.

To *Amol* and *Abhijit* and *Viraj*, I offer many thanks and my appreciations for the emotional support and understanding throughout the years.

Thanks to all the participants who took part in my user studies. Their cooperation and feedback were key to this research.

Thanks to *DC++*, *IMDB*, *CRICINFO* and my motivation for life, *Sourav Ganguly*.

Thank you to all my friends: (in alphabetical order)

Aananda, Abhijit, Amit, Amol, Anuj, Asmita, Bhavani, Chandan, Eshan, Kalyan, Kaushik, Manish, Nayan, Neeraj, Pankaj, Piyush, Poornima, Pranav, Prasanth^{2}, *Rahul*^{2}, *Raju, Ravikant, Ritesh, Romanch, Ruhi, Rupesh, Rutuja, Sandeep*^{2}, *Sai, Saras, Sarat, Siddarth, Srirang, Sudheer, Uma, Vinayak, Viraj...*

Sorry if I missed out on few names...

Lastly, my family and friends have been so incredibily understanding throughout this journey. There have been many missed special occasions, stressed phone calls, and rushed holidays over the course of the degree. Their unwavering support and confidence means a lot to me. To my *Mom* and *Dad*, you are the best.

Notations and Conventions

- **Bold** and *italic* are used for emphasis and to signify the first use of a term.
- The present report is divided in *chapters*. Chapters are broken down into *sections*.
- Where necessary, sections are further broken down into *subsections* and subsections may contain some *paragraphs*.
- *Author references* [Mels04] as well as *web references* [*Soft08*] (note the *italic* style) are tagged inside square brackets.
- As a respect to both genders, *he* and *she* are used interchangeably in this document.

Table of Contents

List of Figures

List of Tables

1
Introduction

"This triumph of human ingenuity is the most audacious, remote, improbable, incredible, — the one that would seem least likely to be regained, if all traces of it were lost, — of all the discoveries man has made".

Oliver Wendell Holmes

American physician, professor, lecturer, and author, 1809 -1894

1.1 A conversation in space

Somewhere in space, two aliens (a master named *Alpha* and his subordinate *Beta*) are talking to each other on a highly classified mission. The recorded conversion is translated in English.

> **Alpha:** So boys, is everything ready?
>
> **Beta:** Pardon me master, there is a small problem.
>
> **Alpha:** What is it now?
>
> **Beta:** Sir, our robots are failing to recognize and remember the destined targets. We tried all advanced technologies that we have, unfortunately no success as yet.

Alpha: (*with anger and disgrace*) What? We are the most advanced and developed planet in the whole universe, and still we can't solve such a simple problem! Shame on us! Go hunt other planets and look whether they have a solution for it. This time, I don't want any excuses, come back only when you have something to show…

Go now…

(*Few days later…*)

Beta: (*holding his breath*) Master, master….

Alpha: Calm down. What happened?

Beta: Master, we just found one planet having the technology we desire.

Alpha: Fantastic! Tell me more.

Beta: The planet is full of supercomputers with supreme capabilities. These supercomputers are well equipped with five extremely fast and accurate input channels, huge memory storage and a very efficient processor working at the speed of thought. Moreover, for years, they have been recognizing and remembering different targets in front of them for variety of reasons. As a result, today they have mastered the art. These are some snaps we have taken of them in action. Master, please have a look.

Alpha: (*looking at pictures*) Very good Beta, but have you tried to learn and understand their technology? See, how easily they are doing it with so much precision!

Beta: Sorry sir, we tried hard, but no luck. Their encoding is so complex that, as of today, no one is able to decode it. However, the good news is, we now know how to convince them to work with us.

Alpha: Great then, what are we waiting for? Let us join hands and carry out the mission…

Most of you must have realized it by now; the aliens were talking about none other than the *planet Earth*, and collaborating with the *natural born supercomputers* on this planet, humans.

Although, computers can perform a variety of tasks that are beyond human capability because of speed, complexity, or dangerous environments, some of the simplest patterns immediately recognizable to the human eye are still elusive to machines (like the alien robots failing to locate the targets in the above fictional conversation). Human visual system on the other hand, is a pattern seeker of enormous power and subtlety. Our eyes collect and store images for use in meeting with our psychological needs. We not only can see things clearly, but are also capable of describing them with precision and remembering them for a long time. Having these capabilities had a major impact on our sustenance, survival and perpetuation as species [Ocon07].

In this thesis, we advance the research in the field of human computation by leveraging human perceptual abilities to solve following two important problems: user authentication, and image annotation.

1.1.1 User authentication

User authentication is a problem for every system providing secure access to valuable and confidential information or personalized services. Since user authentication involves users, it has issues in both security and usability. For example, most systems use passwords to authenticate users, but passwords that are easy to remember such as '*iiit123*' are also easy to guess by dictionary searches, while secure passwords with random characters something like '*ad45$%w*', end up in people either writing them down or forgetting them. Either outcome defeats the purpose of passwords, which is to be secure and memorable at the same time. Ironically, attackers are experts in usability; they exploit these password malpractices, by developing simple yet effective social engineering attacks to steal identity information. An authentication system must therefore, balance the need to remember the password with the necessity of making password as random as possible.

Graphical passwords are viable alternative to text passwords since they are based on proven human ability to recognize and remember previously seen images, coupled with the larger password space offered by images [Cran05].

In this work, we propose and evaluate, Marasim, a novel Jigsaw based graphical authentication scheme, using Tagging. Marasim is aimed at achieving the security of system chosen images with the memorability of self chosen images. Empirical studies of Marasim provide evidence of increased memorability, usability and security.

1.1.2 Image annotation

Modern image search engines such as Google [*Goog09*], Bing [*Bing09*] collect and index images from other sites to provide access to the wide range of images. However, they often struggle to find the right image for a specific need from the large database of images and to reduce the clutter that often comes with the selection. An effective solution and a way to improve the accessibility of images, is by marking the image content with descriptive textual keywords known as tags. However, manual annotation is costly since humans find it tedious despite its benefits in terms of recall and retrieval.

Recently there have been number of attempts to lure humans into image annotations. Notable examples are interactive games like ESP [Vona04], and social tagging like

Flickr [*flic09*]. However, we found that extant methods in their present form are inadequate to result in annotations of high quality.

In this work, we present and evaluate two effective system designs for semantic annotation of images. Both these designs are based on strong emergent semantic theory that ensures annotations are of good quality.

1.2 Context: Human Computation

The area of '*Human Computation*' [Vona05] is an exciting new field of research, aimed at harnessing the combined computational powers of humans and computers to solve computationally hard problems. Most of us perceive that computers make people smarter. Human computation on the other hand, targets the '*wisdom of crowd*' to make computers smarter (as shown in Figure 1).

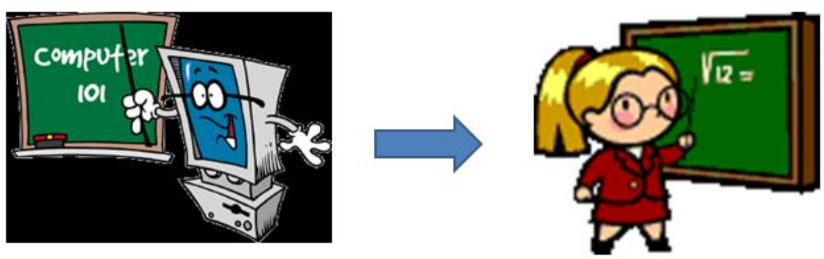


Figure 1: A shift from Traditional Computation to Human Computation

To illustrate, in traditional computation, a human employs a computer to solve a problem; a human provides a formalized problem description to a computer, and receives a solution to interpret. Human computation however, alters this perception by reversing the roles; the computer asks a person or a large group of people to solve a problem, then collects, interprets, and integrates their solutions. Human computation is an effective '*crowdsourcing*' [*Crow09*] technique where human brains are considered as processors in distributed systems each performing a small part of massive computation. However, unlike computers, humans need strong incentive in order to become part of collective computation.

A number of works have shown that by providing proper incentives, people can collectively solve large scale, open problems of computer science. The power of human computation was first demonstrated by simple yet effective game called ESP to label images on the web. The ESP game has been hugely popular with millions of users playing it for fun and collaboratively contributing to image annotation process. Amazon Mechanical Turk (AMT) [Amaz09] is another example that provides financial marketplace to coordinate developers and workers in solving human intelligence tasks. Some other notable examples include Wikipedia [*Wiki09*], SETI@home [*Seti09*] and Flickr [*Flic09*].

1.3 Motivation : The visual experience

If someone asks us a simple question, "*What will you do, if you win a million dollar lottery?*" Along with many other wonderful things, most of us would wish for a trip around the world. We all are fascinated with the thought of seeing the world with our own eyes. A natural question to ask here is "*What could be the reason? Why most of us would want to see the world?*", "*Why can't anyone want to hear all existential sounds, or taste all the edible dishes, or feel every surface?*" (Note that, all these wishes can in fact be fulfilled with the money.)

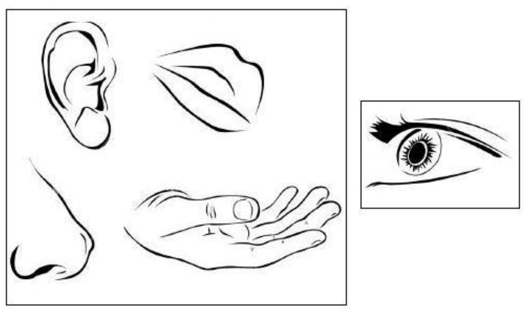


Figure 2: Dominance of Vision over other senses

It appears that vision is the prime input for generating the greatest understanding of the outside world while other senses (Touch, Smell, Hear and Taste) often carry out the supporting roles. The last of our senses to evolve and the most sophisticated, our eyes are truly wondrous windows on the world, sending more data more quickly to the nervous system that any other sense [Barr02]. As a result, we tend to believe more on things that we see (rather than on things that we hear or touch or smell or even taste) and enjoy things that challenge our visual belief (For example, magic shows, optical illusions, Escher's drawing). We thus, often say, "*Seeing is Believing*".

'*To see*' however, does not simply mean "*Having visible wavelength energy stimulating electrical activity in the eyeballs and brain.*" Vision as a process, is composed of *perception* (How we see) as well as *conception* (How we think). Our eyes are just the sensors (or more appropriately "goal oriented detectors" [Fisc87]); the main organ of the vision is the component that does the interpretation, our brain. Our brain combines the information from our eyes with the gathered data from our other senses, synthesizes it, and draws on our past experiences to give us a workable image of outside world. This image orients us, allows us to comprehend the situation, and helps us to recognize significant factors within it. Seeing enables understanding and resolution of problems.

Fischler and Firschein [Fisc87] quote:

> "*When a person says 'I see' after solving a difficult mathematical or conceptual problem, he is voicing a piece of wisdom that we are just beginning to appreciate, that his perceptual machinery... probably played a substantial role in producing the solution.*"

This human perceptual ability to recognize patterns and filter the irrelevant information has been hindering the progress of Artificial Intelligence (AI) for quite some time. How our mind utilizes the information from the visual system to image the external world and to create a meaningful experience is still a mystery. The only device to achieve any significant degree of AI is a digital computer. However, conventional digital computers are sequential symbol manipulators and are primarily suitable for tasks that can be broken down into series of simple steps. Thus, we can only expect them to realize the sequential paradigm of human intelligence (left hemisphere of our brain, Refer Figure 3).

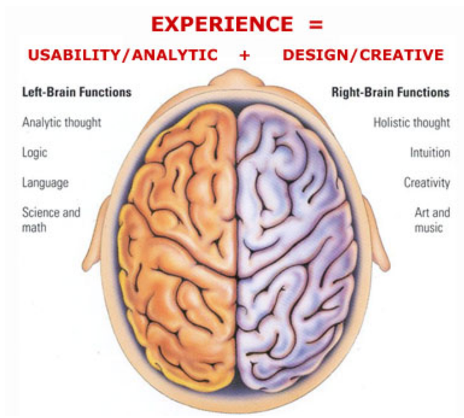


Figure 3: Left and Right Hemispheres of brain are specialized to deal with problem differently

However, many problems of perceptual nature do not allow decomposition and can only be solved by employing gestalt paradigm that can deal with the global information (Right hemisphere of our brain, See Figure 3) [Fisc87]. In effect, despite the impressive advancements in the computer technology, perceptual tasks that are trivial to humans such as image recognition and annotation still challenge most advanced computers. As of today, there is nothing that can replace the perceptual process on so grand and efficient scale.

Rather than devising software or algorithms that replicate human perception, we advocate a novel approach of utilizing human perception ability for solving various kinds of problems creatively and analytically. We call it '*Perceptive Intelligence*', since it originates in the perceptual process and is characteristic of abstract thinking.

(It is just a formal name given to a problem solving technique and we do not mean to relate it with multiple theory of intelligence, proposed by Howard Gardner [Gard93]).

In particular, we explore following two attributes of Perceptive Intelligence (These attributes are discussed in detail in the subsequent chapters: Chapter 2, 3, 7, 8).

- Pattern recognition and Image annotation
- Strong visual memory

1.4 Thesis Statement

The major goal of this thesis is to constructively channel human perceptual abilities to solve two fundamental problems: user authentication and image annotation.

User authentication

This work is aimed at designing an authentication scheme that is memorable, secure and usable. We focused on graphical passwords, because of the proven human ability to recognize and remember previously seen images. The main research question is,

> "*Can we design a graphical authentication scheme that supports both memorability and security, while maintaining usability?*"

The work began with the general investigation, with the new ideas being formed and tested as we progressed with the research. The three main research objectives are described below.

Objective 1: Catalogue existing graphical password schemes focusing on user choices of password images. Identify the key design features that offer maximum benefits in terms of security, memorability, and usability.

(It turned out to be, system chosen images are most secure while personal, self chosen images are easily remembered and recognition based graphical passwords are most usable.)

Objective 2: Propose or identify an authentication design strategy that incorporates the key security, memorability, and usability features found in objective 1.

(This goal ended up in creating a Jigsaw based authentication design that provides the security of system chosen images with the memorability of self-chosen images.)

Objective 3: Create and empirically evaluate an authentication mechanism based on the design strategy devised in objective 2.

(We proposed a working prototype of jigsaw based authentication design, Marasim, using tagging.)

Image annotation

The aim of this work is to study existing techniques of image annotation and to discover effective ways of improving them. The main research question is:

> "*Can we propose effective and enjoyable ways for semantic annotation of images?*"

Three research objectives for this work are described below.

Objective 4: Catalogue existing techniques of manual image annotations focusing on scalability, the quality of the annotation and the enjoyment people get while doing it. Identify problems with the manual annotation process and probe for the reasons.

(It turned out to be, although extant methods are quite successful in luring human to annotate images, the quality of the resultant annotation is still far from perfection. Our investigation found two fundamental problems with manual annotation process.)

Objective 5: Discover an effective approach to manual annotation or alter existing approaches to get quality labels or descriptions for images.

(This goal ended up in presenting an emergent semantics approach to tagging that can effectively solve the problems we found in objective 4.)

Objective 6: Create and test annotation systems based on the approach found in objective 2.

(This goal ended up in creating two intelligent system designs in the form of a game and a CAPTCHA.)

1.5 Overview of the Thesis

The remainder of the thesis is organized as follows (The roadmap of the thesis is shown in Figure 4).

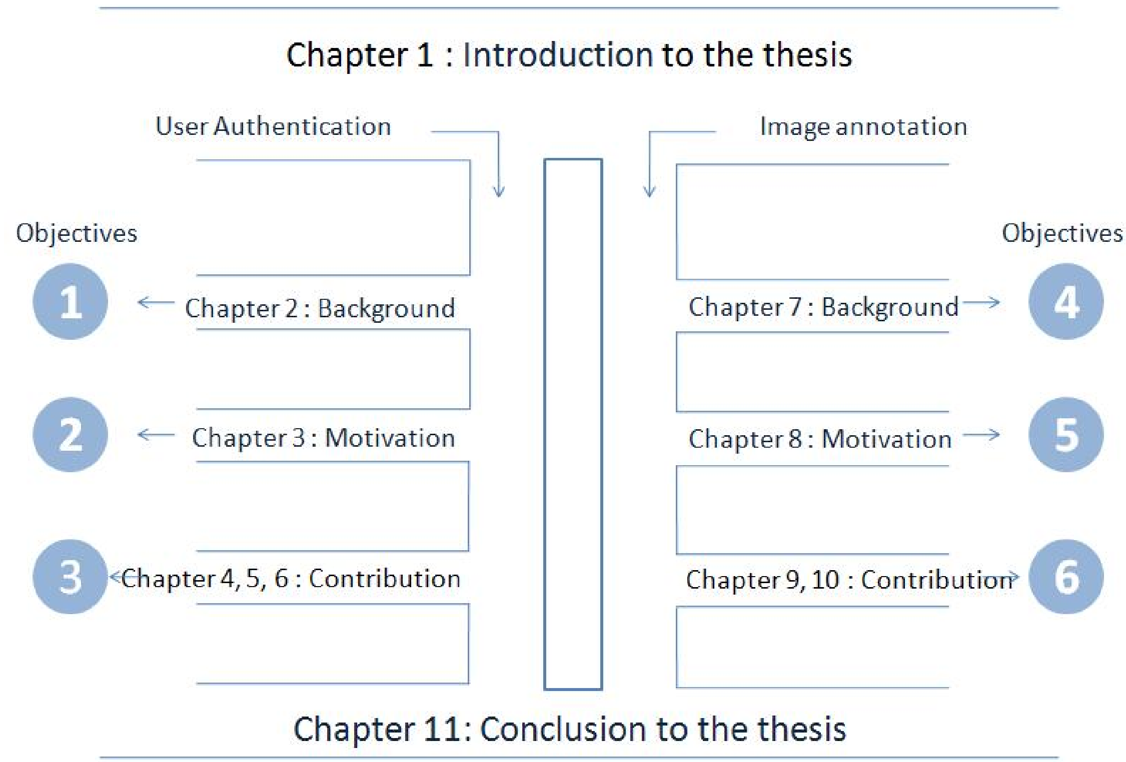


Figure 4: Roadmap of the thesis

This thesis is divided into two parts that corresponds to the two problems that we intend to solve. The first part talks about the user authentication problem and present our solution, Marasim, while the second part mainly discusses the image annotation problem and our proposed designs.

Chapter 2 addresses *Objective 1* of Section 1.4. First half of Chapter 2 presents relevant background on usable security and authentication, while the second half explores existing graphical password schemes by summarizing published analyses of these schemes. The chapter concludes with our rationale for Jigsaw based authentication design.

As a response to the *Objective 2*, Chapter 3 presents a novel authentication strategy in the form of a Jigsaw. Chapter describes our motivation behind such an approach and discusses its potential benefits in terms of security and usability.

Objective 3 required design work and then creation of novel scheme, as well as analysis to determine whether our design was effective. Chapters 4 to 6 contribute to meeting *Objective 3*. Chapter 4 introduces, our proposed graphical password scheme, Marasim. In Chapter 5, we present the user study of Marasim and results of our security and usability analysis. We conclude our discussion on Marasim with its potential benefits and commercial potential in Chapter 6.

The second part of thesis starts with Chapter 7, which targets the *Objective 4* of Section 1.4. This Chapter begins with the introduction to structuring an image collection, and its associated problems. We discuss how these problems can be solved with semantic annotation of images. Later in the Chapter, we review existing literature and methods of manual annotation and evaluate them with respect to the quality of the resultant annotation and the associated fun in doing. We conclude this chapter with our rationale for emergent semantics theory.

As a response to the *Objective 5*, First half of Chapter 8 presents problems associated with the human approach to annotation while the second half presents the solution to these problems in the form of emergent semantic theory. We also present an effective annotation design approach based the emergent semantic theory.

Objective 6 required design work and then creation of novel schemes, as well as analysis to determine whether our design was effective. Chapters 9 and 10 contribute to meeting *Objective 6*. Chapter 9 introduces an interesting multiplayer game GoFish. The lab study of GoFish and the results of our analysis are presented. We present in Chapter 10, an intelligent image based CAPTCHA design, and its corresponding user study.

Finally Chapter 11 discusses contributions of this research, present future directions and offer concluding remarks.

1.6 Main contributions of this research

This research contributes original ideas and knowledge to the field of human computation. We create and evaluate a novel graphical authentication scheme that offers improved security and memorability. We also present two interesting system designs in the form of a game and a CAPTCHA for quality image annotation.

The main contributions of this thesis are enumerated below.

- We reviewed existing graphical password schemes and found inherent weaknesses in the user choices of passwords images. As a solution to this password selection problem, we explored the feasibility of Jigsaw based authentication design. We illustrated how a jigsaw based approach can help to create a portfolio of secure yet memorable images to be used for authentication.
- We proposed and evaluated Marasim, a Jigsaw based authentication mechanism using Tagging. Marasim was aimed at achieving the security of system chosen images with the memorability of self chosen images. This scheme relies on human ability to remember a personal image and later recognize alternate visual representations (images) of the concepts that occurred in that image. We discussed its potential benefits in terms of security and usability. Our system makes a significant contribution in the way it leverages useful

characteristics of cued recall and challenge response schemes. Results of a user study proved the viability of our proposed design.

- We explored known HCI based techniques of manual annotation focusing on the quality of the annotation and the enjoyment people get while doing it. We found that extant methods in their present form appear inadequate to result in annotations of high quality. We therefore, propose an emergent semantic based approach to tagging.
- We presented GoFish, an intelligent system for semantic annotation of images from an online game. GoFish is a web variation of standard Go Fish, a popular playing card game. Behind GoFish game design is a strong emergent semantics theory that ensures superior annotations. We present the complete design of the game and discuss its benefits. Results of a preliminary user study are encouraging.
- We introduced iCAPTCHA, a user friendly and productive CAPTCHA design. Our premise is based on the human ability to recognize images and label them in proper categories. Each time a user solves an iCAPTCHA, he/she is helping to label images in proper categories which will in turn improve image search and retrieval.

1.7 Related Publications

Significant portions of the research presented in this thesis have been peer-reviewed and published in academic venues. I am primary author on the following papers based on work from this thesis. Much of the text in the thesis for these published portions is taken from the publications.

The publication list (Submitted or published):

- Rohit Khot and Kannan Srinathan. *iCAPTCHA: Image Tagging For Free,* In Proceedings of Usable Software and Interface Design (USID) 2009, Bangalore, India.
- Rohit Khot and Kannan Srinathan. *Gofish: Fishing Thousand Words Worth A Picture*. (Conference submission).
- Rohit Khot and Kannan Srinathan. *Marasim: A Jigsaw Based Graphical Authentication Scheme using Tagging*. (Conference submission).

Recognition and public demonstrations

- Early beta version of Marasim, titled 'NAPtune' was showcased in the 17th Annual *HySEA (Hyderabad Software Exporters Association)* event, on 7th March 2009.
- Marasim and GoFish were part of *IIIT- Hyderabad R&D showcase* 2009.

- Marasim is currently been submitted to *National Research and Development Corporation (NRDC) Budding Innovators* award 2009.

Other Publications

Following are the papers that helped directly or indirectly to shape up this thesis.

- Rohit Khot, Nagaraja Kaushik Gampa and Kannan Srinathan. *Let Only The Right One In: Privacy Management Scheme for Social Networks*, In Proceedings of International Conference on Information Systems Security (ICISS) 2009, Kolkata, India.
- Rohit Khot and Kannan Srinathan. *Elliminati: Bringing New World Image Order*. Technical Report, 2008.
- Rohit Khot and Kannan Srinathan. *IDO: One Time Graphical Password Scheme*. Technical Report, 2008.
- Rohit Khot, Ravikant Poola, Kishore Kothapalli and Kannan Srinathan. *Self-stabilizing Routing Algorithms for Wireless Ad-Hoc Networks*. , In Proceedings of International Conference on Distributed Computing and Internet Technology (ICDCIT) 2007, Bangalore, India.

Part I.

Graphical Authentication: Marasim

2

Framing User Authentication

"The problem with securing assets and their functionality is that, by definition, you don't want to protect them from everybody."

Bruce Schneier

American cryptographer, computer security specialist, and writer

2.1 Let only the RIGHT one in

The world is rapidly moving to a cashless society, with much of the personal information is now being held online. This information is an asset for today's organizations and individuals. The disclosure, improper modification, or unavailability of information may incur expenses (loss) or missed profits for the organization or the individual. Hence, most organizations and individuals protect their information to a certain extent. However, it makes no sense to protect the assets from its righteous owner or other trusted individuals. In effect, all security systems must allow the authorized people in (Permission Problem) and at the same time keep the imposters out (Protection Problem) [Schn00, Cran05]. This process normally involves three distinct steps: Identification, Authentication and Authorization.

- **Identification (Who are you?):** Users first must make some claim of their identity using names, account number or email id (something that will uniquely identify them).

- **Authentication (Prove it):** User must provide satisfactory evidences to verify the identity they are claiming to be. The most common method of user authentication is passwords. Users present knowledge of the secret string (i.e. passwords) which is only known to the claimed identity and hence the identity is verified.
- **Authorization (This is what, you are allowed to do):** On successful authentication, system grants access rights to the users (i.e. the set of actions they are permitted to).

Let us understand these steps with an example. If you have a debit card, then the account number identifies you while the signature on the back of card authenticates you. You are allowed to spend only up to the amount in your account (authorization).

2.2 Context: Usable security

Traditionally we measure the strength of a security system in terms of how secure it is against advanced cryptanalysis and other technical attacks. However, most secure systems today fail due to user negligence, implementation errors and management failures [Bros00].

The field of *usable security* is a relatively new area of study combining two areas of computer science: *human-computer interaction (HCI) and computer security*. HCI is "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [*Hewe96*]. Computer security is a discipline concerned with the "ability of a system to protect information and system resources with respect to confidentiality and integrity", and is associated with several concepts: confidentiality, integrity, authentication, access control, non-repudiation, availability, and privacy [Ross99]. The focus of usable security has been on human aspects of computer security. It studies how human behavior affects the security of the system and how interactive design of a security system impacts its users.

Security experts often say that users are the weakest link in a security system [Bros00, Cran05]. Users misunderstand how to use security mechanisms and do not realize the need for such a protection. User behavior is essentially goal driven and security is usually a supporting task. Users are happy to circumvent the security measures, if security measures try to impede their primary tasks. The classic example of this happening example is the case of user authentication.

2.3 User authentication

Authentication means, proving to someone or to the system, that you are indeed, what you say you are. Methods for authenticating users differ significantly from those used for authenticating systems. This is because of the differences in the processing capabilities of system and human. Humans are not like computers. They are instinctively lazy and can not process and store large amount of data. Therefore, designing an authentication scheme for users is often challenging. Existing approaches for user authentication rely on at least one of the following [Cran05] (as illustrated in Figure 5).

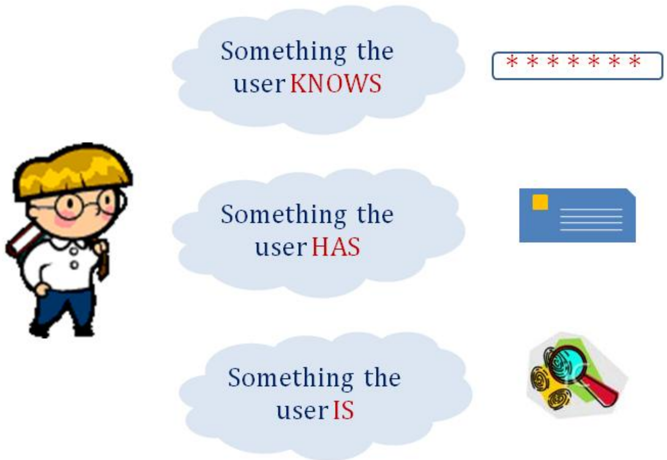


Figure 5: Available Authentication alternatives

Something you know (e.g. Passwords)

This is the most common kind of authentication used for humans. During enrollment, user shares a pre-agreed secret (often a string) with the system and must provide its knowledge in order to login. Anyone who knows or guesses the secret will also be able to authenticate as the original user. Examples include passwords or PINs. Unfortunately, something that you know can become something you just forgot. And if you write it down, then other people might find it. An alternative is proposed in Graphical passwords where users instead of recalling, recognize their secret.

Something you have (e.g. Smartcards)

This form of human authentication removes the problem of forgetting something you know, but some object now must be with you any time you want to be authenticated. And such an object might be stolen and then becomes something the attacker has. Therefore, they are often combined with passwords or PIN to give an extra protection. A typical example of such object is smart card (i.e. a hardware token with embedded microprocessor chip).

Something you are (e.g. Biometrics)

This form of authentication is based on something intrinsic to the principal being authenticated (e.g. fingerprints, retina scan). It's much harder to lose a fingerprint than a wallet. Unfortunately, biometric sensors are fairly expensive and (at present) not very accurate. Moreover, they are not secret and are difficult to revoke (They suffer from the property: Once lost, always lost). Similar to biometrics, behavioral characteristics of a person can be measured and used for authentication. Examples of such schemes are using keystroke dynamics and handwritten signatures.

2.3.2 **Text passwords : Achilles heel of a security system**

Despite the numerous options available for authentication, knowledge based authentication scheme like passwords remains the most common choice for several reasons. Passwords are simple, easy to use and familiar to most users. They are inexpensive and reliable when compared with available authentication alternatives in smartcards and biometrics [Bros00]. Unlike smartcards, passwords are easily portable, and use of passwords does not violate privacy, as biometrics could. However, as Bruce Schneier quotes, "*the whole notion of passwords is based on an oxymoron: having a random string that is easy to remember*" [Schn00]. In other words, following two conflicting requirements must be satisfied in order to effectively use passwords.

- **Usability:** Passwords should be easy to remember and user authentication protocol should be executed quickly and easily by humans.
- **Security:** 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 or stored down in plain text.

Unfortunately, we human beings are instinctively lazy and prefer things that are simple and less complicated. We can not easily remember a random string such as '*r#sL9u5*' and what we do remember something like '*emmy123*' is hardly secure today [Feld90, Flor07, Morr79]. Another problem of concern is our tendency to pick cognitive shortcuts (the path of least resistance) [Dham08]. We are habituated to write passwords down and share them with others. We keep identical passwords on multiple accounts and do not change them even after years [Adam99]. Ironically, attackers are experts in usability; they exploit these password malpractices, by designing social engineering attacks to steal identity information [Dham08].

2.3.3 **Graphical passwords**

Graphical passwords were introduced to overcome the drawbacks of alphanumeric passwords. They are based on recall and recognition of visual information (images, in particular) instead of alphanumeric strings. The interest in graphical passwords is driven by the '*Picture Superiority Effect*' [Nels76, Nels79, Shep67, Stan73]. According to this effect, humans have vast, almost limitless visual memory and pictures are remembered better than the words for a long time. Psychological results show that recall of words declines by *50%* or more over a *72* hour retention period compared with recall of visual objects which drops only by less than *20%* over the same period [Kirk94]. To illustrate, the task of memorizing a list of 15 digits printed on a piece of paper, after a few seconds of inspection, would be difficult for all but small number of people. On the other hand, an aerial picture of '*Taj Mahal'* could be easily memorized so that at some time in future, it could be distinguished from variety of other scenes. A possible explanation to this fact is given by the '*Dual Coding Theory*' [Paiv06] which suggests that verbal and nonverbal memory (image based) are processed and represented differently in mind. Images are mentally represented in way that retains the perceptual features being observed and are assigned perceived meaning based on what is being directly observed [Ocon07]. Text on the other hand, is a form of knowledge representations.

A nice survey of the existing graphical password schemes can be found in [Chia09, Cran05, Suox05]. We can divide graphical passwords into three categories based upon the cognitive activity required to remember the password [Cran05]. These categories are: pure recall, cued recall and recognition.

Pure Recall

Pure recall is considered the most difficult task for memory, since user must remember and reproduce the password from memory without any assistance or cues given by the system. Traditional text based passwords fall in this category. Notable examples of recall based graphical authentication schemes are Draw-A-Secret (DAS) [Jerm99], Pass-Go [Taoh08] and Pass-Doodles [Gold02]. These schemes require user to remember and repeat visual drawing on predefined grid cells. However, results indicate that users prefer to draw symmetric images with less number of strokes which reduces the overall password space [Nali04, Thor07]. Another observation is users often make mistakes in remembering the order and precise grid location, while repeating the drawing. Figure 6 shows working of DAS scheme with a sample password.

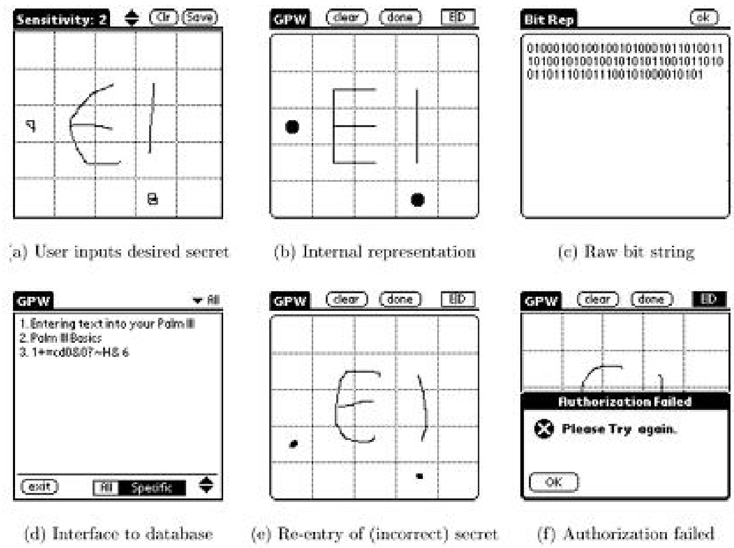


Figure 6: Sample Draw A Secret scheme [Jerm99]

Dunphy et al. introduced background images to DAS scheme to help users remember the location of drawing. Their scheme BDAS [Dunp07] shows that background image reduces the symmetry and offers better overall password space. However, they did not mention other possible attacks on images like predictable patterns and common choices of images.

Cued Recall

In a cued recall authentication scheme, user is given a cue that aids the recall of password from memory. Best example of cued recall scheme is PassPoints [Weid05]. PassPoints is a click based graphical password scheme where password is constructed with series of random clicks on predefined region of an image. To login, user must repeat clicking on the same location (within the system specified tolerance) and in correct order. The image acts as cue for locating the click points. However, PassPoints scheme is vulnerable against dictionary attack as users chose distinct and semantically meaningful regions of an image (known as hot spots) as click points. Example PassPoints scheme design is shown in Figure 7.

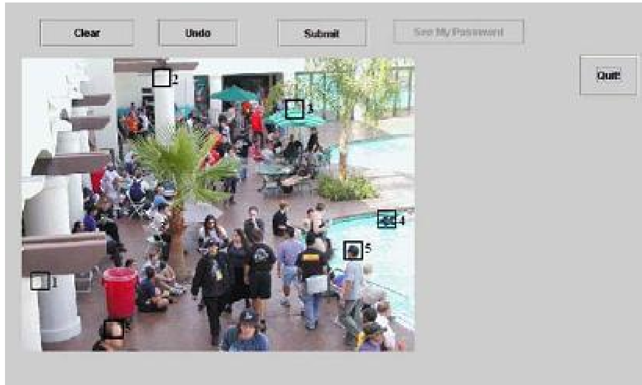


Figure 7: An example of PassPoints scheme [Wied05]

Chiasson et al. proposed an alternative in Cued Click Points (CCP) [Chia07]. In CCP, user clicks on multiple images instead of clicking on multiple regions of single image. Next image to be clicked is displayed based upon the accuracy of the click on the previous image. Though CCP increases attackers overload by increasing the number of images, it does not solve the problem of hot spots completely.

Another cued recall scheme is InkBlot authentication (as shown in Figure 8) [Stub04]. It uses images as cue for text password entry. System presents system generated inkblots and user responds with the correct characters from the word she used to describe the inkblots during registration. Authors suggest that the inkblots should be abstract enough so that an attacker viewing the inkblots would not have an advantage at guessing the password. However, no longitudinal user study has been reported as yet, to measure the level of allowed abstractness and liability of the scheme.

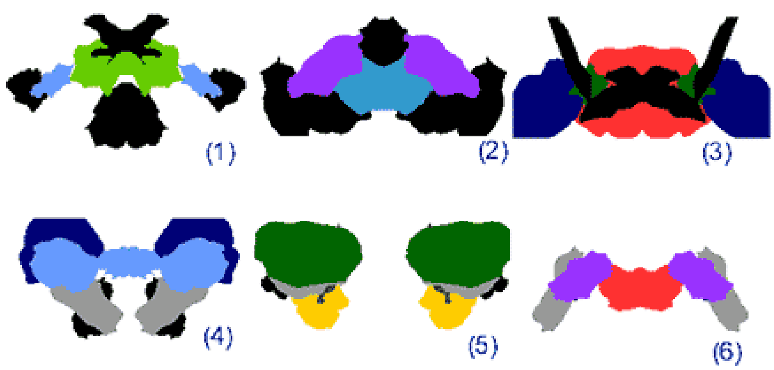


Figure 8: Microsoft's InkBlot Authentication Scheme [Stub04]

The problem with cued recall schemes is in the design of a cue because attacker can see what user can see and understand what user can understand. Ideally, the cue should only help the legitimate user and not the attacker trying to steal the identity information (the cue should be hard to deduce for an attacker).

Recognition

Recognition is easier task than recall [Kint70, Tulv73]. In recognition based authentication scheme, user typically memorizes a portfolio of images during enrollment and must recognize those images among decoy images to login. These schemes can also be viewed as Challenge-Response schemes, where challenge is to correctly identify the images from the set of distractors. Humans are known to be proficient in recognizing images they have seen before [Nels76, Stan70]. Some of the known recognition based graphical password schemes are: PassFaces [Bros00, *Pass09*], DejaVu [Dham00], VIP [Dean05, Monc07] and Use-Your-Illusion [Haya08]. Figure 9 shows representative Déjà vu and PassFaces password scheme.

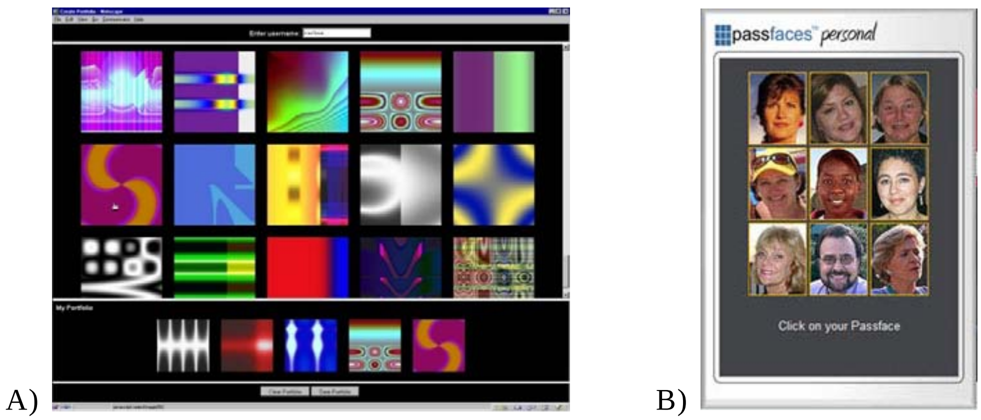


Figure 9: Example of A) Deja Vu password scheme B) PassFaces password scheme

PassFaces uses portfolio of faces and DejaVu uses images of abstract art. VIP is based on recognizing simple natural photographs while Use-Your-Illusion asks users to recognize distorted images. Initial results for most of these schemes are encouraging. However, evaluations indicate common predictable patterns in user chosen images which make the schemes susceptible against dictionary attacks. For example, In PassFaces, users choose attractive faces over plain ones and faces of the people of same gender and race. Weinshall et al. proposed an alternative scheme [Wein06], where password is constructed by computing the 'path' between the shown pictures. This scheme required larger display to be effective and recently been broken by Golle and Wagner [Goll07].

Recognition based schemes discussed above however, do not allow users to provide their own images to create the portfolio. The idea of using personal pictures as passwords was first introduced by Pering et al. [Peri03] and subsequently used by Tullis et al. [Tull05] and Renaud [Rena09]. These schemes allow users to create a portfolio of images from their private photo collection. Since no filtering is done on these images, these images remain closely related to the person and thus are insecure in real life setting. One way to strengthen the schemes is by prohibiting user selection of password images such that all graphical passwords are system chosen. However, system chosen passwords are tough to remember and thus doing so, will have usability concerns [Cran05, Rena09].

A better graphical authentication alternative proposed is 'Story' scheme [9]. In Story scheme, user creates a portfolio of images to make a story and subsequently identifies those images in correct order to login. Images used in Story are of everyday objects like car, ball etc. Preliminary evaluations indicate that Story scheme offers better resilience against guessing compared to Faces (scheme identical to PassFaces [*Pass09*]) [Davi04]. However, users found problems in remembering their story passwords and often forgot the order. Main reason (as the authors note) was very few users actually chose stories, despite being suggested to do so. Instead, they pick images that looked attractive and tried to remember them. The intention behind creating a story is if the sequence of images has some semantic meaning (story), they are likely to be remembered better. A sample Story scheme interface is shown in Figure 10.

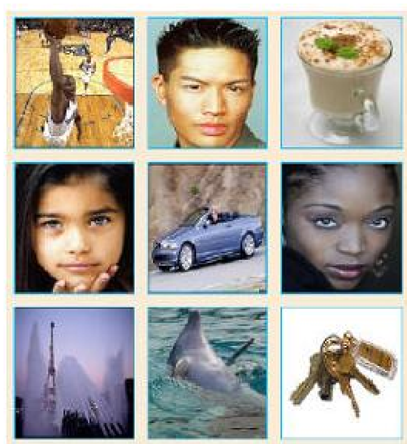


Figure 10: A sample panel of Story Scheme [Davi04]

Strength of a Story based authentication scheme depends on two factors:

- How easy it is, for a legitimate user to create and remember a story?
- How difficult it is, for an attacker to predict that story?

We believe, for a normal user to come up with a story for system generated (provided) images and then subsequently remember it would be difficult. Instead, memorability of Story scheme can be improved by letting users to create story from their own images (i.e. allowing users to provide set of images that already depicts some memorable story). In fact, we show even one image is sufficient to tell a story. Moreover, decoy images should be selected carefully to create overlapping stories to confuse the attacker and improve resilience against guessing.

2.4 Rethinking graphical password designs

When graphical password were first introduced, it was conceived that since images are better remembered than text, and can be chosen from a infinite space of images, users will select and remember graphical passwords that are stronger than the text passwords they normally select [Cran05]. However, this assumption is quickly falling apart, as demonstrated by recent user studies on known graphical passwords scheme we reviewed earlier. We found inherent weaknesses in user choices of graphical passwords. For example, system chosen password images were tough to remember. In contrast, personal or self chosen password images were quite easy to guess for an attacker.

A solution to this problem as we suggested earlier, could be in user education (to create better passwords) or in restricting user selection of passwords. However, either approach can guarantee security only at the cost of usability which is not desirable.

We therefore, ask a question,

> "*How can we achieve the security of system chosen images with the memorability of self chosen images, simultaneously in an authentication scheme?*"

As an answer to above question, we look into the feasibility of Jigsaw based authentication design.

3
A Jigsaw based Authentication design

"There are no extra pieces in the universe. Everyone is here because he or she has a place to fill, and every piece must fit itself into the big jigsaw puzzle."

Deepak Chopra

Endocrinologist, lecturer, celebrity and author

3.1 The world of Jigsaw

We all love to solve puzzles. One of the popular tiling puzzles is *Jigsaw* [*Jigs09*]. In this game, a large picture is broken into numerous small, often oddly shaped, interlocking and tessellating pieces. Each piece has a small part of the picture on it. The challenge is to reassemble the pieces and bring back the original picture. If we have seen the original picture before, then joining the pieces is easy. However, difficulty arises when the original picture is not revealed (known) to us and the pieces are mixed with similar pieces of other decoy images.

We investigate the idea of using personal pictures as passwords with an interesting Jigsaw based Authentication design. We therefore, ask a question:

> "*Can Jigsaw based authentication scheme exist? If yes, what benefits it can offer in terms of security and usability?*"

Let us describe a naïve Jigsaw based scheme.

During registration, we ask the user to upload or choose any memorable image. We then break this image into *p* pieces, such that an individual piece carries no information about the original image. At the time of login we mix these *p* pieces with *n* pieces of other decoy images and present the complete set of *n+p* pieces to the user. User must identify the correct *p* pieces to login. Failure in doing so after *t* attempts will lock the system. The steps are summarized in Figure 11.

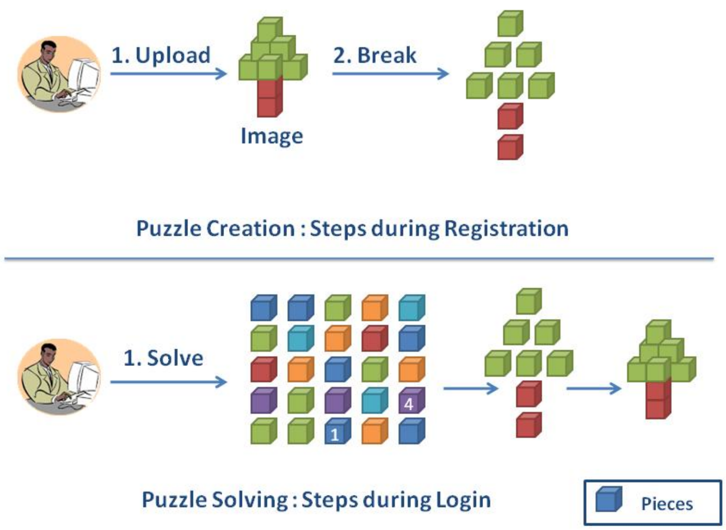


Figure 11: A Naive Jigsaw based authentication design

We first list the benefits of the scheme.

- The scheme allows the use of personal pictures. User is free to choose images from her personal photo collection. Psychological results show that self generated images are better recognized than those that are not [Kinj00]. Moreover, such images are highly meaningful and unforgettable to the users [Tull05].
- We involve the user in the password creation process. Whenever human is actively involved in any cognitive process, an *action event* memory, stronger than the recognition memory is active [Knop05]. It would certainly help in the memorability of passwords.
- The legitimate user knows the original image used in Jigsaw, therefore she can identify the correct *p* pieces with little cognitive effort. On the other hand, the original image is never revealed completely to the attacker. Besides, the correct *p* pieces are jumbled with *n* similar pieces of other images. As a result, it would be difficult for an attacker to find correct *p* pieces in limited attempts.

Although, the above design looks promising, few issues must be resolved before it can be considered as a viable authentication alternative. The most important question is how to break the image so that each piece carries no information about the original image? We certainly can not break the image physically or visually, knowing the availability of many sophisticated pattern matching algorithms (Jigsaw solvers) [Niel08]. We therefore, take help of Web 2.0 phenomena, *Tagging* [Smit08, Tagg08].

Tagging is an emerging approach of organizing information using keywords (called as tags) contributed by ordinary users [Smit08]. People tag their content (for example, images) to easily retrieve them in future [Ames07]. One of the major advantages of tagging is its open ended, nonhierarchical nature, which means that user can assign tags freely based on the cognitive connection between the user and the tagged object [*Sinh05*]. There are already more than 42 million people who have tagged their content and 10 millions are tagging daily which shows growing popularity of tagging [Smit08]. Another example is of a popular photo sharing website, Flickr [*Flic09*] which makes immense use of tagging. Tagging has also been successfully used in CAPTCHAs [Vona05, Chew05] to restrict automated bot (script) attacks. We believe ours is the first attempt (to the best of our knowledge) of using tagging for user identification and authentication.

3.2 Motivation

We are motivated to design an authentication scheme that provides the security of system chosen images with the memorability of self chosen images. Before we proceed to the actual design, it is essential to understand how we perceive images. We often see not what an observance is, but what an observance appears to be. Appearance, as a visual process, is composed of both perception (how we see) and conception (how we think) [Ocon07]. To illustrate, Let us follow a small cognitive experiment. First, look at the twelve images shown in Figure 12. Can you tell how many images are related to each other?

Figure 12: Can you find all the relations among the pictures shown?

It appears that many images among the presented set are related to each other. To name a few, images of *cute girl with rose in hand, birthday tag, rose flower and toys* are related with a *birthday* theme. Another relation we can think of is of a *family or people* where the individual members are shown in the four corner images (See Figure 12). Now, let us try to find the relation among the images shown in Figure 13.

Figure 13: Are these images related to each other? If yes, then how?

After a careful thought, the images in Figure 13 seem to be related with a theme: *a mother and a daughter went to see an art exhibition.* Image of a mirror, here may symbolically denote the act of seeing. Notice that these images are also present in Figure 12. This relation however, seems to be overshadowed by other superior relations described earlier. Let us turn our attention to the image in Figure 14 and try to describe its content.

Figure 14: Describe the things you see in the image (Add tags)

Figure 14 is an image of a cute girl sitting in an art exhibition. We can also see paintings of a beautiful lady (an Indian actress named *Nargis*) and a mirror. Therefore, a possible set of tags for this image can be: *cute girl, art exhibition, nargis and mirror*. Notice that these labels can also be assigned to the four images in Figure 13. In fact, the images in Figure 13 are part of an attempt to recreate the scene we observe in Figure 14. Thus, after we have seen the image in Figure 14, describing the relation among the images in Figure 13 seemed easy.

The idea of Jigsaw based authentication design stuck to us with this small cognitive experiment. Let us reverse the experiment to understand it better.

3.2.2 The Transformation

1. Start with the image in Figure 14, similar to the way we start a Jigsaw puzzle with an image.
2. Describe the image in the form of tags. Let us say we described it as: *cute girl, art exhibition, nargis* and *mirror*. This step is analogous to breaking the image in pieces. However, unlike the original Jigsaw, we break the image into the semantic concepts (tags) occurred in the image.
3. Find similar images for the tags we gave in Step 2. Assume that we found the four images as shown in Figure 13.
4. Mix the four images with other related images as shown in Figure 12, similar to mixing the Jigsaw pieces of many images.
5. Present the set of images to user and ask her to find the correct four images that describe the predefined relation. It is like giving a Jigsaw puzzle to solve.

If we have seen and described the original image in Figure 14, solving Jigsaw puzzle is easy. All we need to do is a visual search for the images that represent the tags given to the original image. In the presented example, we therefore, look for the images of: *cute girl, art exhibition, nargis* and *mirror* among the presented images in Figure 13. However, without the knowledge of original image in Figure 14 and the given tags, solving the Jigsaw (linking the four images) within limited attempts is difficult. The above steps are summarized in Figure 15.

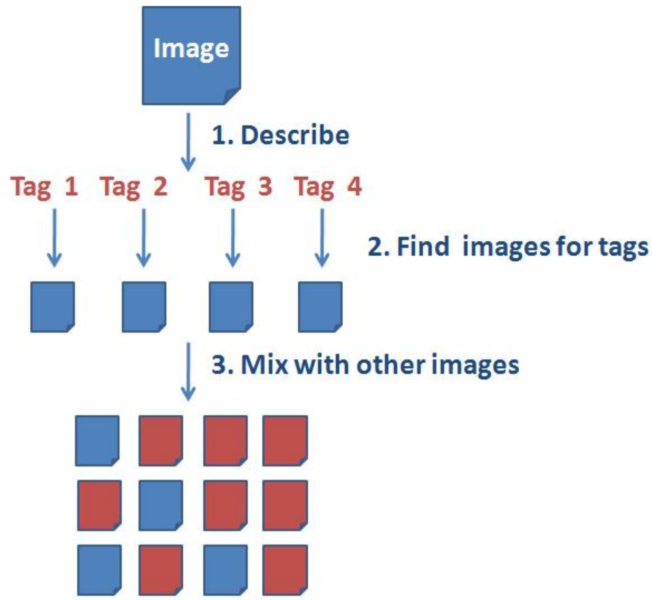


Figure 15: Three step Jigsaw transformation

As a cinematic analogy, this experiment can be viewed as remake of a classic old movie, where story and characters remains the same but the appearances differ.

In the following chapter, we present a working prototype of Jigsaw based authentication design, which we call *Marasim*.

4

Marasim: The Prototype Design

"The real voyage of discovery consists not in seeking new landscapes but in having new eyes."

Marcel Proust

French novelist, essayist, and critic, 1871-1922

4.1 Introduction

Marasim is a novel Jigsaw based authentication scheme using tagging. It allows the use of personal pictures as passwords. Marasim is an *Urdu* word which means *relations or affinity*. Within our context, it denotes the *association among the pictures*. Our scheme is based on human ability to remember a personal image and later recognize the alternate visual representations (images) of the concepts that occurred in that image. These concepts are retrieved from the tags attached to the image by the user. There is no longer any need for users to remember numbers or passwords, all they need to remember is 'one' memorable image from their personal photo collection. Psychologists will tell us that remembering self generated image (autobiographical) is much easier and more natural for the human mind than words or numbers [Kinj00].

4.2 System Architecture

During account setup, user uploads one of her memorable images and provides *p* number of tags for the concepts in that image. These tags are then used to retrieve images that share the same tag from the image database. User picks *p* images (one for each tag) as her graphical password images. At the time of login, user must identify these *p* images presented in the challenge set of *n* images. In the current prototype, we choose the value of (*n*, *p*) as (*25*, *4*) respectively. The sequence of operation is shown in Figure 16 and described below.

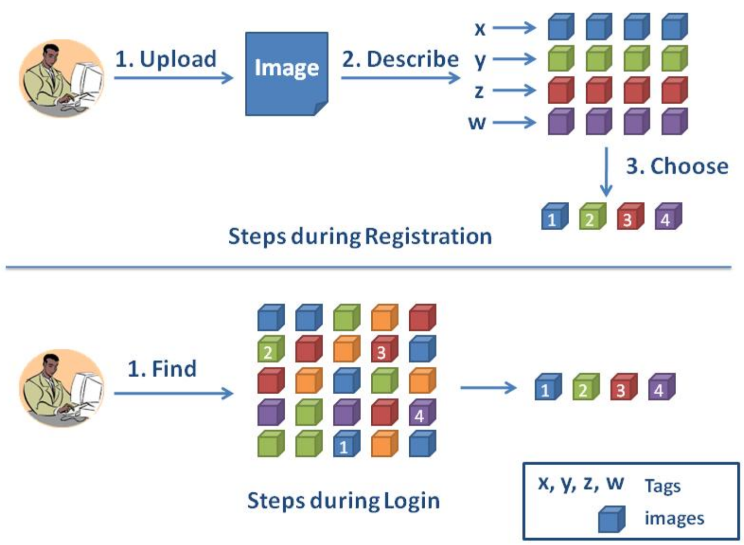


Figure 16: Steps followed during Registration and Login of Marasim

4.3 Registration

Enrolling is a "One Time" event that assigns users a Password and takes them through a process to help them recognize and retain their Password. The entire process takes no longer than 5 minutes. It "must" be completed in its entirety for Marasim to work for users. We assume that registration is done in a secure environment and using trusted channels. Registration involves three steps: Upload, Describe (Tag) and Choose.

4.3.1 **Step 1: Upload**

We ask user to upload or choose a memorable image which is then used to create the portfolio of password images. User is free to upload any image that she can remember for a long time. The image can be from her personal photo collection or can also be chosen from the set of presented images. We suggest that users should not to keep the image used for authentication in insecure or public medium.

4.3.2 Step 2: Describe or Tag

Once the image has been uploaded to the server, we prompt the user to provide four tags for the concepts in the image. A resized version of the uploaded image is visible on the screen (See Figure 17) to help user in describing (tagging).

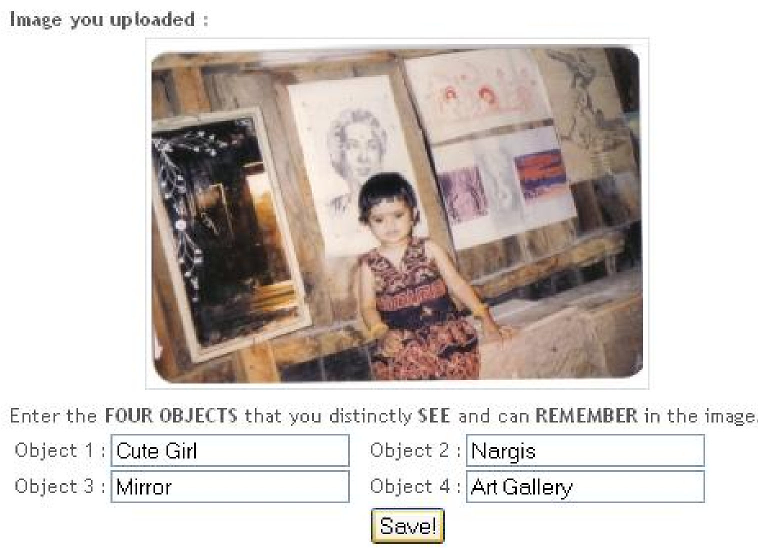


Figure 17: Describe step during Registration : In this step, user describes the content of the image uimage using tags

The concepts can be the distinct objects that user can see and remember from the original uploaded image. User is also welcomed to describe the associated memories with the photo. For example, for the image in Figure 17, user can assign tags that describe the event or place where the photo is taken. There is no restriction on the language or the words used to describe the concepts. Although, we encourage use of simple English words of which visual representation are possible. For example, for the image shown in Figure 17, '*cute girl*' is better tag than a simple tag '*refreshing*'.

4.3.3 Step 3: Choose

We search Google Image Search engine [*Goog09*] with the tags, user entered in the last step. A representative set of images for each tag is then shown to the user. In the current prototype we display the four random images from the search result for each tag as seen in Figure 18.

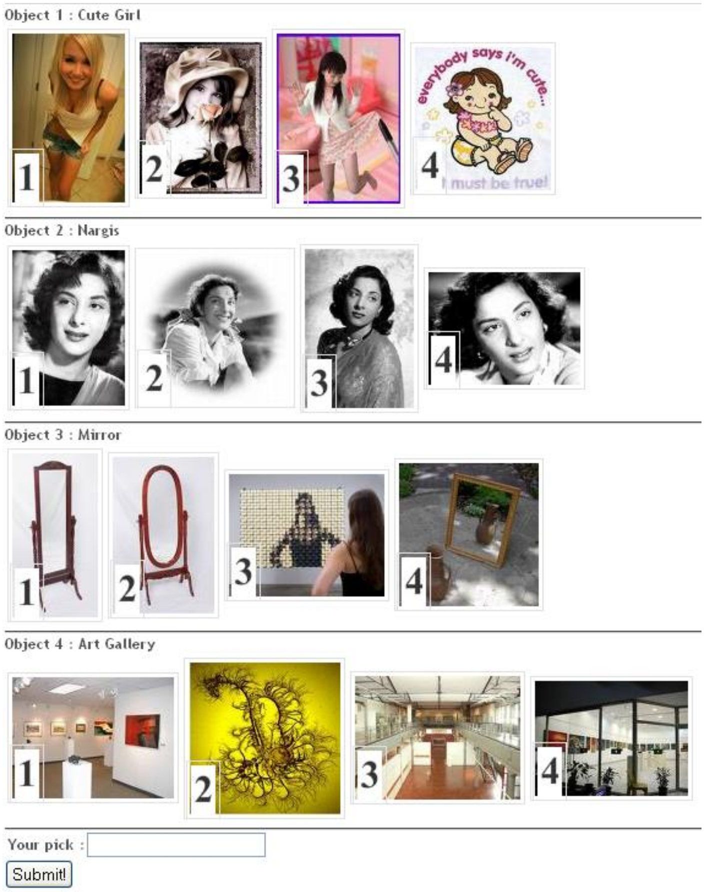


Figure 18: The Choice step during Registration: In this step, user chooses one image from each row that matches the associated tag.

User must select four images (one from each row) that correspond to the four tags. To select a particular image, she must enter the associated number in the textbox below (See Figure 18). For example, to select first image from first three rows and third image from the fourth row, user should enter '*1113*' in the textbox. If the user is not satisfied with the presented images, she can ask for replacing them (all in one go, or one row at a time) and a new set of images is showcased as per request. User also has the option of going back and altering (improve) her entered tags. Both these options are allowed only before the final confirmation (before pressing the '*Submit*' button). Once the user has submitted her choices of images, a confirmation message is displayed about successful completion of registration. The confirmation message displays selected four images and indicates it to the user that these would be her password images. At the time of login, she must recognize them amongst decoy images.

In this prototype, we have used Google Image search to find images for the entered tags. However, Google Image search may not always return perfect image results. Therefore, we recommend the use a human annotated image dataset, like Flickr [*Flic09*] or ESP [*Espg09*] to find proper images for the given tag.

4.4 Post-Registration Processing

Once the user is successfully registered, post-processing begins in which we create a unique image challenge set for the legitimate user. The challenge set consists of *4* password images and *21* decoy images. Ideally, decoy images should be easier for legitimate user to neglect and should be complex enough to add confusion for an attacker.

In Marasim, the four password images share a semantic relationship with respect to the original secret image (i.e. each image depicts a concept occurred in the original image). The password can be vulnerable if the relation is directly visible to the attacker. Therefore, we must add confusion by overriding this relation with other superior relations. There are two ways of doing it:

1. Find all related tags for the password images and the entered tags (using *Natural Language Processing* techniques) and add images of them.
2. Use images of existing relations from the user image dataset.

In the present prototype, we choose the second option and create set of decoy images from the images of other users. We maintain a secure database of images that are retrieved from Google Image Search and selected by the users during enrollment. Once the Image challenge set per user is created, it remains consistent across all login sessions (i.e. No new image is added or deleted).

4.5 Login

At the time of login, system presents the challenge set of *25* images randomly placed in a *5×5* grid. Each image has a number (between *0* to *9*) associated with it. To login, user must identify her password images and enter the corresponding number in the textbox below (similar to what we did in the *choose* step during registration (See Figure 18). User is allowed to enter the password images in any order. The numbers associated with the images keep changing with each login session that becomes *one time access code* for a particular login session. The screenshot of the login session is shown in Figure 19.

Figure 19: Screenshot of the Login session

In the example shown in Figure 19, let us say, during registration user has composed her password with images of *cute girl, nargis, art gallery and mirror* (Refer Figure 17, 18). Now to login, she must locate those four images and enter the associated numbers in any order. (in the preceding example, she should enter '*1963*').

User can recall his/her password images in two ways:

1. Recalling his/her secret password image and doing a visual search for the contents (tags) described in the original image.
2. Recognizing the four images that he/she saw and learnt during enrollment process.

5 Evaluating Marasim

"A few weeks after 9/11, a reporter asked me whether it is possible to prevent a repetition of terrorist attack, 'Sure …' I replied, 'Simply ground all aircrafts.'."

Bruce Schneier

American cryptographer, computer security specialist, and writer

5.1 Security of Marasim

We recommend that Marasim, should be implemented and deployed in systems where offline attacks are not possible and where number of guess attempts are limited per account in a given time period (For example, ATMs). We assume that all communication between the user and the server is made secure through SSL, thereby avoiding simple attacks based on network sniffing. Below we list countermeasures for the possible attacks on Marasim.

5.1.1 Brute force attack

Simplest of the attack against any authentication scheme is to randomly guess the correct password. Attacker can try to randomly guess the one time access code. The access code is constructed using four digits (from 0 to 9), giving password space of 10^4=10,000. However, the one time access code changes with every new login session, therefore brute force for one time access code does not seem economical. Alternatively, attacker can try to guess randomly the password images. In the current prototype, Marasim presents a challenge set of 25 images which contain the four password images. There are 12,650 possible combinations of choosing 4 images out of 25 images. Therefore, the probability that a single random guess succeeds is 1/12,650.

In addition, as we suggested earlier, we can use a counter *c* that limits the number of unsuccessful attempts. Thus, the probability that attacker succeeds in randomly guessing the password within *c* trials, will be c/12,650.

Another way to strengthen the security is by increasing the size of the image challenge set as well as the number of portfolio images (which is currently 25 and 4 respectively).

5.1.2 **Dictionary attack**

Dictionary attack is more sophisticated attack than brute force. Instead of random guessing, attacker tries to crack the password using a dictionary of most common passwords. Marasim password consists of four independent images that are related to each other with respect to one secret image. Therefore, to crack Marasim password, an attacker can try to find all possible relations between the images shown in the challenge set. She can also exploit and use previously obtained information about the user, e.g. through social engineering. However, we argue that it would be difficult for an attacker to find the correct relation (and thus, the four password images) within limited attempts. We present following countermeasures to defend such attacks.

- Original secret image is never revealed to the attacker. Attacker also does not know the tags given to the image. Therefore, without prior knowledge of the image and the tags, guessing the relation between four independent images is difficult.
- Images in the challenge set are public images are unique per user and show no easily recognizable relation with the legitimate user. Attack on Marasim has to be a dedicated one.
- We override the relation between the images with multiple superior relations as described in post-registration processing step. As a result, challenge set consists of images which share many overlapping relations. Predicting the correct relation (the one user has picked) is hard within limited attempts.

Our proposed scheme, Marasim can be vulnerable in cases where user uploads a publically known image and provides tags that are closely related to each other and to the user (with no superior overlapping relation). For example, user uploads an image of a *birthday party* and assigns four tags as: *cake, balloon, candle,* and *gifts*. Such relation can easily be identified among the presented images. In such cases, we plan to give relevant feedback to the users about their password strength. Although in the current prototype design, this feature is not included and left as a future work. We explain below how it can be done.

Password Strength meter (Future Work)

Flickr API [*Fapi09*] has useful function called *GetRelated* that displays all related tags for a given tag based on cluster usage analysis. We plan to use these function to calculate relative strength of a password based on the relatedness of the four tags. If all the four tags are related to each other then password is considered as weak, whereas if none of the four tags are related to each other, password is strong. We define a medium strong category of password that has two related tags. For example, A password comprising (*cake, balloon, candles, gifts*) is a weak password while (*girl, mirror, art gallery, nargis*) is relatively strong password.

5.1.3 Social engineering attacks

Social engineering [Jaga07] includes any technique used to trick people into divulging their credentials or private information to untrustworthy parties. It is often easier to obtain a password or credentials from the legitimate user than trying to break into a system by other means. Some of the popular social engineering techniques are:

Shoulder-surfing:

Shoulder-surfing [Tari06] are targeted at capturing passwords during authentication through direct observation, or through external recording devices such as video cameras, while the legitimate user enters the information. Availability of high-resolution cameras with telephoto lenses and surveillance equipment make shoulder-surfing a real concern if attackers are targeting specific users and have access to the same geographic location as these users. This is especially problematic in public environments, but may not be as serious a threat in other more private environments.

Phishing

Phishing [Dham06] attacks involve tricking users into entering their credentials (username, password, credit card numbers, etc.) at a fraudulent website that is masquerading as a legitimate site. Users normally reach these phishing websites through spam email enticing users to click on an embedded link that directs them to a website designed to look like a site for which they have a legitimate account. When users attempt to log in, attackers record the user's credentials and subsequently use them for fraudulent purposes.

Malware

Malware (i.e., malicious software) [*Mala09*] includes any unauthorized software that is installed without a user's informed consent. Such software has a malicious purpose, and can include viruses, worms, and ActiveX or JavaScript components [Prov08, Ross08]. One category of malware is intended to gather confidential information, including user credentials, from the computer on which it is installed. For example, key-loggers record keyboard input, while mouse-loggers and screen scrapers capture mouse actions and the contents of screen memory, then either send this information back to the attacker or otherwise allow attackers to retrieve it.

To defeat popular social engineering attacks, we employ the idea of Probabilistic One Time Passwords (POTP) [Bedw08]. It involves association of single number with multiple images. Explaining the theory behind POTP is beyond the scope of this thesis. Thus, we illustrate the use of POTP in our paper with a small example. Consider a 5×5 grid (similar to Marasim login grid) as shown in Figure 20. For simplicity, we have not shown the actual images. Assume that the password images are the one shown in orange color. Therefore, one time access code for the user is corresponding numbers with the four password images, which is 4367 in this example.

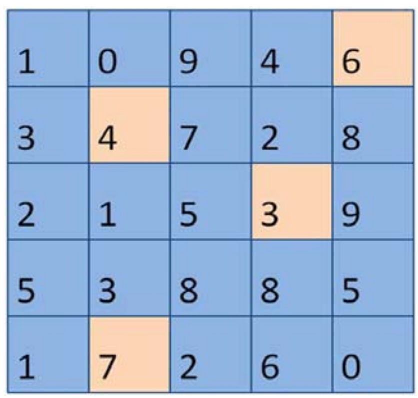


Figure 20: A sample *5×5* grid similar to Marasim Login grid. User identifies the password images and enters the associated numbers to login

An attacker can grab the one time access code by following simple social engineering techniques.

- **Via Malware attacks:** Using key loggers and Screen scrappers.
- **Via Shoulder surfing:** Capturing by either looking over the shoulder or with the camera.
- **Via Phishing:** by fraudulent emails or websites.

However, the numbers in the grid are repeated multiple times. Moreover, one time access code is unique only for the ongoing session. Therefore, attacker can not easily correlate the grabbed one time access code with the shown password images (since, many images correspond to the same number). For example, the numbers *4*, *7*, and *6* repeat twice while the number *3* repeats thrice within the login grid shown in Figure

8. It means that attacker would require at most $2 \times 2 \times 2 \times 3 = 24$ attempts to find the correct four password images.

In general, an attacker needs $a \times b \times c \times d$ attempts to find the user password from an obtained one time access code, where a, b, c, d denotes the frequency of occurrences of four numbers from the one time access code within the grid. The best case occurs when all the four numbers in the one time access code repeats three times giving a total of 81 attempts, while the worst case happens when all numbers in the access code repeat only twice, needing 16 attempts to crack the password. We therefore recommend blocking the account, after a few unsuccessful attempts to avoid such attacks similar to ATMs.

5.2 Usability study of Marasim

We conducted a formal user study aimed at testing the usability of Marasim. Is it simple and convenient to use? Can user remember Marasim passwords? What aids the recognition, is it the original image or tags? If answers to above questions are affirmative then our scheme can aid to memory benefits of earlier graphical authentication schemes.

Another goal of the study was to learn the characteristics of the user chosen passwords. We are particularly interested in knowing the kind of image user uploads to create her password and the relatedness of the assigned tags.

5.2.1 Participants and setup

A formal user study was conducted within the university campus. We recruited 30 participants by sending invitation emails. All 30 participants were graduate students from the university. 23 participants were male while 7 were female. Their ages ranged from 20 to 28 years with a median age of 25. All participants reported use of authentication schemes before for emails and financial reasons, but none of the participants knew or were familiar with graphical password design.

5.2.2 Procedure

Our usability test spanned over three months and consisted of five sessions. A web based prototype of the Marasim was created and is available at (Link removed due to anonymity reasons). All the participants were sent an email with the URL of the site and instructions for the usage of the website. Online help explaining the working of the scheme was also kept at the site. Table 1 shows details of each session which includes the task that participants needs to complete in that session.

Table 1: Details of each user study session

Session No.	Date	Tasks
1	First day	Registration and Training
2	One day later	Authentication
3	One week later	Authentication
4	One month later	Authentication
5	Three months later	Authentication and Questionnaire

In the first session, Participants created their password by uploading one of their memorable images and subsequently tagging and choosing the four password images. A confirmation message was displayed on successful completion of registration. Each participant required approx. three minutes to register. The training involved two back to back login sessions to get the user familiar with the authentication process and image challenge set. To login, participant must identify her four images out of the set of 25 challenge images. In the second, third, fourth and fifth sessions (which happened after one day and one week and one month and three months of registration respectively), we asked participants to authenticate themselves by sending them a reminder mail (i.e. to login to their accounts by recognizing their password images amongst distractors). At the end of the fourth session, we requested participants to fill out questionnaires for the sake of qualitative analysis.

5.3 Results

We report the viability of Marasim in terms of [Cran05]:

- Accuracy (Number of successful Logins).
- Efficiency (Time required to login).
- Predictability (Password Strength)
- User satisfaction.

5.3.1 Accuracy and Efficiency

We start with reporting the accuracy and efficiency of Marasim. Authentication is considered successful, if the participant is able to login by correctly identifying her password images. Each participant was given a maximum of three attempts to login. Table 2 shows the combined results of Login Success rate and the mean time needed to login. We also report the number of users who required more than one attempt to login.

Table 2: Login Success rate and Mean time needed to login

Time interval	Number of Successful Logins	Number of users with first attempt failed	Mean time (Std. Deviation)
Session 1	100% (30/30)	0	24.7 (3.9)
Session 2	100% (30/30)	0	18.1 (4.4)
Session 3	100% (30/30)	4	17.6 (4.1)
Session 4	97% (29/30)	6	21.3 (5.0)
Session 5	93% (28/30)	7	25.2 (4.7)

Results show that, all the participants were able to login successfully for the first three sessions. Only four users needed more than one attempt to login to their respective accounts in the third session. After one month, one user found problems in remembering his password. At the end of the user study (i.e. after three months), two users made mistakes in their login, while seven participants needed more than one attempt to login. We investigated the cause of failure for the participants and found mismatch between their password images and entered tags. That is, the image they chose as password did not directly match with the tags. For example, one user picked an image of a girl for a tag '*divine*', which after one month delay he was unable to identify correctly (he later admitted that he remembered the tag, but could not able to recognize the image). The reason behind this kind of failure can also be from the use of imperfect Google Image Search engine. We therefore, plan to use a superior human annotated image set so that each image is correctly described by the tag.

Table 2 also provides the details of the time required to login. As we can see, login times do not vary significantly across all the sessions. The maximum delay occurred in the first session when participants were using the scheme for the first time and in the last session which was after three months gap. Although, login time of 20 sec seems inappropriate for practical applications, we suggest that it can be sufficiently reduced if the images are small, of same size and are stored locally. Moreover, we can also clicking instead of typing the Marasim password.

5.3.2 User choices and Password strength

During enrollment, users upload one of the memorable images and assign tags to create the portfolio of password images. We were interested in knowing the categories of the uploaded images and the relatedness of the tags. These findings are important to determine the predictability of the passwords. Figure 21 a) shows the image categories while Figure 21 b) shows the relatedness of the tags.

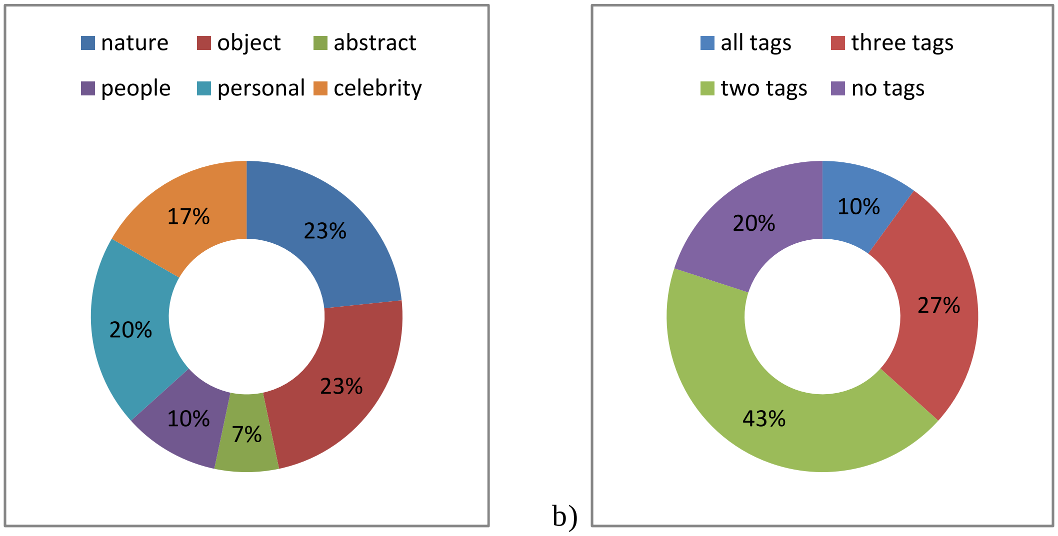
a) b)


Figure 21: a) Common categories of the uploaded image b) Relatedness of the assigned Tags

We particularly identified six different categories from which users choose their memorable image. These categories are: 1) Nature scenery 2) Everyday objects 3) Abstract art 4) People and Animals 5) Personal images 6) Celebrity photos.

As we can see from Figure 21 a), none of the image categories was dominant in the selection. The most frequent categories are Nature scenery and Everyday objects while the least frequent category is of abstract art images. 20% of the users also chose their personal photos as passwords.

We have also evaluated the relatedness of the tags as seen from the Figure 21 b). If the tags are closely related to each other, then the relation among the password images can be easily identified. Therefore, lesser number of related tags in a password is advisable. We found most of the password images have at most two tags that are related to each other, whereas only 10% of images have all related tags.

5.3.3 User Satisfaction

Upon completion of the study, the users were requested to answer a set of questionnaire providing feedback about the scheme and to write down any specific comments. Four multiple choice questions were: 1) Marasim password is easy to remember 2) No one can easily guess my password 3) I can prefer over text password and 4) I can use Marasim effectively with practice. Available responses to these questions were *Yes*, *No* and *Neutral*. Figure 22 shows the responses of the participants to these questions.

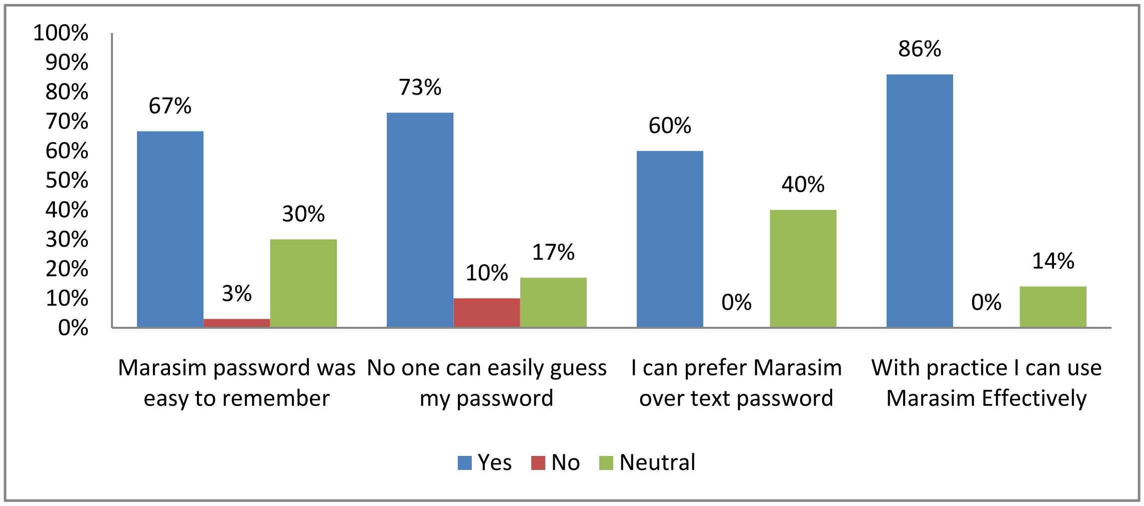


Figure 22: Questionnaire responses to Marasim design.

All questionnaire questions had median value of neutral of higher, showing high levels of satisfaction. When asked about how they were able to remember their passwords, 43% users said that the original image and the tags helped them to recall the password. 27% users said they simply recognized the four password images, while rest 30% users responded with combination effect of original image and recognition of four images.

5.3.4 **Summary**

To summarize, results of the user study show good improvements in terms of memorability with only two participants had problems in remembering their password after three months. Authentication protocol executed fast and there are not many variations in login time. Users were satisfied with the scheme both in terms of usability (most of the participants said they can easily remember Marasim password) and security (73% of participants were confident that there password is secure referring to question 3, refer Figure 22). We believe these evidences are enough to show the viability of proposed scheme.

6

Contribution Summary of Marasim

"Our world is divided into facts, because, we so divide it."

Susanne Langer

American philosopher of art

Marasim is a robust graphical authentication scheme, strong enough for banking, finance and e-commerce. Its strength lies in its simplicity and unique 'Jigsaw' way of working. We provide the ideal - a simple-yet-strong system, which is easy enough for users to grasp - but offering huge levels of security to keep the fraudsters at bay. The main contributions of our work are enumerated below.

6.1 Usability Features:

Improved memorability:

Users often struggle in recalling their passwords composed of "*cold random*" string of alphanumeric characters (e.g. "*s$a3#2rk*") and what they do remember something like "*rohit123*" is hardly secure today. We propose a novel solution where user just remembers one (memorable) image from her personal photo collection and recognizes the content within that image to login. Psychological results show that personal images are highly meaningful and unforgettable to the users and are better remembered than text.

Cognitive scalability:

Our proposed scheme is language independent. Use of pictures and symbols makes the scheme ideal for use by the people of all abilities and age with any level of literacy. However, the visual way of working may not be suit visually impaired persons. To accommodate them, we provide an alternate audio based challenge where image tags are read aloud along with the associated numbers and the user respond orally with the numbers of her password images.

Faster, simple and stress free login experience:

Our proposed scheme achieves the desired security without the aid of any extra hardware or token. Marasim also do not need costly software installations or dedicated hardware to run. The login interface is intuitive and specially designed by keeping the cognitive abilities of the users in mind. The authentication process executes faster without putting much load on the user.

Software as a service (SaaS):

Marasim can easily be integrated into current secure online authentication architecture and can replace passwords or serve as a second form of authentication. It is compatible (Adaptive) across various financial domains and transaction types like ATM, e-commerce, mobile commerce.

6.2 Security features:

Strength of two factor authentication without an extra hardware or token:

Most authentication solutions today rely on an alternate factor (often a hardware token or smartcard) to strengthen the security. However, use of hardware token poses an extra overhead in terms of cost and usability (user must carry the card always to login). Moreover, Hardware token can also be stolen, tampered and damaged. Instead, we propose a “one of a kind” solution which provides the strength of a two factor without an extra hardware token.

Resistance against brute force attack:

Simplest of the attack on password based authentication scheme is randomly guessing the password. To crack Marasim password, attacker must guess the four password images from *5x5* grid. However, in Marasim, there are *12,650* ($^{25}C_4$) possible patterns of selecting four images out of *25* images. It is sufficiently greater than normal PIN based security (which is *10,000*). Furthermore, security of Marasim can be strengthened by limiting the login attempts.

Mitigating dictionary attack:

A common attack on passwords is a “dictionary attack” where attacker tries to attack with a dictionary of most common passwords. Personal images are easy to guess given some knowledge about the person. However, In Marasim, each user authenticates himself/herself with a unique set of “system chosen random” images. The personal image used in constructing the image set, is never revealed to the attacker. Moreover, it is extremely difficult for an attacker to figure out correct password from set of presented images. In short, we blend the security advantages of “system chosen images” with the memorability gains of “user provided (personal) images” to create a robust authentication design. It is “big gain” in terms of security and usability over existing authentication mechanisms.

Security against social engineering attacks with WYSWYE strategy:

We employ ‘Where You See is What You Enter (WYSWYE)’ strategy to defeat prevalent forms of identity theft. At every login, images are randomly placed in the grid and has a different number (0-9) associated with it. With every login, the positions of images within the grid and the numbers keep changing, making the password unique per session. Thus identity theft using following social engineering techniques is extremely difficult.

Resistance against Shoulder surfing:

User never actually selects her true password images, by clicking on it. All she does it is to enter four numbers that corresponds to her four password images. So anyone who is piping over the shoulder or even with hidden cameras can only be able to see your one time token password of four digits. Since the same number actually correspond too many images within the grid, it is hard for an attacker to predict the correct images in limited attempts. The token password changes with each login session. Thus shoulder-surfing is not effective.

Protection against (key logging / screen scraping) malware attacks:

Our proposed scheme offers no advantage with key loggers as the same key pattern points to many images within the password grid. Even if someone captures and records the screen, he will still not be able to deduce your password as every time grid pattern is randomly generated and the password that is formed with given pattern is used only once.

Anti-Phishing:

By phishing a legitimate user into revealing the password, attacker will only get her one time token password corresponding to the presented challenge grid. Attacker must record multiple sessions (phish user multiple times) and try out several combinations before she can successfully deduce the correct password.

Scalability and Flexible design:

Marasim can easily be scalable to desired level of security by increasing the grid size and choice of images (password length).

Marasim can add strength to any device and system. For example it can be made Two Factor by separating the login entry from the grid. (For example, login entry can be at the ATM, but grid can be transferred to the secure mobile or hardware token or vice versa)

6.3 Commercial potential

Users (Value on Experience):

- Easy, intuitive and stress free login experience.
- No need to carry an extra hardware or token.
- Marasim password images are easy to remember images than normal text based password. The authentication process executes fast and with ease.
- Suitable for people of all ages and abilities. Language independent and simplified login process.
- Security Assurance: Trusted environment to carry transactions without the fear of social hacking.
- Cognitive flexibility and user centric control: Marasim provides both textual and pictorial support to accommodate people with different cognitive abilities.
- Alternate audio based challenge can be set specially for “visually impaired” customers.

System (Value on Investment):

- Software as a service: Marasim is a hosted solution that can easily be integrated into existing security architecture. It requires no software setups and dedicated hardware to run.
- Low total cost: There is no user software or certificates to install. Fewer failed login attempts, reduced customer support calls and cost per authentication is in pennies; maintenance is minimal.
- Builds customer trust and confidence by preventing prevalent forms of social hacking and identity theft.
- Flexible: Marasim can be layered with other security mechanisms to strengthen the login and provide a strong second factor of authentication.

- Configurable: Various attributes of Marasim (number of images, length of the access code etc) can be customized to increase security and meets the specific needs of the customer.
- Advertising opportunities: Customers have the opportunity to advertise on the Marasim and use sponsored images. The ad images can be tailored to meet the needs of the user demographic.
- Verified and tested performance: Marasim scheme is fully tested and acknowledged by the users.

6.4 Summary

We proposed a novel jigsaw based authentication design that provides the security of system generated images with the memorability of self chosen images. The novel contribution of this work is in the construction of image portfolio, to be used for graphical authentication. We designed and tested a web based prototype of Marasim. We discussed possible attacks on our scheme and we could defend against each of them. Results of the user study provide evidence for improved usability and memorability. Our future work includes providing relative feedback on the password strength and testing the scheme with large audience of all ages.

Part II.

Image Annotation: GoFish and iCAPTCHA

7

Image Annotation & Human Participation

"There is no way of determining in advance which detail is relevant to an aesthetic interest; every detail can and ought to play a part."

Roger Scruton

English conservative philosopher, writer, activist and composer

7.1 Introduction

Today, picture making is almost a routine part of life. With the proliferation in digital capturing devices and decreasing storage costs, people are motivated to communicate with other online by sharing photos, videos and thoughts (blogs). As a result, visual information (images and videos) is widely available on diverse topics and from multiple sources. Creating a structured collection of images is a tedious task especially when it is going to be viewed by others and they have given the right to access something they specifically need in the collection. Modern image search engines such as Google [*Goog09*] collect and index images from other sites to provide access to the wide range of images. However, most of the search engines are word matching tools that can only retrieve images that match the words in a keyword based query. Such engines often struggle to find the right image for a specific need from the large database of images and to reduce the clutter that often comes with the selection.

For example, a given the popularity of cricket in India, a search query '*cricket*' on Google image search engine returns a collection of over 15 millions of images (Refer Figure 23 shown below). Not surprising, it is a diverse set of images which include live match captures, image of cricket team, stars photos and even the screen captures of cricket video games. Notice that even the images of cricket insects are also present. Will the avid cricket fan, feel happy with such result?

Figure 23: An example showing the diversity of Image search results

7.1.2 **Problems with Image Search**

For those, used to viewing well indexed collections of quality images, the results of large automated image search engines will probably disappoint. The poor offering of images is not surprising, since it reflects the randomness and unevenness of the web. However, we also believe following factors also contribute to such offering.

1) Query Dependency:

Current image search engines require users to be specific in terms of the search query while seeking for the visual targets. Most of the times however, it is hard for users to express the need in words. As a result, search query tends to be short, too general and sometimes ambiguous. If the query is not detailed enough, search engine returns plenty of information (images) consisting all subcategories. User then needs to laboriously browse through all information or keep on refining query to get the desired result (image).

For example, while searching for images of old Indian actress Amrita Singh, it is better to type "*Amrita Singh*" as query than a general query "*Amrita*", which would result in set of images mostly dominated with images of "*Amrita Rao*", another popular Indian actress. User therefore, must know complete name of the actress (detailed query) to get the desired result, which many users may not know.

2) Talking with words when we mean images:

Image search engines are word matching tools which analyze the metadata associated with the image (e.g. tags, keywords, and text in same page) for indexing and categorization of images. They assume that content of an image is related to adjacent text appearing in the page. Unfortunately, the language of the web, HTML is clunky and that clunkiness permeates the majority of the Internet. Therefore, the text adjacent to images is often scarce and can be misleading sometimes [Cars96]. Results show that, only 20% of the images on web have proper 'alt' tags [Vona04].

3) Inefficient Content Based Image Retrieval (CBIR)

Current computer vision algorithms try to extract meaning by analyzing the visual content (e.g. shape, color, and texture) of the image. However, such approaches have found limited success only in specialized setting and are yet to match the performance of humans in image recognition and understanding [Datt08, Smeu00]. There are two main reasons for CBIR's lack of success as a technology in commercial image search engines. First, they are struggling to bridge the '*semantic gap*' [Datta08] between low level visual features and high level semantics. And the other reason is low efficiency.

Above pitfalls restate that image should possess meaningful textual metadata (in the form of tags) to facilitate its access. If the image collection has been extensively annotated, technique such as faceted search will help user filter down a collection and show potential targets for browsing [Whit06]. However, the only method currently available to obtain precise image description is through manual labeling or tagging. The reason is simple. When it comes to labeling and organizing images, man has traditionally outperformed machines for most tasks. As Datta et al. quote [Datt08]

> "*This distinction is due to fact that text is man's creation while typical images are mere replica of what man has seen from birth, concrete descriptions of which are relatively elusive. The interpretation of what we see is therefore hard to characterize and even harder to teach to machine…*"

Humans have little difficulty in describing the image. However, motivating them to annotate images is tedious since they find the task laborious and not particularly engaging.

7.2 Productive procrastination

Humans, unlike computer processors, require some incentive to become a part of a collective computation. Humans have mostly avoided tagging despite its benefits in recall and retrieval. Thus, different attempts have been made to lure humans in annotating (tagging) images. Three most prominent approaches are: embed tagging in social activity, provide monetary incentives, and design special purpose games.

7.2.1 **Social activity:**

Ludicrop Inc. [*jeff06*] realized very early the human potential for tagging and developed a social tagging system, Flickr [*flic09*]. Flickr is a popular image hosting and sharing website (service). Its popularity has been fueled by its organization tools: mainly *tagging*. Tagging allows user to attach set of textual labels known as tags to images and browse with it. People come together, share their photos, and tag them collaboratively the way they want. Currently, Flickr claims to host more than 4 billion images [*Bfli09*]. However, in Flickr, tagging is a choice; it leaves many unlabelled images from uninterested users. Since we can't force users to tag even the images of their choice, users trying to increase their exposure will only tag, and beyond community there is little reason for an average user to tag the images properly.

7.2.2 **Money:**

Another option used for tagging is to pay monetary incentives. The concept was introduced with Amazon Mechanical Turk (AMT) [*Amaz09*] which co-ordinates workers and developers in solving human intelligence tasks like tagging, for a small payment in return. Some recently launched search engines such as TagCow [*Tagc08*] utilize AMT. They pay $1.20 per hour to participants tagging images. However, current image search engines do not have an alternate source of generating revenue like Advertisements on image search ages, therefore, paying humans to tag images is not well justified from the business point of view. Moreover, getting unbiased, precise description for images from unknown contributors is also a problem of concern.

7.2.3 Entertainment or Fun (games):

Each year, people around the world spend billions of hours playing computer games. What if all this time and energy could be channeled into useful work? What if people playing computer games could, without consciously doing so, simultaneously solve large-scale problems? Sometimes people like to think and be challenged; sometimes it is just for pastime. Online games are thus seductive methods for encouraging people to participate in a collaborative work such as tagging. Such games constitute a general mechanism for using brain power to solve large scale problems. In fact, designing such a game is much like designing an algorithm—it must be proven correct and efficiency must be analyzed [Vona05]. People play such games for entertainment, and not because they want to voluntarily tag images. Existing human algorithmic games designed for tagging are ESP [Vona04] and Phetch [Vona06].

ESP (Extra Sensory Perception):

ESP is the first human computation game. It has been hugely successful: millions of image tags have been collected via playing the game, and even after a many years of its launch, people are still interested in playing the game.

In this game, two randomly paired players try to agree on labels for single image. On match, both players score points. The word or tag on which the two players agree, then becomes the taboo word for that particular image. Next time, when the same image is shown to two new players, they can use this taboo word to describe in the game. Authors argue that if the image has generated an extensive list of Taboo words (words that player can not use) and pairs are unable to agree upon new label and preferring to pass the image, then image can be considered as completely labeled [Vona04].

However, problem with the ESP game is it encourages users to assign the obvious labels, which are most likely to lead to an agreement with the partner. For example, let us assume the image in Figure 24 is shown to players in ESP game.

Figure 24: What tags will result in easy agreement in ESP game?

Even if one or both the players know the name of the person in the image (that is, Gary Oldman in this case) they will not tag the image as '*Gary Oldman*'. Since the player is not sure that his random partner also knows the name and tags similarly. Therefore, their obvious guesses or tags are: '*man*' and '*spectacles*'. These tags are not wrong in any sense, but the question arises whether one has to rely on humans to obtain them.

Further ESP game gives players an easy option of passing on difficult image and difficulty is kept up to the user to decide. Therefore to score high in the game, player will prefer to pass the image rather than applying the mind to extract new meanings. It can also be seen in their game statistics as only 1023 of 293760 images have five or more labels.

Phetch

Phetch [Vona06] game is aimed at fetching natural language descriptions for images, which help blinds to navigate images and the web. In this multiplayer game, one player (Describer) describes the image to other players (Seekers) and they try to find the image from a search engine for given description. On success, all players get points. However, we found some problem with the game play.

First to enjoy the game play it is "must" for either Describer or seeker or both to possess complete knowledge about the given image. If Describer doesn't describe the image properly to the seekers, it is hard for seekers to find desired image. For example consider the image in Figure 25 (similar image is also shown in original paper [Vona06]).

Figure 25: To search for this image, players must know the names of the celebrities

If this image is given to describer, he must know that persons in the images are Justin Timberlake and Janet Jackson. A general query describing the scene like “two singers, a man and women in a concert when man ripped a piece of women’s shirt” is hard to search with. For such description seeker should also be knowledgeable enough to replace the man and women with Justin Timberlake and Janet Jackson Respectively. This assumption is too strict and will therefore fail to attract global audience.

7.3 Rationale for semantic annotations

The main problem or difficulty associated with extant human annotation approaches is that person who is tagging complete ignores the possibility that a searcher for the image may not know what he is looking for or may not able to recollect what he wants. Seeing and saying may have meaning to one observer, but the same visual experience may not have the same meaning to another observer.

To illustrate, a person while tagging follows his own interpretation of image and tag accordingly. He may not be aware of other possible or complete interpretations of the image, or can simply ignore them. For him, ‘*the meaning of image is what struck to his eyes*’. In effect, the tagged image remains accessible only to him and to people who also interpret the image similarly. Any other person, who wishes to find the same image but queries differently, will not possibly find the image due to mismatch in interpretation with the owner or tagger.

We must realize that an image itself does not have any meaning. It is merely a rectangular shape with colored amorphous blotches of various sizes. While looking at image, we interpret and compare the blotches to objects or situations we encountered before. The cumulative of all visual clues in an image gives us the ability to constitute context and meaning. These interpretations vary from person to person. O’Conner et al quote [Ocon07],

> “*The veracity of an image lies in the viewing engagement with the image, not in its description*”

8
Emergent Semantics

"To see in limited modes of vision is not to see at all."

S. I. Hayakawa

Canadian-born American academic and political figure

8.1 Motivation

Although humans are very good in describing images, the resulting annotation may not always be precise and of good quality. We believe following two fundamental problems must be addressed with manual annotations.

8.1.1 Fallacies of misplaced concreteness

Let us first understand how we generally describe an image. The way we describe the image is as much function of "*how we see*" as it is function of "*how we think*" or more appropriately "*how we are made to think*" [Ocon07]. To illustrate, let us describe the image we see in Figure 26.

Figure 26: Can you describe (tag) this image?

Figure 1 shows a familiar pattern of oval shape and green color. Therefore, we reply as "*It is an Apple*". Now, consider the same image, within a group of other images as shown in Figure 27 and try describing it.

Figure 27: Is 'Apple' a good tag for the first image, now?

Our discriminating mind can see that all images are of Apple, yet we know, the first one is different from the rest. We therefore, look for the features that separate this image from the rest of the images and describe the first image in Figure 27 as "*It is an apple fruit.*" If we progress in the same way and assemble the same image with other images of Apple fruit, our description becomes even more precise and we say "*It is a green apple fruit, partially eaten from left*."

This small exercise in the prequel shows that, '*we do not always say what we see*'. We all saw and knew from the start, the features present in the image, but we never felt the need to express it completely. We described the image with an abstract notion of an '*Apple*'. The problem with manual annotation is our tendency to oversimplify things. We tend to get lost in what Alfred North Whitehead called as, '*Fallacies of misplaced concreteness*' [*Fall09*]. If we do not describe the image precisely (i.e. communicate the complete cognitive experience through language), the image under view remains hidden in the crowd of similar images and needs cumbersome browsing for retrieval.

8.1.2 **Crippled Viewer Syndrome**

Greisdorf and O'Connor in their book [Ocon07] coined the term *Crippled viewer syndrome* which refers to the cognitive disconnect a viewer experiences on seeing an unknown image. Without the background knowledge necessary to interpret the image, a viewer can describe the image only in terms of the signs it contains.

Therefore, the true (intended) meaning of the image is often does not get communicated in the annotation. For example, let us look at the image in Figure 28. If we do not know that the person in the middle is '*Barack Obama*', then our description would probably be limited to '*Basketball team*', '*black kid among white kids*', '*group photo*' etc.

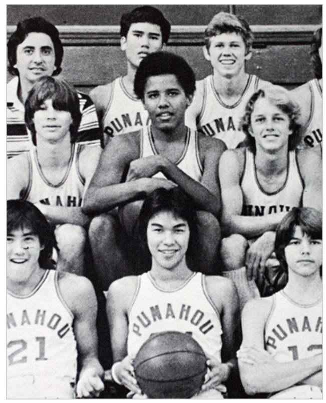


Figure 28: An example image of Crippled Viewer Syndrome

These problems with manual annotation exist because, to most individuals, applying word descriptors to a textual document and to an image appear to be the same sort of activity. However, they cannot be the same activity, since describing the document text is an extraction process, and there are (usually) no words to extract from an image (photo, painting, etc.) Creating a structured collection of images based on words requires an underlying framework that connects the collection to its viewers through purposeful communications [Ocon07].

8.2 Emergent Semantics

A solution to the annotation problem lies in the theory of *Emergent Semantics* [Sant01]. According to this theory, image in general does not have meaning, but the meaning emerges from the interaction with the user and by placing the image in the context of other images. The small exercise in the prequel is a proof for the same. When the image of the apple is placed in the context of other images as shown in Figure 27, we are able to describe (or made to describe) the image more precisely.

To elaborate, emergent semantics theory [Sant01] reveals that:

Meaning of the image is contextual

It depends on the particular condition under which the annotation is done and particular user that is annotating the image.

Meaning of the image is differential

Meaning of the image can be made manifest by differentiation between an image which possess that meaning and image which do not. Further, Meaning of the image can also emerge by association between different images that share that meaning.

Meaning of the image is grounded in action

Meaning of the image can also be established from the user actions when the image is presented to her.

8.3 Emergent semantics approach to image annotation

Using the emergent semantics theory, we present a novel approach for annotation of images (Refer Figure 29). Our approach is a recursive way of extracting new meanings of the image by repeatedly placing the image in the context of other similarly described images (similar to the exercise followed earlier in this chapter).

The procedure

The steps are as follows:

1. Present user with an image *A* to describe.
2. Get the description D_0 for the image *A*.
3. Find all images from the database that corresponds to the given description D_0.
4. Present the original image *A* along with images found in the step 3 and ask user to describe the original image again with respect to other shown images.
5. Get the new description D_1 for the image *A*.
6. Repeat the steps 3 to 5 using the new description D_1.

The process stops when the user can not able to differentiate between the images and attach a new description (tag) to the image. At the end, each image will have a rich set of *n* image tags, $D = \{D_0, D_1, \dots, D_n\}$.

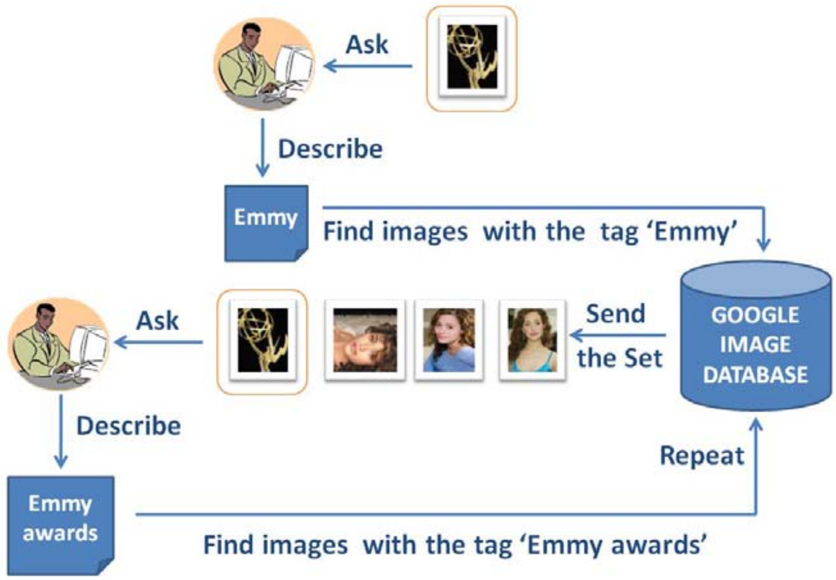


Figure 29: An emergent semantic based approach to image annotation

8.3.2 Benefits of the approach:

Productivity:

If an image A receives a description or tag *D*, then the tag D not only describes the image A, it also tells us that tag D differentiates the image A from the rest of the presented images. These accompanying images therefore, will not have the same tag *D*. As a result, annotation is done faster and on the complete set of presented images.

Features ranking:

With every new round, the image *A* receives a new description $D_{i.}$ In the first round people describes the most striking feature of the image, all subsequent rounds, next most striking features about the image are introduced. This hierarchical way of tagging helps in ranking the received tags.

However, success of the above approach depends upon active human participation, which can only happen if humans find this task engaging and fun.

Following two chapters present two intelligent system designs for tagging images using the above mentioned emergent semantics approach. First one is an interactive fun game called 'GoFish' while other one is the image based CAPTCHA design named 'iCAPTCHA'. We stress that our method is not meant to compete against existing techniques of image search and retrieval. Results of our designs can be combined with these techniques to provide a powerful solution.

9

GoFish: A Game With A Purpose

"For Surrealists, photographs were full of meanings that resulted from the intersection of unexpected happenings, and the artist's objective was to stimulate the emotions with the element of surprise."

W. Naef

The J. Paul Getty Museum Handbook of the Photographs Collections

9.1 Introduction

We present GoFish, an intelligent system for semantic annotation of images from an online game. GoFish is a web variation of standard Go Fish, a popular playing card game. Behind GoFish game design is a strong emergent semantics theory that ensures superior annotations. Our technique like previously proposed games [Vona05], is not dependent upon computer vision techniques, but on people's existing perceptual abilities and desire to be entertained.

9.2 GoFish: A popular playing card game

Go Fish is popular card game [*Gofi09*], played among two to five players with a deck of *52* playing cards. One of the players is chosen as a *Dealer*, who first shuffles the cards and distributes them equally among all the players including him. Objective of the game is to win most Books of cards where, a Book is a collection of four cards of same rank. For example, four kings, four aces, etc. Figure 30 shows interface of an online version of GoFish game [*Gofg09*].

Figure 30: Online version of card based Go fish game [*Gofg09*]

Player to the left of the *Dealer* starts the game. He asks any one of the players, for a card of specific rank and from a specific suit (Hearts, spades, clubs and diamonds). For example, "*John, Do you have '6 of hearts'?*" However, in order to ask, the player himself must have at least one card of the same rank, i.e. '6' in this case. If *John* has the requested card, he has to give it to the player who asked for it. Whenever, the request for the card is successfully fulfilled, the same player continues asking for other cards. But if the player addressed i.e. *John* in this case, does not have the requested card, he says "*Go Fish!*" It means, player who asked for the card, looses his turn for asking and now John can start asking for cards. Once the player collects all the four cards of specific rank to complete one Book, he shows them to all and keeps them face down on the table. The game proceeds in the same manner until all the thirteen books of cards are won. The player with most number of books is declared as the "*Winner*".

Transformation

Let us see, how we can transform this game into a game with a purpose of efficient tagging of images. As a first thought, the transformation seems easy. Just replace the playing cards with images. The catch here is, whenever a player asks for a card, he has to describe it in plain text. If we capture all such description, it will solve the problem of describing the images. However, we must ensure that cheating is minimal and generated descriptions are accurate. We describe below the modified version of the GoFish.

9.3 GoFish: our proposed game

GoFish is a turn based game played among four players. One of the players is chosen as *Narrator* and others are *Seekers*. The deck of the cards is a collection of eight images those we wish to tag. The replica of the entire deck is always visible at the bottom of the game screen as shown in Figure 31.

Figure 31: Screenshot of GoFish game in action

Figure 31 shows a snapshot of GoFish game window. The four players are seen at the four corners of Figure 31.

Narrator starts the game by shuffling and then distributing the cards equally among all the players (including him). Therefore, each player holds two cards (they are marked with orange bubble. Refer Figure 31). The players do not know other players' cards. The objective of the game is to win (collect) all the cards.

9.3.2 **Game play**

Narrator gets the first chance to ask for a card. In order to ask, Narrator must enter proper description of that card which can be approved by Seekers. He describes the card (that he wants to ask) in the plain text and sends the description to all the Seekers. All the Seekers on their respective turn try to identify the card that matches the received description (all the cards are visible at the bottom of the screen, refer Figure 31). Seeker scores points for finding the correct card. If majority of Seekers is able to find the correct card, Narrator gets points for the valid description. However, if none of the Seekers is able to find the correct card, then Narrator is penalized for incorrect description. Narrator will lose his turn of asking after two such penalties.

Once the card has an agreed description, Narrator picks any one player and asks him for the described card. If the player has that card, he has to give it to the Narrator. Narrator then continues asking for more cards. However, if the player does not have the requested card, he says "*Go Fish!*" and becomes the Narrator. He can then ask for the cards. Present Narrator takes his position as Seeker. The game continues in the same way till one player wins all the eight cards, he is then declared as "Winner".

9.3.3 **Strategy**

GoFish maintains a scoreboard which lists top scorers and players who won maximum number of games on the current day, week and till date (all time winners). We present below strategy for winning the game and scoring good points.

Winning a game:

To win the game, a player needs to collect all the cards. At the start of the game, player does not know which player has which card (Probability of correct guessing is 1/3). For a player to improve his chances, a good strategy is to pay attention to who seeks which card He can then capture those cards in the next turn if he can remember whom to ask.

Scoring High points:

Narrator scores points for every card he correctly describes, while Seeker gets points for every correct card he finds. If card is correctly described every time and all the Seekers are able to find the card then everybody gets equal points and none gets chance to become Narrator for the next game. To beat other players in scoring, Narrator can opt to give description to which minimal number of Seekers agree (Minimum number of Seekers will be able to find the correct card). Narrator can not give wrong description, to which no Seeker will agree. Therefore, a better strategy would be to describe the image in more specific details, hoping that not all the Seekers knew about it. It is a gamble, but worth taking. Similarly, for Seeker to score high he should look to find the correct card each time and follow the above strategy if he gets chance to become Narrator after wards.

9.4 Description quality

A *proper* description is *correct* if it makes sense with respect to the image and *complete* if it gives enough information about its content. The description becomes *superior* if it conveys beyond what can be seen from the image.

9.4.1 **Accuracy**

We argue that descriptions generated by playing GoFish will always be accurate. We list following points in support of it.

- All the players are randomly grouped from all the players online to avoid colluding.
- Narrator can not give description that does not correspond to any image (irrelevant) or more than one image (incomplete). In both the cases, he will lose points as Seekers may not be able to find the correct card.
- For an easy agreement with the Seekers, Narrator might want to describe the position of the card as discriminating factor, for example, "*First image*", "*second from right*". We make sure that no match (agreement) is possible by randomizing the order in which cards are laid on every player's screen. Therefore, the first card from left on Narrator's screen may not be the same on any other player's screen. Similarly distinction based on colors can be avoided by using grey scale images.
- If Seeker select wrong image for given description, he gets negative points. Since the players are randomly grouped, the probability that all the seekers choose same image which is different from the one Narrator has picked is low.

9.4.2 **Completeness**

We expect that Narrator will describe the image with only features that separates the image from the rest of the image. These features may not be sufficient to describe the image completely. We therefore follow emergent semantics theory discussed earlier and group the image with other similarly described images in a new game instance of GoFish. Now, Narrator of the new game can not give the same description as before, as it will now correspond to two or more images (The previous description can not separate the image from the rest of the images). Therefore, Narrator must explain the image further to score points in the game. An image can be said to be completely described if Narrator can no longer able to distinguish the image from the rest of accompanying images and asks for replacements.

9.4.3 Superiority

We call a description Superior, if it conveys beyond what is conveyed (what we can see) in the image. Players who are playing the game are viewers, not the originator of the images. Getting superior description by player (viewer of the image) compared to the originator of the image is difficult. The viewer may not know the meaning of what he is seeing, yet he knows what he is looking at. We intend to help him by giving supportive images which are similarly described to obtain the desired background knowledge. Even then a player may not give superior description for an image. To make him do that, we introduce the strategy of beating other player in points as we explained in Section 9.3.3. Player (Narrator) can attempt to give superior description, hoping that not all, only one Seeker knows that the description holds (makes sense) for the selected image.

9.5 Implementation and user study

GoFish is implemented in Adobe Flash [*Adob09*] and Smart fox server [*Smar09*] is used for socket connections. Upon completion of the game, server records all activities of the player in the database for future analysis. Currently, GoFish is in beta stage. GoFish is made available within the university campus for testing. We present below the results of preliminary study conducted.

A total of 30 players played the game over the period of 2 weeks. All participants were students from the university campus with their age in the range of 19 to 28. A tutorial is provided to let them learn the rules of the game. To resolve the cold start, monetary incentives were provided for winners. We used image dataset comprising of top 20 results of 10 popular search queries from Google. Each game lasted for roughly 14 minutes. Total number of games played was 33 that generated 231 descriptions for the 150 images. 78% of the received descriptions were precise. 60% of the players played the game more than once, while 8 players played the game for four or more times. Upon completion of the game, the users were requested to answer a set of questionnaire providing feedback about playing the game and to write down any specific comments. GoFish received on average a score of 7.2 on the 10 point scale. Most of the players said they would love to play the game again.

9.5.1 Discussion

GoFish is a game that tests not only player's ability to distinguish among images but also his memory and above all his luck. To score high points, player needs to describe image correctly, while to win the card, he needs good luck and concentration in the game. A factor like randomness which comes with luck was missing from the earlier games like ESP and Phetch [Vona04, Vona06].

One criticism on GoFish is not simple like ESP. However, GoFish is adopted from an existing popular game and we kept the game play nearly same as the original. Therefore we believe players who loved original Go Fish card game, will also appreciate this design for its novelty. Our user study indicates that most users easily understood the rules and liked the game with only 20% needed hands on demo.

With GoFish, we introduce competitive factor in the Game with a Purpose design [*Gwap09*]. While earlier games like ESP and Phetch are collaborative in nature.

GoFish in aimed at receiving more search specific tags. Although, we can also obtain good tags with the existing games, but it will require an alternation of the game, which we believe, will take away the fun.

9.6 Summary

Until recently, research in image annotations has largely been centered on development of effective techniques such as games to lure human into annotation. However, less focus was given on the quality of the resultant annotation. We discussed the problems with extant methods and presented a different perspective on annotation using semantic theory of images. We introduced an intelligent annotation scheme GoFish. We explained the design with potential benefits. At the end, we gave a preliminary user study that show the game is fun to play. Although the game is not released to public, we hope in the near future, our game will help annotate majority of the images.

10

iCAPTCHA: A Productive CAPTCHA

"For Surrealists, photographs were full of meanings that resulted from the intersection of unexpected happenings, and the artist's objective was to stimulate the emotions with the element of surprise."

W. Naef

The J. Paul Getty Museum Handbook of the Photographs Collections

10.1 Accessibility of CAPTCHAs

Internet has been one of the best things happened in the last few decades. It has transformed the way we live and look at the world around us. Most of us start the day by checking emails and reading news articles over the web. From train reservations to online shopping, from chatting to file sharing, almost all essential services are now available online and at free of cost. However, we never had thought that one day we would have to prove our intelligence (that we are human) before we can access any of these services. For example, most of us surely would have encountered a similar crazy image shown in Figure 32, while accessing a popular site Google.

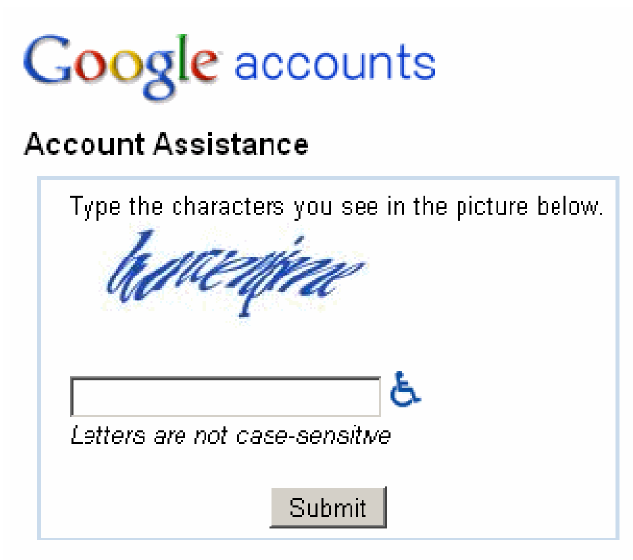


Figure 32: Can you read this CAPTCHA?

This image contains several distorted characters that we must correctly identify and type, in order to access the site content. Since current computer programs like Optical Character Recognition (OCR) are yet to achieve the accuracy of human eye while reading a distorted text, this image represents an instance of a test which is easy for humans to solve and yet difficult for computers to solve. This test is widely known as CAPTCHA [Vona03]. A CAPTCHA stands for '*Completely Automated Public Turing Test to tell Computers and Humans Apart*'. It is a program that can generate and grade tests that humans can pass but current computer programs cannot [Vona03].

The need for such differentiation or the CAPTCHA arose from the vulnerabilities of online forms and data entry. Before CAPTCHA, there was no easy way for a system to verify that the form is filled by a user (human) and not by some automated program running on behalf of the user. A classic cited example is of online polls conducted by Slashdot website [*Slas09*]. The website conducted a poll to determine the best graduate school in USA. At the end of the polls, MIT and CMU not surprisingly, stood tall in terms of the gathered votes against all other colleges. However, the real reason behind this success was the execution of automated programs, giving plenty of fake votes to MIT and CMU. Polluting an online voting system was just one example which shows the power of automated script attacks. Other examples include creating fake email accounts, spreading plenty of junk emails etc. CAPTCHAs work as sentries against these attacks, since solving CAPTCHA is difficult for automated programs and is relatively easy for humans.

Today, most of the popular websites like Google, Yahoo, and Wikipedia use CAPTCHAs as a standard security mechanism to defend automated script attacks. As a result, their online services are now not directly accessible. A user must solve the CAPTCHA to access the service. However, solving a CAPTCHA requires a substantial human cognitive effort. Based on the type of cognitive effort required to solve CAPTCHA, CAPTCHAs can be classified into three categories.

- **Text based CAPTCHAs:** They require users to read and type distorted text rendered in an image.
- **Audio based CAPTCHAs:** They rely on sound or speech recognition by the users.
- **Image based CAPTCHAs:** They ask users to perform an image recognition task.

Text based CAPTCHAs are the most popular of the three, considering they are easy to deploy, intuitive and they have potential to offer reasonably good security. However, many of the existing text based CAPTCHA implementations [Mori03, Yanj07] have been broken recently. It has prompted the CAPTCHA designers to create more complex (distorted) CAPTCHAs (like the one in Figure 1) taking away its usability. As we can see in Figure 1, the shown CAPTCHA image is barely readable by human eye, causing strain to the eye and fatigue by unnecessary multiple solving attempts. Therefore, CAPTCHAs are effective only if they are robust (computers can not solve them) and usable (humans can solve them) [Yane08]. Unfortunately, text based CAPTCHAs fails to achieve both robustness and accessibility (usability) simultaneously which prompt us to look for other possible alternatives. Image based CAPTCHA is one such alternative because recognizing images are far better and fun than reading complex distorted text. This approach was first proposed by Tygar et.al in [Chew04] where they discussed alternate image recognition CAPTCHA designs. Other attempts in creating image based CAPTCHA include Assira from Microsoft [Elso07] and hotCAPTCHA from HotOrNot website [Hotc009]. However, all the proposed image based designs were created only as suitable alternatives to text based CAPTCHAs. On the other hand, we are also interested in tapping the human effort spent in solving CAPTCHA into a useful work.

10.1.2 **Motivation**

People around the world, solve millions of CAPTCHAs everyday, if put together, will easily amount for hundred or thousand hours of human effort per day [Vonc08]. Although the main purpose of CAPTCHAs is to prevent automated script attacks, the effort humans put in to solve them is otherwise getting wasted. We thus ask a question:

"*Can we channel the wasted human effort into some productive work? If yes then how?*"

The idea of productive CAPTCHA was first introduced by Luis Von Ahn, the man who also invented the CAPTCHA mechanism. He proposed a novel CAPTCHA design called as reCAPTCHA [*Reca09*], which helps in reading and archiving old textbooks. The OCR (Optical Character Recognition) software used in reading books, can not effectively interpret text from the old books that has become pale, dirty and yellow over the time. On the hand, human eye can easily pick and figure out what the text is. In reCAPTCHA, user is presented with CAPTCHA consisting of two text words to interpret. Verifying system knows answer for one of the two words, while the other word comes from the old text book, which system can not read. This fact is never revealed to the user. He therefore must read both the words and enter them correctly to access the web content. As a result, each time he is solving a reCAPTCHA, he is helping the system to read and digitize books.

Figure 33: A reCAPTCHA helps to digitize books

We take inspiration from the reCAPTCHA design and aim to solve the problem of image annotation and in doing so; we wish to improve the image search and retrieval.

10.2 iCAPTCHA: Overview

We present iCAPTCHA, a user friendly CAPTCHA design. Instead of annotating images fresh from start, we try to improve the default labels the images have got. That is we attempt to obtain the more proper labels (subcategories) for an image. For example, with our design, we improve the label from general category such as '*apple*' to more specific as '*apple fruit*'. Our premise is based on the human ability to recognize images, label them and put them into proper categories. Figure 34 shows the overview of the scheme.

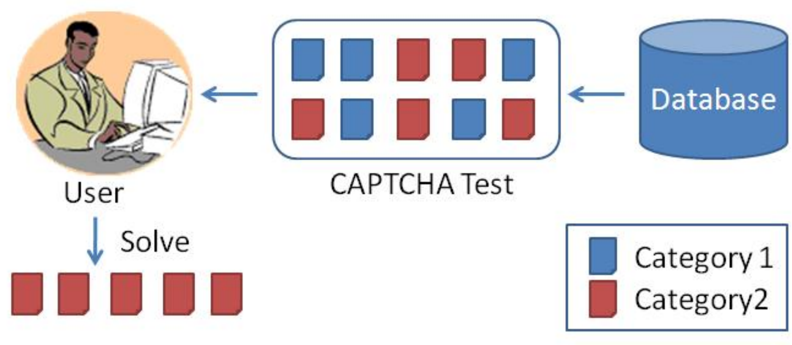


Figure 34: Overview of iCAPTCHA scheme

We pick randomly a set of 12 images belonging to the two different image categories from the image database and present them as a CAPTCHA test. The task for the user is, to identify all the images belonging to one specified category. We can explicitly tell the category name or show a representative image belonging to the category. If the user correctly identifies all the images belonging to the desired category, he/she is considered to have 'passed' the test. On the hand, failure in recognizing the correct images will mean that user has failed the CAPTCHA test.

The use of images makes iCAPTCHA, language independent, less stressful and suitable for people of all ages and at any level of literacy.

10.2.2 iCAPTCHA: System Architecture

Before proceeding to the actual design it is essential that we understand the concept of 'tagged database' and 'test database'. Related to them are the concepts of 'category' and 'sub category'. We first briefly explain them.

Category and subcategory:

A category represents a short or ambiguous search query (e.g. 'apple') which when fired on popular search engine, normally results in images of many subcategories mixed together (e.g. 'apple fruit', 'apple logo', 'apple iPod' are subcategories for a category 'apple').

Test database:

CAPTCHA image test database is prepared by crawling the web for different image categories (as defined above). All resulting images are stored according to their respective categories (image queries) in a secure database at the server side.

Tagged database:

We recruit people or ask some trusted volunteers to describe (tag) the subcategories of few representative images, chosen at random from the test database. All labeled images are then stored according the described subcategories in a separate database called ‘tagged database’ at server side. The support of volunteers is needed only once at the beginning, the tagged database gets updated after each successful iCAPTCHA test.

The concept of ‘*test database*’ and ‘*tagged database*’ is analogous to the concept of test data and training data in the field of Content Based Image Retrieval.

10.3 iCAPTCHA: Proposed design

iCAPTCHA test comprises of 12 images. First, we fix one category say ‘apple’ from ‘test database’ and two related subcategories say ‘apple fruit’ and ‘apple logo’ from ‘tagged database’. We retrieve few images at random say ‘*n*’ (minimum 1 and maximum 11) from the ‘tagged database’ corresponding to the selected subcategories. Rest *’12-n’* images we select from ‘test database’ that belong to the selected category. We shuffle the selected images and present them to the user in a *2x6* matrix (two rows containing six images each) as shown in Figure 35.

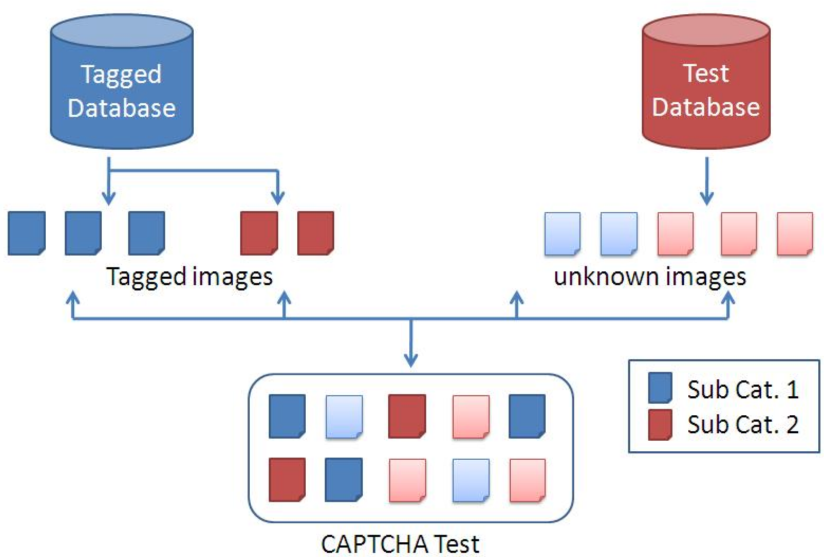


Figure 35: iCAPTCHA test generation process

Users then must identify all the images that belong to the specified category say ‘apple fruit’. Since user does not know which images are from ‘tagged database’ (i.e. already tagged) and which are not, the best option for him/her is to recognize and correctly select all the images of the required image category. The selected images would not be just from the ‘tagged database’ but could also be from the unlabelled ‘test database’. Therefore, each time user is solving an iCAPTCHA, he/she is actually helping in labeling the images from ‘test database’ that are part of the given test. The evaluation process is shown in Figure 36.

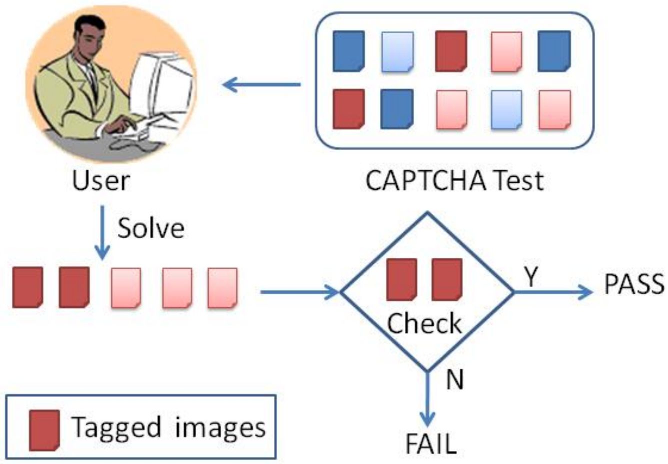


Figure 36: iCAPTCHA test evaluation

10.3.2 iCAPTCHA: Implementation

A working prototype of iCAPTCHA is created in Adobe Flash with PHP at the back end. MySQL is used for data storage. We present two alternate designs:

Design 1

In the prototype, the desired sub category is specified in words. For example, Figure 37 shows a sample iCAPTCHA test where user is asked to identify all images of '*apple fruit*'.

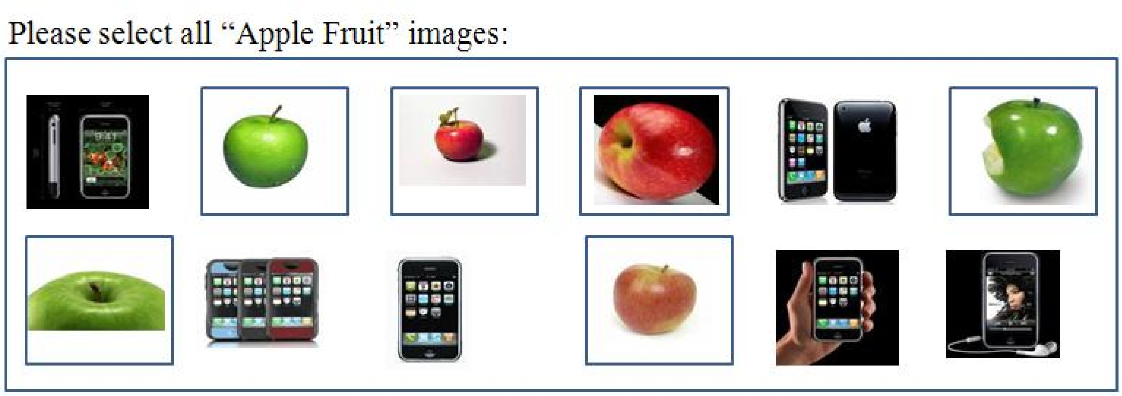


Figure 37: prototype design 1 of iCAPTCHA

As we can see in the Figure 37 that user has correctly selected all the 'apple fruit' images. But how we know it? Consider that, the first four images (first two images from each row) are from the 'tagged database' and rest eight images are the from 'test database'. For evaluation, we therefore check whether user has selected two correct images (the second image from first row and first image in the second row) and he has not selected the two wrong images (first image from the first row and second image from the second row). These four images are from 'tagged database' whose sub category we already know. Since user has correctly done that, he/she has successfully passed the test. Note that, in the process we also acquired the knowledge about the sub categories for the rest eight images (which are from 'test database'). That is, we came to know the images that belong to the specified category (i.e. third, fourth and sixth image from first row and fourth image from the second row are also images of 'apple fruit'. See Figure 37)

Design 2

In the prototype, the desired sub category is shown in with a representative image. For example, Figure 38 shows a sample iCAPTCHA test where user is asked to identify all images that are similar to the challenge image shown in the left (The challenge image was given is of an actress named '*amrita rao*'.

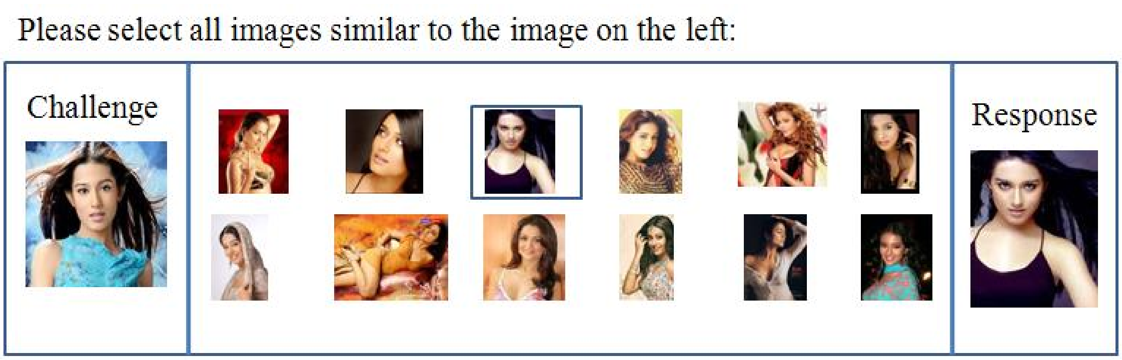


Figure 38: Alternate design of iCAPTCHA

User must accurately pick all the images that she thinks resembles the person shown as a challenge. We verify the answer with the entries in a tagged database and based on the answer, user is declared as pass or fail.

10.3.3 Security: Attacking iCAPTCHA

Attacking iCAPTCHA is difficult as computer programs are not yet advanced to automatically detect and label images in particular categories. An alternate attack can be by storing and searching for the images in Google image search engines. However, Google image search engine pages are dynamic in nature, which means the image that exist and ranked today may not be ranked in the same manner tomorrow.

WE further take necessary measure such as no two iCAPTCHA tests are similar in nature both in terms of the kind of images that it has and to whom it is given. As a result, attacker, same as user will receive a random iCAPTCHA test each time that has not completely similar to the tests he/she solved before. We recommend that large image database should be constructed from Google image search with large number of categories to avoid any database attacks.

10.3.4 Usability study

To test the liability of the proposed design, we conducted a preliminary lab study with eight participants. All the participants were from university campus with their age in the range of 22 to 28. Two participants were female while rest six participants were male. We fixed five sample categories, those are: *Apple, Cricket, Sachin tendulkar, Amrita and Rahul* (with which all users were familiar with). Task for each of the participants were to solve five iCAPTCHA tests. All the participants successfully completed all the five tests. Early feedbacks were extremely positive with most of them reporting satisfaction with the proposed approach and design. We know the numbers are not be satisfactory in terms of the population they represent, therefore, as a future work, we are in the process of conducting a large scale field study with the diverse population.

10.4 Summary

We described a novel CAPTCHA design, based on human ability to recognize images, label them and put them into proper categories. Benefit of our approach is getting the work of categorization and image annotation at virtually no cost. However, in doing so, we specially had taken care that the basic principles of CAPTCHAs like robustness and usability will not get affected. As a future work we are planning to launch an open source plug in of our proposed CAPTCHA design, and conduct a large scale field study.

11 Conclusion

"If the doors of perception were cleansed every thing would appear to man as it is, infinite. For man has closed himself up, till he sees all things through narrow chinks of his cavern."

William Black

English poet, painter, and printmaker, 1757-1827

11.1 Research contributions

The general research topic discussed in this thesis was whether perceptive intelligence could help in solving computationally hard problems. In particular, we studied two important problems: user authentication and image annotation.

User authentication

This work started with an overall aim of designing an authentication scheme that is memorable, secure and usable. We focused on graphical passwords, because of the proven human ability to recognize and remember previously seen images. The main research question was,

> "*Can we design a graphical authentication scheme that supports both memorability and security, while maintaining usability?*"

The identified three main research objectives summarized below.

Objective 1: Catalogue existing graphical password schemes focusing on user choices of password images. Identify the key design features that offer maximum benefits in terms of security, memorability, and usability.

(It turned out to be, system chosen images are most secure while personal, self chosen images are easily remembered and recognition based graphical passwords are most usable.)

Objective 2: Propose or identify an authentication design strategy that incorporates the key security, memorability, and usability features found in objective 1.

(This goal ended up in creating a Jigsaw based authentication design that provides the security of system chosen images with the memorability of self-chosen images.)

Objective 3: Create and empirically evaluate an authentication mechanism based on the design strategy devised in objective 2.

(We proposed a working prototype of jigsaw based authentication design, Marasim, using tagging.)

Image annotation

This work was intended at discovering effective ways for semantic annotations of images. Three research objectives for this work were described below.

Objective 4: Catalogue extant manual techniques of image annotation focusing on scalability, the quality of the annotation and the enjoyment people get while doing it. Identify problems with the manual annotation and probe for the reasons.

(It turned out to be, although extant methods are quite successful in luring human to annotate images, the quality of the resultant annotation is still far from perfection. Our investigation found two fundamental problems with manual annotation process.)

Objective 5: Discover an effective approach to manual annotation or alter existing approaches to get quality labels or descriptions for images.

(This goal ended up in presenting an emergent semantics approach to tagging that can effectively solve the problems we found in objective 4.)

Objective 6: Create and test annotation systems based on the approach found in objective 2.

(This goal ended up in creating two intelligent system designs in the form of a game and a CAPTCHA.)

We first present how our primary research contributions address the objectives set forth in this thesis. We then highlight some minor contributions.

11.2 Main contributions

To meet the first objective we examined the existing graphical password schemes, and discovered user tendencies of choosing predictable password images since they are easy to remember. In contrast, system chosen images are tough to remember and users often make mistakes in entering them in right order. In terms of usability, recognition based graphical passwords are better choice than recall or cued recall based graphical password scheme. We therefore, concentrated on creating a novel recognition based graphical password scheme that offers both security and memorability benefits.

To address the second objective, we investigated the feasibility of jigsaw based authentication design. The motivation behind jigsaw based approach was to blend together the security benefits of system chosen images with the memorability gains of self chosen images. Significant contribution of a Jigsaw based authentication approach is in password registration phase, where such an approach helps in creating a portfolio of password images that are both secure and memorable.

The third objective was met by designing, prototyping and testing a novel scheme, Marasim, based on the Jigsaw based authentication approach discussed earlier. Marasim leverages human ability to remember a personal image and later recognize the concepts occurred in that image. We conducted an extensive user study for three months to evaluate Marasim. Results of the user study provide evidence of improved memorability and usability.

To meet the fourth objective, we reviewed popular annotation techniques focusing on the enjoyability and the quality of the annotations. We found that although, these methods are quite successful in annotation process, the resultant annotations for the images may not be of a high quality. We investigated the reasons and discovered two fundamental problems with manual annotation process.

To address the fifth objective, we studied the emergent semantics theory. The study revealed that meaning of an image can be manifested by association between the images that share the same meaning. Based on the findings, we proposed a practical approach for semantic annotation of images.

The sixth objective is met by designing and testing two intelligent system designs for semantic annotation of images based on the emergent semantic approach we discovered earlier. First of the proposed design was a effective interactive game GoFish. GoFish is based on a popular playing card game with the same name. Another design is productive image recognition CAPTCHA, with a name iCAPTCHA. The iCAPTCHA design was also an improvement in terms of the accessibility compared with the traditional text based CAPTCHAs. We conducted preliminary user studies for both the designs. Results of user studies were encouraging.

11.2.1 Minor Contributions

This research also produced some minor contributions. Though, they were not directly mandated by our objectives. We list them below.

New Methodology of transforming popular games:

Existing approaches towards creating Games with a purpose has always been fixing the problem *X* and then building the game *G* that solves the problems. However, in such approaches, following points must be taken care to yield a viable solution.

1. *G* must solve the problem *X* correctly.
2. Introduce the elements of fun, and game play that overshadow the labor work of solving the problem *X*.

In short, the success of the game is proportional to the doing the labor work, while fun and the entertainments are the side products.

We introduce a novel approach of transforming an existing popular game into a game with a purpose. The advantage of this approach is we can capture the existing user database and there would be no need to introduce the fun and the game play.

Accessibility of the CAPTCHA

Our proposed CAPTCHA design, iCAPTCHA is more accessible than the traditional text based CAPTCHAs since recognizing images is easier than reading distorted text.

11.3 Research directions

This research has contributed to human computation and knowledge based authentication literature, but it has also raised further questions. In this section, we describe other projects resulting from this thesis. Members of our research lab are currently working on some of these projects, while other projects are yet to be undertaken.

Password strength meter

Most of the times, Users do not know their chosen password is secure or not. Providing a relative feedback on the password strength is good way to make them aware of the associated risks. Until recently there has not been much work in this regard. Our next logical step is to design a password strength meter for Marasim and conduct a full scale field study analyzing its impact on users.

Backup authentication:

Our proposed authentication scheme, Marasim can also be used as a backup authentication in case users forget their normal text passwords. However, to validate the claim, we must evaluate whether users can remember Marasim password images even after a sufficiently long time. Results of our user study showed that 93% users successfully remembered their password after three months. We are planning to take the study further, and verify the results after six or eight month's time gap between two logins. It can provide valid evidence for the use of Marasim as a backup authentication mechanism.

Multiple passwords inferences

One of the main reasons for users having troubles in remembering their passwords is there are too many password to remember. All these passwords put load on the memory, which results in users writing their password down or forgetting them frequently. We are currently conducting the field study for Marasim to test whether multiple password affect the memorability of Marasim.

Audio based challenge

Our proposed scheme in its present form will not work for visually impaired person since it is based on recognition of images. To accommodate them, we can present an alternative in the form of audio based challenge. When a visually impaired user is trying to login, the system will read aloud the tags for the images along with associated numbers. User then responds with the correct numbers that corresponds to the four tags she selected during registration. We are planning to test the working of this approach and associated tradeoff in terms of login delay and accuracy with user study with visually impaired persons.

Game theoretical proofs

GoFish game is classic example of how to transform an existing popular game into a game with a purpose. Although, results of the preliminary user study are encouraging and provide good evidences that the game could be fun and productive at the same time, the remaining thing to explore is a game theoretical proof for the same [Jain08] and an investigation of underlying design strategy.

Image ranking and categorization

In this thesis, we presented intelligent designs for semantic annotation of images. The same design with a little modification can also be used for ranking and categorizing images. We have just completed the design of a game for image categorization and planning for a user study for the same.

11.4 The last words…

The World is an amazing web of opportunities. On one side people are trying to build a human like thinking entity, while on the other side, people are also exploring the entertainment values on the World Wide Web, for impatient and idle youth. Human computation grabs both these opportunities, by leveraging human energy and skills for solving the problems that computer can not yet solve and providing them the desired entertainment in return.

In this thesis, we explored the possibilities of utilizing human perceptive intelligence towards solving computationally hard problems, with an overall aim of advancing the research in the field of human computation. We presented an authentication design in Marasim that offers improved security and memorability. Additionally, we described two alternate designs for semantic annotation of images.

Finally, what science has discovered about visual perception and the many ways that this knowledge can be applied to human computation is beyond the thesis. I hope, my work as a whole will provoke thinking in ways that advance the discourse in this arena.

Part III.

Appendix

A
Project Website

Personal home page

http://cstar.iiit.ac.in/~rohit/

Product Prototypes

All the product prototypes discussed in the thesis are available for testing and review at my personal website.

The individual page listing is as follows:

System or product	Address
Marasim: Graphical password scheme	http://cstar.iiit.ac.in/~rohit/marasim/
GoFish: Game	http://cstar.iiit.ac.in/~rohit/gofish/
iCAPTCHA design	http://cstar.iiit.ac.in/~rohit/icaptcha/

Table 3: Individual page listing at my webpage

References

[Adam99] Adams, A., and Sasse, M. A. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (Dec. 1999), 40-46.

[Ames07] Ames, M. and Naaman, M. 2007. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2007). CHI '07. ACM, 971-980.

[Barr97] Barry, A. M. S. 1997 *Visual Intelligence : Perception, Image, and Manipulation in Visual Communication.* New York Press.

[Bedw08] Bedworth, M. A Theory of Probabilistic One-Time Passwords. In *Proceedings of the 2008 International Conference on Security & Management, SAM*, (2008), 113-118.

[Bros00] Brostoff, S., and Sasse, M.A. Are Passfaces™ more usable than passwords? A field trial investigation. *Proceedings of HCI on people and Computers XIV*, (HCI 2000), 405-424.

[Cars96] Carson, C., and Ogle, V.E. Storage and Retrieval of feature data for vey large online Image collection. *IEEE Computer Society of the Technical Committee on Data Engineering*, 1996, Vol. 19 No 4.

[Chel05] Chellapilla K., Larson K., Simard P., and Czerwinski, Designing Human friendly human interaction proofs" ACM CHI 2005.

[Chew05] Chew, M., and Tygar, J. D. Image Recognition CAPTCHA. Technical Report No. UCB/CSD-04-1333. University of California, Berkeley 2005.

[Chia09] Chiasson, S. Usable Authentication and Click based Graphical passwords. Phd Thesis, Carlton University, Ottawa, Canada. Jan. 2009.

[Chia07] Chiasson, S., van Oorschot, P. C., and Biddle, R. Graphical Password Authentication Using Cued Click-points. In *12th European Symposium On Research In Computer Security (ESORICS)*, 2007.

[Cran05] Cranor, L., and Garfinkel, S. 2005. *Security and Usability: Designing Systems that People can use*. O'reilly Media.

[Datt08] Datta, R., Joshi, D., Li, J., and Wang, J.Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *IEEE Computing surveys*, 40(2008), 83-85.

[Davi04] Davis, D., Monrose, F., and Reiter, M. K. 2004. On user choice in graphical password schemes. In Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13 (2004). 11-11.

[Dean05] De Angeli, A., Coventry, L., Johnson, G., and Renaud, K. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies*, 63(1-2):128–152, 2005.

[Dham08] Dhamija, R., and Dusseault, L. 2008. The Seven Flaws of Identity Management: Usability and Security Challenges. IEEE Security and Privacy 6, 2 (Mar. 2008), 24-29.

[Dham00] Dhamija, R., and Perrig, A. 2000. Déjà Vu: a user study using images for authentication. In *Proceedings of the 9th Conference on USENIX Security Symposium - Volume 9* (2000). 4-4.

[Dham06] Dhamija, R., Tygar, J. D., and Hearst, M. 2006. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '06. ACM, New York, NY, 581-590.

[Diri07] Dirik, A. E., Memon, N., and Birget, J. 2007. Modeling user choice in the PassPoints graphical password scheme. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (2007). SOUPS '07, 20-28.

[Dunp07] Dunphy, P., and Yan, Y. Do background images improve "Draw a Secret" graph-
ical passwords? In *14th ACM Conference on Computer and Communications Security (CCS)*, October 2007.

[Elso07] Elson, J., Douceur, JR., Howell, J., and Saul. J. Asirra: a CAPTCHA that exploits interest-aligned manual image categorization. *Proceedings of the 14th ACM conference on Computer and communications security (CCS)*, 2007.

[Feld90] Feldmeier, D. C., and Karn, P. R. 1990. UNIX Password Security - Ten Years Later. In *Proceedings of the 9th Annual international Cryptology Conference on Advances in Cryptology* (1989). 44-63.

[Fisc87] Fischler, M. A., and Firschein O. *Intelligence: The Eye, the Brain and the Computer*. Addison-Wesley, 1987. ISBN: 0201120011.

[Flor07] Florencio, D., and Herley, C. 2007. A large-scale study of web password habits. In *Proceedings of the 16th international Conference on World Wide Web* (2007). WWW '07. ACM, 657-666.

[Gard93] Gardner, H. (1993) *Multiple Intelligences: The Theory In Practice*. New York: Basic Books.

[Gold02] Goldberg, J., Hagman, J., and Sazawal, V. Doodling our way to better authentication (student poster). In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2002.

[Goll07] Golle, P., and Wagner, D. 2007. Cryptanalysis of a Cognitive Authentication Scheme (Extended Abstract). In *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (2007). 66-70.

[Govi07] Govindarajulu, N., and Madhavnath, S. Password management using doodles. In *9th Int'l Conference on Multimodal interfaces. (ICMI)*, 2007.

[Haya08] Hayashi, E., Dhamija, R., Christin, N., and Perrig, A. 2008. Use Your Illusion: secure authentication usable anywhere. In *Proceedings of the 4th Symposium on Usable Privacy and Security* (2008). SOUPS '08, 35-45.

[Jaga07] Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. 2007. Social phishing. *Commun. ACM* 50, 10 (Oct. 2007), 94-100.

[Jain08] Jain, S., and Parkes D. C. A game-theoretic analysis of games with a purpose. In WINE'08, (2008), 342-250.

[Jerm99] Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., and Rubin, A. D. 1999. The design and analysis of graphical passwords. In *Proceedings of the 8th Conference on USENIX Security Symposium. 8*(1999). 1-1.

[Kinj00] Kinjo, H., and Snodgrass, J. G. 2000. Does the generation effect occur for pictures? *Amer. J. of Psych. 6(2000)*, 156-163.

[Kint70] Kintsch, W. Models for free recall and recognition. In D. Norman, editor, Models of human memory, chapter Models for free recall and recognition. Academic Press: New York, 1970.

[Kirk94] Kirkpatrick, E.A. (1894). An experimental study of memory. *Psychological Review, 1,* 602-609.

[Knop05] Knopf, M., Mack, A., Lenel, S., and Ferrante, S. Memory for action events: findings in neurological patients, *Scandinavian Journal of Psychology*. 46(2005), 11-19.

[Monc07] Moncur, W., and Leplatre, G. Pictures at the ATM: Exploring the usability of multiple graphical passwords. In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2007.

[Mori03] Mori G., and Malik J. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Vol 1, June 2003, pp.134-141.

[Morr79] Morris, R., and Thompson, K. Password security: A case history. *Communications of the ACM* (CACM Nov 1979), 594-497.

[Nali04] Nali, D., and Thorpe, J. Analyzing user choice in graphical passwords. Technical report, TR-04-01, School of Computer Science, Carleton University, May 2004.

[Nels76] Nelson, D. L., Reed, U. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 523-528.

[Nels79] Nelson, D.L. Remembering pictures and words: appearance, significance and name. *Levels of Processing and Human Memory.* (1979) 45-76.

[Niel08] Nielsen, T., Drewsen, P., and Hansen, P, Solving jigsaw puzzles using image features. *Pattern Recognition Letters*, 14(2008), 1924-1933.

[Ocon07] O'Connor, B. C., and Griesdorf, H. F. 2007. *Structures of Image Collections: From Chauvet-Pont-d'Arc to Flickr*, Libraries Unlimited.

[Paiv06] Paivio, A. *Mind and its evolution: a dual coding theoretical approach*. Lawrence Erlbaum: Mahwah, N.J., 2006.

[Peri03] Pering, T., Sundar, M., Light, J., and Want, R. 2003. Photographic Authentication through Untrusted Terminals. *IEEE Pervasive Computing* 2, 1 (Jan. 2003), 30-36.

[Prov08] Provos, N., Mavrommatis, N., Abu Rajab, M., and Monrose, F. All your iFrames point to us. In 17th USENIX Security Symposium, 2008.

[Rena09] Renaud, K. 2009. On user involvement in production of images used in visual authentication. J. Vis. Lang. Comput. 20, 1 (Feb. 2009), 1-15.

[Ross99] Ross, S. Unix System Security Tools. McGraw-Hill, 1999.

[Ross05] Ross, B., Jackson, C., Miyake, N., Boneh, D., and Mitchell, J. Stronger password authentication using browser extensions. *In 14th USENIX Security Symposium*, Baltimore, August 2005.

[Salt75] Saltzer, J., and Schroeder M. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[Sant01] Santini, S., Gupta, A., and Jain, R. Emergent semantics through interaction in image databases. *IEEE Trans. Knowl. Data Engg.* 13(2001), 337-251.

[Schn00] Schneier, B. 2000. *Secrets & Lies: Digital Security in a Networked World*. 1st. John Wiley & Sons, Inc.

[Sens07] Sen, S., Harper, F, Lapitz, A., and Riedl, J. The quest for quality tags. In *CSCW'07*, (2007), 371-380.

[Shep67] Shepard, R. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*. 6(1967), 156-163.

[Smeu00] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(12): 1349-1380 (2000).

[Smit08] Smith, G. 2008, *Tagging: People powered Metadata for the social web (Voices that matter),* New Riders Press.

[Stan73] Standing, L. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25 (1973), 207-222.

[Stub04] Stubblefield, A., and Simon, D. Inkblot Authentication. *Technical Report,* MSR-TR-2004-85, Microsoft Research, 2004.

[Suox05] Suo, X., and Zhu, Y. Graphical passwords: A Survey. In *Proceedings of Annual Computer Security Applications Conference. (2005).*

[Taoh08] Tao, H., and Adams, C. Pass-go: A proposal to improve the usability of graphical passwords. *Int'l Journal of Network Security.* 7(2008), 272-292.

[Tari06] Tari, F., Ozok, A. A., and Holden, S. H. 2006. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the Second Symposium on Usable Privacy and Security* (2006). SOUPS '06, 56-66.

[Thor07] Thorpe, J., and van Oorschot, P. C. 2007. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium* (2007). 1-16.

[Tull05] Tullis, T. S., and Tedesco, D. P. 2005. Using personal photos as pictorial passwords. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (2005). CHI '05. ACM, 1841-1844.

[Tulv73] Tulving, E., and Watkins, M. Continuity between recall and recognition. *American Journal of Psychology*, 86(4):739–748, 1973.

[Vona03] Von Ahn, L., Blum, M., Hopper, N., and Langford, J. CAPTCHA: Using Hard AI Problems For Security. In *Proceedings of Eurocrypt.* (2003) 294-311.

[Vona04] Von Ahn, L., and Dabbish, L. Labeling images with a computer game. In *CHI'04*, (2004), 319-326.

[Vona05] Von Ahn, L. 2005 *Human Computation*. Doctoral Thesis. UMI Order Number: AAI3205378., Carnegie Mellon University.

[Vona06] von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. 2006. Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '06. ACM, New York, NY, 79-82.

[Vona08] Von Ahn, L., and Dabbish, L. Designing games with a purpose. *Commn. ACM*, 51(2008), 58-67.

[Vonc08] Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, September 12, 2008. pp 1465-1468.

[Wein06] Weinshall, D. Cognitive Authentication Schemes Safe Against Spyware. In *Proc. 2006 IEEE Symposium on Security and Privacy (S&P),* May 2006.

[Wied05] Wiedenbeck, S., Waters, J., Birget, J., Brodskiy, A., and Memon, N. 2005. PassPoints: design and longitudinal evaluation of a graphical password system. *Int. J. Hum.-Comput. Stud.* 63, 1-2 (Jul. 2005), 102-127.

[Whit06] White, R.W., Kules, B., Drucker, S.M., and schraefel, m.c. (2006). Supporting Exploratory Search, Introduction to Special Section of Communications of the ACM, Vol. 49, Issue 4, (2006), pp. 36-39.

[Yanj07] Yan, J., and El Ahmad, A., S. Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms, in Proc. of 23rd Annual Computer Security Applications Conference (ACSAC'07). FL, USA, Dec 2007. IEEE Computer society. pp 279-291.

[Yane08] Yan, J. and El Ahmad, A. S. 2008. Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Proceedings of the 4th Symposium on Usable Privacy and Security* SOUPS '08, vol. 337. ACM, New York, NY, 44-52.

[Yanj08] Yan, J., and El Ahmad A., S. A Low-cost Attack on a Microsoft CAPTCHA, School of Computing Science Technical Report, Newcastle University, England. Feb, 2008.

Referenced Web Resources

[*Adob09*] Adobe Flash 8 http://www.adobe.com/products/flash/.

[*Amaz09*] Amazon Mechanical Turk http://aws.amazon.com/mturk/.

[*Bing09*] Bing Images. http://www.bing.com/images/.

[*Bfli09*] 4 Billion Photos on Flickr from Flickr Blog. http://blog.flickr.net/en/2009/10/12/4000000000/

[*Crow09*] Crowdsourcing. http://en.wikipedia.org/wiki/Crowdsourcing/.

[*Espg09*] ESP Game Dataset. http://www.cs.cmu.edu/~biglou/resources/.

[*Fall09*] Fallacy of misplaced concreteness. http://en.wikipedia.org/wiki/Fallacy_of_misplaced_concreteness/.

[*Fapi09*] Flickr API. http://www.flickr.com/services/api/.

[*Flic09*] Flickr. http://www.flickr.com.

[*Goog09*] Google Image Search. http://images.google.com/.

[*Gofi09*] Go Fish. http://en.wikipedia.org/wiki/Go Fish/.

[*Gofg09*] Online Game Gofish. http://www.littlebrowniebakers.com/games/gofish/gofish_play.htm.

[*Gwap09*] Games with a purpose. http://gwap.com/.

[*Hewe96*] Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., and Verplank W. *ACM SIGCHI Curricula for Human-Computer Interaction*. http://www.sigchi.org/cdg/index.html, 1996.

[*Hotc09*] Hot CAPTCHA. http://hotcaptcha.com/.

[*Jeff06*] Jefferson, G. Flickr of idea on a gaming project led to photo website. *USA Today*. http://www.usatoday.com/tech/products/2006-02-27-flickr_x.htm.

[*Jigs09*] Jigsaw Puzzle. http://en.wikipedia.org/wiki/Jigsaw_puzzle/.

[*Mala09*] Malware. http://en.wikipedia.org/wiki/Malware/.

[*Pass09*] PassFaces™. http://www.realuser.com/.

[*Reca09*] reCAPTCHA http://recaptcha.net/

[*Seti09*] SETI@Home. http://setiathome.berkeley.edu/

[*Sinh05*] Sinha, R. 2005. *The cognitive analysis of Tagging*. http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/.

[*Slas09*] Slashdot. http://slashdot.com/.

[*Smar09*] Smart Fox Server http://www.smartfoxserver.com/

[*Tagc08*] Image Recognition Problem Finally Solved: Let's Pay People To Tag Photos. http://www.techcrunch.com/2008/03/29/image-recognition-problem-finally-solved-lets-pay-people-to-tag-photos/

[*Tagg09*] Tag. http://en.wikipedia.org/wiki/Tag_(metadata)/.

Index