

Cracking Captchas For Cash: A Review Of CAPTCHA Crackers

Merrill Serrao
St. Francis Institute of
Technology, Mumbai

Shanu Salunke
St. Francis Institute of
Technology, Mumbai

Amrita Mathur
(Assistant Professor)
St. Francis Institute of
Technology, Mumbai

Abstract

The CAPTCHA [Completely Automated Public Turing test to tell Computers and Humans Apart] was developed order to prevent automated bots from performing actions that were meant for human users. However, ever since the birth of CAPTCHAs, they have been broken for both educational and malicious purposes. There are many strategies that can be employed to break CAPTCHAs, each with their own advantages and disadvantages. This document reviews various methods of bypassing CAPTCHAs as well as popular commercial software used for this purpose.

1. Introduction

The Internet is a limitless source of resources. Ever since its conception thirty years ago, it has grown into this web of service providers and consumers offering and using a wide variety of both free and commercial services. Today, the traditional business model has been replaced by the internet based business model with entrepreneurs dealing with goods via an online marketplace. Many people earn their living by dealing in information through blogs, websites, ebooks, etc. There is also a class of businessmen who earn their livelihood by using underhand marketing techniques, now called spam. As the Internet and computer science evolved, programmers were able to write scripts that could sign up for free services and then use these to send mail to a large number of people. This practice has buried hundreds of internet businesses, forcing many to shut their doors.

These automated scripts have been used not only to send spam, but also to bias the outcome of online polls, to spam forums and even in chat rooms.

In the year 2000, four students from Carnegie Mellon University came up with an innovative method to prevent bots from performing functions as users. They called it the CAPTCHA. The CAPTCHA is a challenge-response test that asks a user a question that only a human user would be able to answer. The

original CAPTCHA was an image of a scheme of letters that was displayed and the user was asked to read the letters on the screen. This worked beautifully because of a machine's inherent inability to "read" like human beings do. With advances in image processing and machine learning techniques, programmers were able to easily bypass this test. In retaliation, the CAPTCHA was strengthened and branched out into the various forms of CAPTCHAs found today – image based, audio based, interaction based, etc.

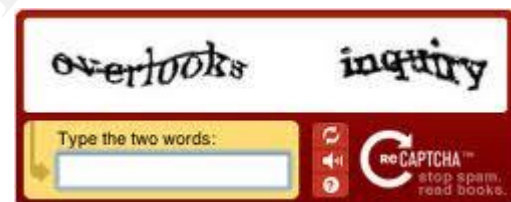


Figure 1: A sample CAPTCHA^[1]

2. Cracking CAPTCHAs

Cracking of CAPTCHAs refers to writing scripts that can identify the text in the CAPTCHA image and hence, pass the test. Bypassing CAPTCHAs refers to those mechanisms that allow a user to completely avoid having to deal with the CAPTCHA test. There are three main methodologies to cracking or bypassing CAPTCHAs.

2.1 Optical Character Recognition

Optical Character Recognition [OCR] is the process of understanding the contents of an image. This was developed in order to digitize large quantities handwritten and printed documents. OCR is virtually solved thanks to advances in Image Processing and Machine Learning techniques.

The general structure of an OCR mechanism has three stages – pre-processing, segmentation and recognition. The pre-processing stage attempts to clean the image of stray lines, background noise and any other unnecessary details. It usually binarizes the image

so that only the most important information is passed for further processing. The segmentation stage is generally the most difficult stage of OCR. Here, one attempts to break down the image and identify Regions of Interest (ROIs). These may be words, letters or any other region that will be processed further. The last stage – recognition involves identifying the region and classifying it as one from the domain of possible answers. For text documents, this could be the set of letters and alphabets. The recognition stage may employ feature extraction as an intermediary step or use some other recognition procedure. Machine Learning structures like Neural Networks and Support Vector Machines are often used in this stage.

2.2 CAPTCHA Farms

A CAPTCHA Farm refers to a large group of people who answer CAPTCHA tests for the benefit of some paying party. Many services employ human beings to answer CAPTCHAs and then sell these answers by volume. A simple API is used to integrate the client's script to an online human solver. Given the sheer number of people who use the Internet daily and the fact that breaking most CAPTCHAs comes easily to human beings and requires no training, this is a simple and economical solution for most CAPTCHA cracking service providers.

2.3 CAPTCHA Smuggling

CAPTCHA Smuggling is a technique wherein a human user is tricked into solving a CAPTCHA by a malware that injects CAPTCHAs into the user's regular surfing session. Most users do not keep track of the CAPTCHAs that they are solving. A slight increase in the volume of CAPTCHA tests goes unnoticed. These solved CAPTCHAs can then be used by a paying client for regular use or malicious purposes.

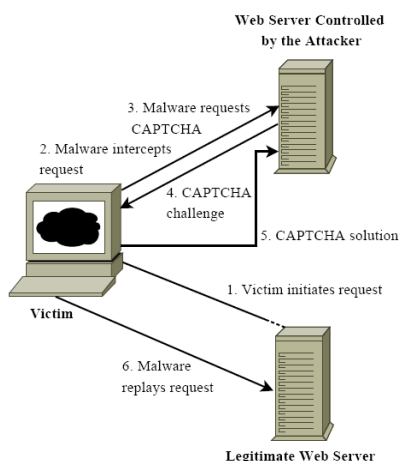


Figure 2: CAPTCHA Smuggling^[2]

3. Commercial CAPTCHA Cracking Services

Captcha bypass services offer users to bypass any CAPTCHA they encounter and are paid for by volume or time period subscription.

3.1 Death By Captcha^[3]

DeathByCaptcha is the most popular CAPTCHA bypassing service on the internet. It uses a combination of OCR and human solvers, offering to decode over 1000 CAPTCHAs for just over a dollar. This service is used by implementing their API in one's project and passing the CAPTCHA to it. Advantages of using DeathByCaptcha:

- It is a hybrid system composed of sophisticated OCR and a 24/7 team of human solvers.
- It has an easy to use API in a wide variety of programming languages.
- DeCaptcha and Antigat API support is provided to make mitigation to DeathByCaptcha easier.
- The user is only charged for the correctly identified CAPTCHAs.

DBC charges \$1.39 for every 1000 CAPTCHAs correctly solved. The average response time is 15seconds and it boasts of an accuracy rate of 90%.

Use of DBC for spamming and other illegal activities is prohibited as per their terms of use.

3.2 CaptchaSniper^[4]

CaptchaSniper detects CAPTCHAs by using the popular DeathByCaptcha service via HTTP post using the DeathByCaptcha API implementation and other services. It returns the decoded CAPTCHA to the software that placed the request. CaptchaSniper also includes a command line interface.

Advantages of using CaptchaSniper:


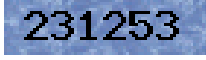




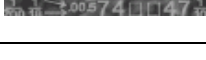

- CaptchaSniper solves CAPTCHAs faster and without charging per CAPTCHA
- It solves hundreds of different CAPTCHAs for popular platforms like wordpress, Movable Type and many others. It works with popular tools like Scrapebox, Sick Submitter and more.
- It is easy to use as no code integration or API is required.

CaptchaSniper is available after a one time purchase and is valid for one PC. It runs on Windows XP, Vista and Windows 7.

CaptchaSniper was developed for use by research groups and for the visually challenged. Using this service for illegal activities is prohibited by their terms of service.

CaptchaSniper currently supports the following platforms and their associated CAPTCHAS:

Table 1: Success Rate of CaptchaSniper^[4]

Platform/ Footprint	CAPTCHA Image	Success Rate
Wordpre4s s Blogs		76%
Lifetime Blogs		100%
ArticleMS Directories		64%
phpBB2 Forums		61%
Kohana		65%
SMF Forums		64%
Vivvo CMS		99%
Generic Bookmark/ RSS Directory		28%

3.3 Automated CAPTCHA Bypass (ACB)^[5]

ACB can be used to automatically translate an image into the underlying text. One needs to specify the image URL in the script and the result of the recognition is given at the output. It requires Perl support. However, ACB bypasses only the CAPTCHAs shown below:

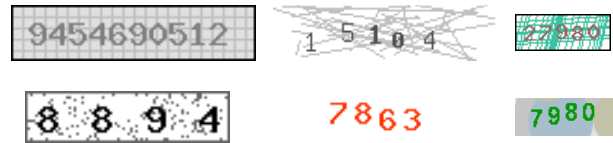


Figure 3: CAPTCHAs Bypassed by ACB

Advantages:

- Easy to use
- Bypasses some popular CAPTCHAS
- Setup requires only image url
- No need to install any special modules or software

3.4 Captcha Monster – Automatic CAPTCHA Filling Firefox Add-On^[6]

The Captcha Monster is an add-on available with the Firefox Browser which solves CAPTCHAs without the user asking it to. It was designed for the benefit of dyslexics, the visually challenged and anyone who did not want to be served with a CAPTCHA test.

The Captcha Monster works in three modes:

- Solves CAPTCHAs only on request (default)
- Solves the CAPTCHA when the website is loaded
- Solves the CAPTCHA when the user starts filling a form with a CAPTCHA

Advantages:



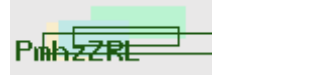




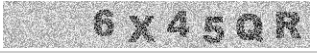

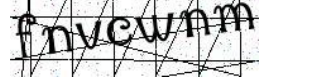
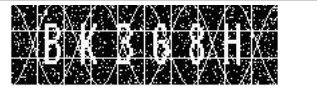

- Works with a wide range of CAPTCHAs
- Complete Automation
- Dedicated Support
- High success rate

The basic version costs \$2.99 monthly. It takes an average 8.5 seconds to solve a CAPTCHA and has a 98.94% success rate.

3.5 PWNTCHA – Captcha Decoder^[7]

PWNTcha stands for “Pretend We’re Not a Turing Computer but a Human Antagonist”. The project’s goal was to demonstrate the inefficiency of many CAPTCHA implementations. PWNtcha was started as a personal research project and has not been updated for the past three years.

Table 2: CAPTCHAs Defeated By PWNTCHA^[7]

Origin	Samples	Efficiency
Authimage		100%
Clubic		100%
linuxfr.org		100%
Live Journal		99%
lmt.lv		98%
Ourcolony		100%
Paypal		88%
phpBB		97%
Scode and derivatives		100%
Slashdot		89%
vBulletin		100%
Xanga		49%

4. Conclusion

The CAPTCHA is the most widely used mechanism to control spam on the Internet. Most current CAPTCHAs have been broken and new ones are being developed everyday. CAPTCHAs are routinely broken using both targeted attacks as well as commercial software. This paper attempts to review the popular CAPTCHA cracking services available on the Internet today. These services use a combination of Optical Character Recognition and social engineering to bypass all types of CAPTCHAs.

5. References

- [1] "CAPTCHA: Telling Humans and Computers Apart Automatically", www.captcha.net accessed on 12/1/2013.
- [2] Manuel Egele et. al "CAPTCHA Smuggling: Hijacking Web Browsing Sessions to Create CAPTCHA Farms", Proceedings of the 2010 ACM Symposium on Applied Computing
- [3] "CAPTCHA Bypass Done Right", www.deathbycaptcha.com accessed on 5/1/13
- [4] "Captcha sniper", <http://www.captchasniper.com> accessed on 5/1/2013
- [5] "Guruperl - Automated CAPTCHA bypass, ACB", www.guruperl.net/products/captcha-bypass accessed on 7/1/2013
- [6] "Captcha Monster - CAPTCHA solving add-on", <http://www.captchamonster.com> accessed on 8/1/2013
- [7] "PWNTcha - Captcha Decoder", <http://caca.zoy.org/wiki/PWNTcha> accessed on 5/1/13
- [8] Hacker Intelligence Initiative, Monthly Trend Report #11 "A CAPTCHA in the Rye" ADC Monthly Web Attacks Analysis, June 2012.
- [9] K Chellapilla and P simard, "Using Machine Learning to Break Visual Human Interactive Proofs (HIPs)", Advances in Neural Information Processing Systems 17, Neural Information Processing Systems (NIPS), MIT Press, 2004
- [10] G Mori, J Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA" in Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 134-141, 2003
- [11] M blum, A. von Ahn and J Langford, The Captcha Project, "Completely Automated Public Turing Test to tell Computers and Humans apart", Dept of Computer Science, Carnegie Mellon University, 2000
- [12] Chellapilla, K., Larson, K., Simard, P.Y., Czerwinski, M.: Designing Human Friendly Human Interaction Proofs (HIPs). In: van der Veer, G.C., Gale, C. (eds.) CHI, pp. 711–720. ACM (2005)
- [13] Imsamai, M., Phimoltare, S.: 3D CAPTCHA: A Next Generation of the CAPTCHA. In: Proceedings of the International Conference on Information Science and Applications (ICISA 2010), Seoul, South Korea, April 21-23, pp. 1–8. IEEE Computer Society (2010)
- [14] Li, S., Shah, S.A.H., Khan, M.A.U., Khayam, S.A., Sadeghi, A.-R., Schmitz, R.: Breaking e-Banking CAPTCHAs. In: Gates, C., Franz, M., McDermott, J.P. (eds.) ACSAC, pp. 171–180. ACM (2010)
- [15] Macias, C., Izquierdo, E.: Visual Word-based CAPTCHA using 3D Characters. IET Seminar Digests 2009(2), P41–P41 (2009)
- [16] OCR Research Team. Teabag 3D CAPTCHA, <http://ocr-research.org.ua>
- [17] Ross, S.A., Halderman, J.A., Finkelstein, A.: Sketcha: a CAPTCHA based on Line Drawings of 3D Models. In: Rappa, M., Jones, P., Freire, J., Chakrabarti, S. (eds.) WWW, pp. 821–830. ACM (2010)

- [18] Yan, J., Ahmad, A.S.E.: Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms. In: ACSAC, pp. 279–291. IEEE Computer Society (2007)
- [19] Yan, J., Ahmad, A.S.E.: A Low-Cost Attack on a Microsoft CAPTCHA. In: ACM Conference on Computer and Communications Security, pp. 543–554 (2008)
- [20] Yan, J., Ahmad, A.S.E.: Usability of CAPTCHAs or Usability Issues in CAPTCHA Design. In: Cranor, L.F. (ed.) SOUPS. ACM International Conference Proceeding Series, pp. 44–52. ACM (2008)

IJERT