

UNIVERSITY OF CENTRAL FLORIDA  
FRONTIERS IN INFORMATION TECHNOLOGY  
COP 4910



## CLASS FINAL REPORT

### **Abstract**

This report brings together the final papers presented by the students in the Frontiers in Information Technology class, COP 4910 during the Spring-2015 semester. In addition, it is worth mentioning that this semester the students attended 76 talks and each student gave 4 presentations. In each talk they had to present the technical aspects of the selected topic along with its social impact, ethical aspects, and professional impact.

# CAPTCHA

## The Technology and its Applications

John P. Redmond and Wesley Mull  
COP4910  
University of Central Florida, Orlando, FL

**Abstract** — CAPTCHA is a modern security device used to determine whether a user is human or not. It is commonly found on websites that ask for user inputs, and it helps prevent spam and abuse by automated bots. CAPTCHA was developed on the premise that what can be easy for a human brain to decipher and infer can be nearly impossible for a computer. It has been an invaluable tool for businesses and website designers as the internet has developed, but it is not without its flaws. Common issues include tests that are too difficult and annoying for humans to solve, as well as the advancement of the computer technology to analyze and counter CAPTCHA's efforts at security. Despite these hindrances, CAPTCHA is still a necessary and developing feature of the internet, and there are many jobs that deal with the use of CAPTCHA both from a user's perspective and a security perspective.

**Keywords** — CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart); OCR (Optical Character Recognition); Google reCAPTCHA;

### I. INTRODUCTION

The inception of the internet brought with it a multitude of ways to quickly and easily communicate with many people across the world. Unfortunately, this great leap in communication ability also brought about new dangers with transactions, accounts, and information. Less than reputable users might try to attack other user accounts, distribute spam, or simply break into systems to acquire personal information. One technique created to help battle spam and brute force account hacks is CAPTCHA. According to their official website, CAPTCHA is “a program that protects against bots by generating and grading tests that humans can pass but current computer programs cannot” [1]. CAPTCHA is an acronym that stands for Completely Automated Public Turing test to tell Computers and Humans Apart. Figure 1 displays a common example of a reCAPTCHA problem. CAPTCHA has several security applications, including preventing comment spam in blogs, protecting website registration, protecting email addresses from scrapers, and preventing dictionary attacks.

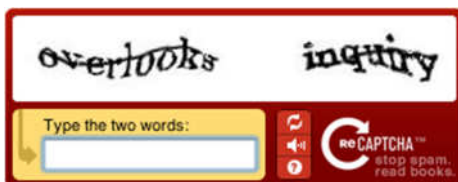


Figure 1: Common reCAPTCHA Example

Like many technologies, CAPTCHA finds its origins in the hacking community. To discuss topics of questionable legality, hackers devised a method of conversation which converted words to look-alike characters. An example of this is converting HELLO to |-|3|\_|\_(). In the 1980s, this method of symbol ciphers eventually became known as leetspeak (also called Leet or 1337) [2]. In 1997, the AltaVista team comprised of Lillibridge, Adabdi, Bharat and Broder, began work on a system to prevent internet bots from adding active URL's to the AltaVista search engine platform. To do this, the AltaVista team worked to prevent OCR (Optical character Recognition) attacks by building puzzles and images which would cause OCR attacks to fail. Essentially the AltaVista team worked to create a system of varied typefaces, backgrounds, type style and size which would fool OCR readers [2]. Later on in 2003, the process was refined and the term CAPTCHA was coined. This paper will detail CAPTCHA's technology, applications, related careers, and social impacts as it is ubiquitously used across the internet by many different entities as a reliable and basic form of user validation.

In section II, this paper will go over technical issues associated with CAPTCHA. Section III, will discuss the impact on society that CAPTCHA has had over the years with section IV, covering the ethical issues associated with the technology. Next, Section V, will be the professional impact of CAPTCHA and related job markets. Section VI, describes careers involved with CAPTCHA, and the requirements to get into this field. Lastly, section VII, summarizes the topics covered and provides a brief glimpse into the future of CAPTCHA.

### II. TECHNICAL DETAILS

CAPTCHA protects against attacks where automated programs attempt to flood an interface with requests or submissions [3]. Most CAPTCHAs rely on a visual test. This is the easiest way to make something a human can understand that a computer cannot. It is also important to note that CAPTCHAs must present different elements to different users. If the image was static, it would be simple to write a program for a computer to solve that one CAPTCHA. For non visual CAPTCHAs, a series of spoken letters or numbers is usually presented. It is common for the speakers voice to be distorted, or for background noise to be included. This helps to thwart

any voice recognition programs. Another type of CAPTCHA is to have the user read a passage and answer questions about the meaning of it. Computers can pick out key words, but are not good at understanding the meaning of them. This variation is not used often, however, as it takes the user a prolonged amount of time to complete [3].

The reason CAPTCHA is so effective is that humans can recognize abstract shapes or symbols and match them to known ones much easier than a computer. As seen in Figure 2, the left box is a rectangle of alternating grey blocks, and both humans and computers could easily recognize this pattern and explain it [12]. With the black removed, we can see many of the confusing elements that might make up a kind of CAPTCHA like test. Where a computer would only be able to process the image in terms of pixel color and location, humans can make out implied 3D shapes like the cylinders and boxes through the subtle color differences and shape configurations [13]. Humans can describe the abstract sensation of depth associated with layering different colored objects on top of and behind each other, while computers cannot yet create these abstract descriptions.

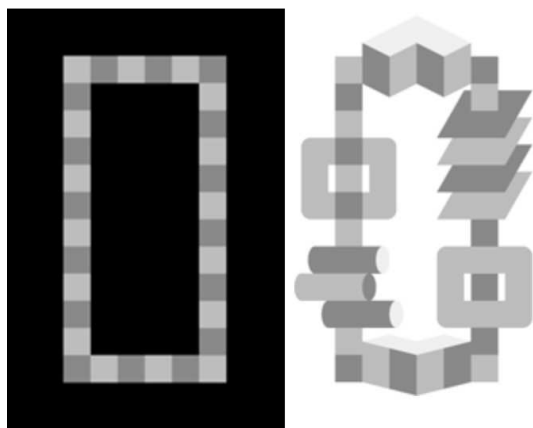


Figure 2: Easy and Hard Pattern Recognition

#### A. Applications of CAPTCHA

One way to easily generate CAPTCHA problems is by digitizing books. This works for the most common CAPTCHA type, distorted letters and numbers. An application called reCAPTCHA harnesses users responses in CAPTCHA fields to verify the contents of a scanned piece of paper. Because computers aren't always able to identify words from a digital scan, humans have to verify what a printed page says. Then it is possible for search engines to search and index the contents of a scanned document. First, the administrator of the reCAPTCHA program digitally scans a book. Then, the CAPTCHA program selects two words from the digitized image. The application already recognizes one of the words. If the visitor types that word into a field correctly, the application assumes the second word the user types is also correct. That second word goes into a pool of words that the application will present to other users. As each user types in a word, the application compares the word to the original answer.

Eventually, the application receives enough responses to verify the word with a high degree of certainty. That word can then go into the verified pool.

#### B. Types of CAPTCHA

As the internet grew so did the need for CAPTCHAs. In addition, the advances made in computer processing power and text, audio, and image recognition meant that CAPTCHA needed to evolve along with this technology. There have been many iterations of CAPTCHA services, starting simply with words distorted or overlaid on top of each other [7]. As this became easier for computers to decipher, more alterations and variations were added to muddle the problems further. First the problems were changed from preexisting words to randomized strings of characters. Numbers were added to increase the pool of characters that needed to be recognized to successfully complete the CAPTCHA problem [15]. These text and number CAPTCHAs are still used abundantly today for their simplicity and ease of implementation, but further developments were made to try and trick more sophisticated computers.

Image recognition is another type of prominent CAPTCHA problem, and is extremely effective in tricking computers. As previously mentioned, humans are able to much more easily infer and interpret shapes or symbols in images that computers have a hard time seeing as more than a sequence of pixels. Common images used are address numbers, text, or matching similar objects [9]. An extension of image recognition CAPTCHAs are animation, video, and 3D CAPTCHA. These are less common than images, and may require more resources to implement making them less desirable, but they have the potential to be stronger than simpler problem as well.

Last is audio recognition. These are another uncommon type of CAPTCHA, but potentially a very secure one. These use distorted audio as a problem for the user to solve. These are more difficult to implement than most CAPTCHA types simply due to the increased number of factors that may inhibit a user from correctly answering it. Some users might be hard of hearing, might not have adequate speakers or headphones, or be in an environment that would allow them to hear the prompt.

### III. SOCIAL IMPACT

As has already been mentioned, CAPTCHA is a fantastic tool for preventing spam on websites, moderating account registration, preventing certain automated attacks, and helping secure certain online activities. While many users might view CAPTCHA as a minor annoyance, most would agree it is a necessary addition to the modern internet [14]. However, a recent 6 month study by Casey Henry for the site moz.com shows that CAPTCHA might also hinder some human interaction. The study monitored a blog for six months, three of which used CAPTCHA in the comments section, and three of which had CAPTCHA turned off. The study found that while spam was reduced by 88% with CAPTCHA (11 messages vs. 91 with CAPTCHA off), there were also 159 failed postings of the 2,156 total the time CAPTCHA was on. Henry calls these "failed conversations", and while some of them may have been spam, it is also possible that some

people could not figure out the CAPTCHA phrase and gave up on their post [4]. Figure 3 provides a visual representation of the number of failed conversations relative to successful ones.

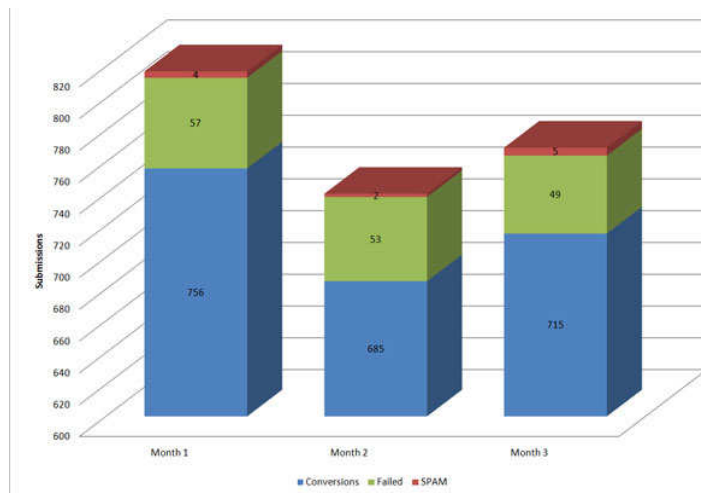


Figure 3: CAPTCHA Successful vs. Failed Conversations

#### A. CAPTCHA Effects on Conversation

While the study from reference [2] targeted a blog, the results can be applied to any online activity that requires a CAPTCHA. Account registration, commenting, participating in polls, or making purchases. Ultimately it is up to the individual website or company to decide if the implementation of CAPTCHA for its aforementioned benefits outweighs the loss of several customers, sales, or conversations.

#### IV. THE ETHICS OF CAPTCHA

CAPTCHA was designed from the beginning to help defend against irregular and illicit activity on the web. On one hand, it can be an incredible simple but potent tool in preventing malicious activity in any number of applications. In this way, CAPTCHA helps enforce an ethical standard of accepted internet practices, usually impeding the use of automation or bots to rapidly interact with website elements, and ensuring that the user is in fact human.

Recently however, Google came under fire for using CAPTCHA as a means to decode street numbers for its Google Maps site. The ethics of such a practice were called into question with regards to privacy and the unknowing participation of users in this interpretation of information. In 2012, Google is quoted as saying in response [6]:

"We are currently running an experiment in which characters from Street View images are appearing in CAPTCHAs. We often extract data such as street names and traffic signs from Street View imagery to improve Google Maps with useful information like business addresses and locations. Based on the data and results of these reCAPTCHA tests, we will determine if using imagery might also be an effective way to further refine our tools for fighting machine and bot-related abuse online."

It is still up for debate as to whether this is a breach of privacy or otherwise since without context the numbers and words are meaningless, but it shows that CAPTCHA has the potential to be used simultaneously for security and as a human optical character recognition platform [5].

#### V. PROFESSIONAL IMPACT

There are several professional impacts of CAPTCHA. One is that many companies feel compelled to have some sort of CAPTCHA if they are collecting accounts. Because so many things are done digitally, more and more companies are allowing online purchases [10]. This only grows the need for CAPTCHAs, which in turn grows the amount of jobs associated with them. This also affects current technical professionals, compelling them to need to know more about the subject of CAPTCHAs.

Due to its prevalence, it would be hard to have a professional career in technology and know nothing about CAPTCHA. As a whole, CAPTCHAs have made a lasting mark on the professional community and will continue to be used for the foreseeable future.

##### A. Job Market

Currently there are several versions and variations of the software, including number, word, and picture recognition. As far as marketing of CAPTCHA goes, it is a ubiquitous and well known feature of the internet.

There is the shadier subset of the job market associated with CAPTCHA. These jobs involve people solving CAPTCHA problems for very little pay for larger employers. Many of these employers are using these solved CAPTCHAs to send spam to emails, message boards, comments sections, and more, or for more sinister purposes such as hacking into accounts through password restoration tools. These jobs can usually be taken from a mobile device or any computer, and simply require the user to solve as many CAPTCHAs as they have time for. However the pay is understandably low, on the order of \$1 for every thousand correctly solved CAPTCHA problems.

##### B. Necessary Background

Careers related directly to CAPTCHA are few and far between, as the system undergoes few revisions. Google has been developing a less intrusive CAPTCHA system, though this has been done somewhat behind closed doors to help prevent security issues [16]. While careers specifically developing CAPTCHA systems are rare, there are many professions that make use of CAPTCHA.

#### VI. CAREERS

While there is minimal active marketing for CAPTCHA, it is widely known due to its prevalence on the internet. Companies, small business owners, website administrators, or anyone else with any kind of input on their websites are aware of, and in many cases expected to use CAPTCHA. A brief overview of several CAPTCHA-related careers, including common skills and average salary, can be found in Table I.

Web developers are the most obvious related field to CAPTCHA. Skills needed to work as an independent contractor web developer range from knowing HTML5/CSS3 and Javascript to Ruby, PHP, and MySQL. It is a large and still growing field, and expected salaries range from \$65,000-\$85,000 [8]. Knowing how to implement CAPTCHA is a necessity in most modern web design, as many companies or contractors will expect the added security on their site or form.

Another very relevant job related to CAPTCHA solutions is the security testing engineer. Through penetration testing, secure code analysis, hacking skills, CAPTCHA breaking, and other means, you are expected to help find flaws in security so future attacks can be prevented or avoided. The average salary for a security testing engineer is anywhere from \$70,000-\$120,000 [8]. This kind of job would provide a very behind the scenes view of the inner workings of website security and CAPTCHA, but requires the unique skillset needed to intentionally break websites and security systems.

#### RELATED FIELDS OVERVIEW

CAPTCHA Related Careers	RELATED FIELDS OVERVIEW	
	Common Skills	Average Salary
Web Developer	HTML5/CSS3, C++/C#, Javascript, PHP, MySQL	~\$70,000
Web Designer	UI Design, Adobe Products, HTML5/CSS3	~\$59,000
Penetration Tester	Web Security & Encryption, Network Security	~\$77,000

#### VII. CONCLUSION

Simply put, CAPTCHA is an elegant solution to a pervasive problem on the internet. It is a system that uses the brain's power of inference to verify that a user is who they say they are. It can be seen as a minor annoyance to many who browse the web, but it prevents spam and brute force attacks on many sites that require user input. There are not many careers specifically in the development of CAPTCHA itself, but there are several related to its use such as web developers or security testing engineers. While its implementation can vary from simple to complex, and it might be used for other information gathering in addition to its security like in the case of Google's reCAPTCHA, it is now a universal part of the internet.

Further developments of CAPTCHA technologies are in testing today. Video CAPTCHA is rare, but could be one of the more common CAPTCHA variants in the future [11]. Google is rolling out a new version of reCAPTCHA as of early 2015. This new version of CAPTCHA improves vastly on every aspect the service provides. It aims to be more accurate in detecting automated attacks and bots, while simultaneously being less obstructive to users. It does this through analyzing how a user fills out forms or interacts with a website. This new form of CAPTCHA could be as easy as interacting normally with the site or form. Google hopes to

provide advanced spam and abuse security, while making user interaction effortless.

#### References

Carnegie Mellon University. "CAPTCHA: Telling Humans and Computers Apart Automatically". Internet: <http://www.CAPTCHA.net/> 2000. [Feb. 18, 2015]

Yale, Brad. "The CAPTCHA: A History, A Problem, Possible Solutions". Internet: <http://www.informit.com/blogs/blog.aspx?uk=Why-Are-CAPTCHAs-So-Awful> Sept. 10, 2014. [Feb 18, 2015]

Kanimozhi R, Jagadeesan D. Authenticating a webpage using CAPTCHA image. *International Journal Of Advanced Research In Computer Science* [serial online]. September 2014;5(7):133-135. Available from: Applied Science & Technology Full Text (H.W. Wilson), Ipswich, MA. [ Mar. 18, 2015]

Plotts, Andrea. "How Captcha can Affect Conversion Rates". Internet: <http://www.semgeeks.com/blog/how-CAPTCHA-can-affect-conversion-rates#.VOp8GfnF8rU> Nov. 5, 2013. [Feb. 19, 2015]

Lung, Jonathan. "Ethical and Legal Considerations of reCAPTCHA: Part I". Internet: <http://www.cs.toronto.edu/~lungj/blog/?p=1307> Jun. 19, 2012 [Feb. 18, 2015]

Menon, Vineet. "Is it ethical to use ReCAPTCHA to decode house addresses?" Internet: <http://security.stackexchange.com/questions/60879/is-it-ethical-to-use-reCAPTCHA-to-decode-house-addresses> Jun 12, 2014 [Feb. 18, 2015]

Vu Duc N, Yang-Wai C, Susilo W. On the security of text-based 3D CAPTCHAs. *Computers & Security* [serial online]. September 2014;45:84-99. Available from: Business Source Premier, Ipswich, MA. [ Mar. 22, 2015]

Author Unknown. "Web Developer Salary". Internet: <http://www.indeed.com/q-CAPTCHA-jobs.html> 2015 [Feb. 18 2015]

Luis von Ahn, Manuel Blum, Nicholas J. Hopper, John Langford , *Advances in Cryptology*. Springer-Verlag Berlin Heidelberg; 1 ed, Vol. 1. 2003.

Steffen Zorn, Pedram Hayati, Marketing Dynamism & Sustainability: *Things Change, Things Stay the Same....* Springer International Publishing; 1 ed , Vol . 1.2008.

Anjali Avinash Chandavale, A. Sapkal, *Recent Trends in Computer Networks and Distributed Systems Security*. Springer Berlin Heidelberg; 1 ed , Vol . 1. 2008.

Mori, G. Malik, J. , "Recognizing objects in adversarial clutter: breaking a visual CAPTCHA" Vol . 1. 2003.

Sungho Kim Gi-Jeong Jang, Wang-Heon Lee, In So Kweon, "How human visual systems recognize objects - a novel computational model." Vol . 1. 2004.

Jeff Yan, Ahmad Salah El Ahmad , "Usability of CAPTCHAs or usability issues in CAPTCHA design". Vol . 1. 2008.

Starostenko O, Cruz-Perez C, Uceda-Ponga F, Alarcon-Aquino V. Breaking text-based CAPTCHAs with variable word and character orientation. *Pattern Recognition* [serial online]. April 2015;48(4):1097-1108. Available from: Academic Search Premier, Ipswich, MA. [ March 22, 2015]

Google " reCAPTCHA: The FREE anti-abuse service." Internet: <https://developers.google.com/reCAPTCHA/> Dec. 3, 2014 [Mar. 22, 2015]